

# Search Engine Yearbook **2003**

**Free Version : : March 2003**

Previously known as "The MOTHER of all Search Engine Reference Books"

Presented by André le Roux ([andre@pandecta.com](mailto:andre@pandecta.com))

**Published & distributed by Pandecta Magazine**

If this is your first time reading a book in Acrobat Reader, we have some handy tips prepared that will save you time. [Click here.](#)

## **Important links to the web:**

SEY 2003 Order Page: <http://www.pandecta.com/sey.html>

Pandecta Magazine Homepage: <http://www.pandecta.com/>

If you have ideas & suggestions for SEY 2004, please [tell us](#).



## **Text colors in the book & what they mean:**

- Black** = Normal text
- Red** = Highlighted / emphasized text
- Green** = Highlighted / emphasized text
- Blue** = Links to the web
- Orange** = Internal links (links to other sections in this book)
- Purple** = Content only available in the full version

# Foreword

CLICK HERE TO  
JUMP DIRECTLY  
TO THE TABLE  
OF CONTENTS

## What's In This Free Version Of SEY 2003?

I wanted to give you a taste of the **full version** without actually giving you the entire 401-page book for free, so I simply took that book, **removed certain sections and replaced the Foreword with the one you're reading now.** Other than that, the free version is exactly like the real thing.

## So Do I Really (*REALLY REALLY*) Need The Full Version?

When you cut away all the **hype and bull**, understanding the search engine game is the one thing that we all **HAVE** to get right.

**Once you understand search engines,** You will be able to **find what you need** when you need it and you will be able to **attract visitors** to your web site. With both of these abilities in your arsenal, the Internet is at your fingertips.

The **full version** of SEY 2003 delivers **both.**

Besides, if you order the full version and you're not 100% blown away, **one** e-mail to Pandecta support gets you a **full, immediate and unconditional refund.** You can try it and get your money back if you don't like it.

## Updates

Only the full version is regularly updated. This free version is **updated about once a year**. Want to know when we update this free book? Just send a blank e-mail to this address: [mother-update-subscribe@topica.com](mailto:mother-update-subscribe@topica.com) and I'll add your name to the list of people who get a note from me each time a new free book is ready.

You can keep up with changes in the search engine industry by subscribing to my free EnginePaper newsletter. It goes out only when something significant changes in the search engine world.

To subscribe, simply send a blank e-mail to [send-ep-subscribe@topica.com](mailto:send-ep-subscribe@topica.com)

## Ordering The Full Version Of This Book

To say thanks for trying this free version, I'll give you the **full version at a 15% below what everyone else is paying.**

### How much is that exactly?

That depends on **when** you order. I started the book at the ridiculous price of \$17 when I first published it. As the book becomes more popular, so the price increases. At the time of writing it's at \$29. It might be more by the time you read this. The thing is not to wait too long. **The sooner you act, the less you'll pay.** And remember: You still get 15% off whatever the price is when you order.

[Order now via this link to qualify for the 15% discount](#)

## Summary info

(CLICK HERE FOR THE TABLE OF CONTENTS)

This is the **free version** of the **Search Engine Yearbook 2003**

For more on the full version, check the previous page.

To order the full version (@ 15% discount), click [here](#).

The order page for the full version (@ 15% discount) is <http://www.pandecta.com/se-15.html>

To receive a note when we **update** this book, send a blank e-mail to [mother-update-subscribe@topica.com](mailto:mother-update-subscribe@topica.com)

To stay up to date on **changes in the search engine industry**, subscribe to the "EnginePaper" newsletter. Send a blank e-mail to [send-ep-subscribe@topica.com](mailto:send-ep-subscribe@topica.com)

This book is © Copyright 2003, [Pandecta Magazine](#). You may redistribute this book freely, in electronic format or otherwise, provided that it is not changed in any way and not sold. If you paid for this book, drop us a line at [legaldesk@pandecta.com](mailto:legaldesk@pandecta.com). Thanks.



# Table of Contents

## SUMMARY:

<b>Section 1</b>	(page 10 to 159):	<a href="#">The Search Engines</a> --- Info like URLs, stats, relationships etc.
<b>Section 2</b>	(page 160 to 174):	<a href="#">Resources for Search Engine Users</a> --- Helping you find info
<b>Section 3</b>	(page 175 to 247):	<a href="#">Search Engine Optimization</a> --- All about getting visitors to your site
<b>Section 4</b>	(page 248 to 269):	<a href="#">SEO Resources</a> --- Webmaster help (tools, tutorials etc.)
<b>Section 5</b>	(page 270 to 301):	<a href="#">Outsourcing Search Engine Optimization</a> --- Possible pitfalls
<b>Section 6</b>	(page 302 to 394):	<a href="#">The Search Engine Dictionary</a> --- 335 search engine terms explained
<b>Section 7</b>	(page 395 to 400):	<a href="#">General Information</a> --- About SEY 2004, about Pandecta etc.

## SECTION 1 – THE SEARCH ENGINES

### PAGE

1.1	<a href="#">How Search Engines Work</a>	11
1.2	<a href="#">Shortcut Page To The Major Search Engines</a>	15
1.3	<a href="#">The 10 Major Search Engines Reviewed</a>	16
1.3.1	<a href="#">Google</a>	16
1.3.2	<a href="#">AltaVista</a>	20
1.3.3	<a href="#">Yahoo</a>	27
1.3.4	<a href="#">Overture</a>	30
1.3.5	<a href="#">DMOZ (ODP)</a>	33
1.3.6	<a href="#">Excite</a>	37
1.3.7	<a href="#">Lycos</a>	39
1.3.8	<a href="#">AlltheWeb</a>	41
1.3.9	<a href="#">Teoma</a>	44
1.3.10	<a href="#">Ask Jeeves</a>	48
1.4	<a href="#">Google Spotlight</a>	51
1.4.1	<a href="#">Google Today</a>	51
1.4.2	<a href="#">Google Features</a>	53
1.4.3	<a href="#">Google Power Player (Interview with Sergey Brin)</a>	56
1.4.4	<a href="#">AdWords</a>	60
1.4.5	<a href="#">PageRank</a>	62
1.4.6	<a href="#">Do's And Don'ts</a>	70

Purple shows links to sections that are only available in the full version of the book. Click anywhere in this block to order your copy of the full version

**Remember:**  
**Orange** = internal links.  
 Click orange links to flip to that section in the book.

**Want to print this TOC?**  
 Try this [printer-friendly HTML version](#) available on the Pandecta web site.

1.4.7	The Google Dance	74
1.4.8	Freshness & Everflux	76
1.4.9	More Google Resources	78
1.5	About Inktomi	80
1.6	About AOL Search	81
1.7	About MSN Search	82
1.8	About LookSmart	83
1.9	About HotBot	85
1.10	About Wisenut	86
1.11	The 117 Search Engines & Directories Worth Knowing About	87
1.12	Topical Search Engines & Directories	100
1.13	252 Country-Specific Search Engines	111
1.14	Important, New Search Engines	128
1.15	Other Noteworthy Search Engines	130
1.16	Spiders & Robots	132
1.17	Stats: Relative Database Sizes	134
1.18	Stats: Estimated Total Database Sizes	136
1.19	Stats: Average Speed	138
1.20	More Search Engine Statistics	139
1.21	Search Engine Relationships	141
1.22	Search Engine News	143
1.23	Telephone Directories	145
1.24	Meta Searching	146
1.25	The Future of the Search (by Detlev Johnson)	150
1.26	Who Will Be The Next Google? (by Jill Whalen)	155

**SECTION 2 – RESOURCES FOR SEARCH ENGINE USERS**

2.1	Internet Search Strategies: An Internet Search Tutorial	161
2.2	More Tutorials on Internet Searching	168
2.3	Articles on Internet Searching	171
2.4	General Resources for Search Engine Users	173

**Purple shows links to sections that are only available in the full version of the book. Click anywhere in this block to order your copy of the full version**

**Remember:**  
**Orange** = internal links.  
 Click orange links to flip to that section in the book.

**Want to print this TOC?**  
 Try this **printer-friendly HTML version** available on the Pandecta web site.

SECTION 3 – SEARCH ENGINE OPTIMIZATION (SEO)

3.1	Overview of the Search Engine Industry	176
3.2	Overview of Web Marketing Techniques	178
3.2.1	Search Engines	178
3.2.2	Link Building	180
3.2.3	Word Of Mouth	181
3.2.4	Online Advertising	182
3.2.5	Offline Advertising	183
3.3	SEO Facts	184
3.3.1	Content Is (Still) King	184
3.3.2	Keyword Targeting	185
3.3.3	Invisible Text	188
3.3.4	Resubmission	189
3.3.5	Search Engines That Matter	190
3.3.6	Domain Names	192
3.3.7	Cross-Linking	195
3.3.8	Dedicated IP Addresses	197
3.3.9	Robots.txt and the Robots Meta Tag	198
3.3.10	Link Building	201
3.4	SEO “Maybes”	206
3.4.1	Getting Doorway Pages Right	206
3.4.2	Updated Thinking On Meta Tags	210
3.4.3	Submission Software	216
3.4.4	Cloaking	219
3.5	Getting Listed at DMOZ (ODP)	224
3.5.1	Before You Submit	226
3.5.2	Finding The Right Category	227
3.5.3	About Regional Sites	228
3.5.4	About Adult Sites	229
3.5.5	About Affiliate Sites	230
3.5.6	Your Submission	231

Purple shows links to sections that are only available in the full version of the book. Click anywhere in this block to order your copy of the full version

**Remember:**  
**Orange** = internal links.  
 Click orange links to flip to that section in the book.

**Want to print this TOC?**  
 Try this **printer-friendly HTML version** available on the Pandecta web site.

3.6	Getting Pay-Per-Click Marketing Right	234
3.7	Why Can't I Get My Site Listed?	238
3.7.1	Browser Requirements	238
3.7.2	Frames	240
3.7.3	Automatic Redirects	241
3.7.4	Google Minimum PageRank	242
3.7.5	Free Space	243
3.7.6	Blocking Spiders	244
3.8	If You Can't Beat'em, Delete'em	245

**SECTION 4 – SEO RESOURCES**

4.1	SEO Tutorials	249
4.2	SEO Articles	254
4.3	SEO Tools	255
4.3.1	Keyword Tools	255
4.3.2	Log File Analyzers	258
4.3.3	Search Engine Position Checkers	259
4.3.4	Link Popularity Tools	260
4.3.5	Other Useful Tools	261
4.4	SEO Newsletters / E-zines	263
4.5	SEO Discussion Forums	265
4.6	Other SEO Resources	266
4.7	Other Ways To Promote Your Site	267

**SECTION 5 – OUTSOURCING SEARCH ENGINE OPTIMIZATION (SEO)**

5.1	Introduction: The Importance Of Proper Search Engine Optimization	271
5.2	Basics of Search Engine Optimization	273
5.2.1	Types Of Search Engines	274
5.2.2	How Search Engines Work	277
5.2.3	Keyword Targeting	280

**Purple shows links to sections that are only available in the full version of the book. Click anywhere in this block to order your copy of the full version**

**Remember:**  
**Orange** = internal links.  
 Click orange links to flip to that section in the book.

**Want to print this TOC?**  
 Try this **printer-friendly HTML version** available on the Pandecta web site.



5.2.4	Submitting Your Site	284
5.2.5	Tracking And Improving Results	286
5.3	Should You Outsource Search Engine Optimization?	287
5.4	The Truth About Search Engine Optimization Providers	289
5.5	Four Warning Signs	291
5.6	Questions To Ask SEO Providers	293
5.6.1	Link Popularity	294
5.6.2	Keyword Targeting	296
5.7	About Guarantees	298
5.8	About The Contract	299
5.9	Finding SEO Providers	300
5.10	How To Report Dishonest SEO Providers	301

**SECTION 6 – THE SEARCH ENGINE DICTIONARY**

6.1	About The Search Engine Dictionary	303
6.2	The Search Engine Dictionary: 335 Terms Explained	306

**SECTION 7 – GENERAL INFORMATION**

7.1	About SEY 2004 And Your 25% Discount	396
7.2	How To Earn A FREE Copy of SEY 2004	397
7.3	Priority Customer Support	398
7.4	About The Author	399
7.5	About Pandecta Magazine	400

	Copyright Notice & Disclaimer	401
--	-------------------------------	-----

**Purple shows links to sections that are only available in the full version of the book. Click anywhere in this block to order your copy of the full version**

**Remember:**  
**Orange** = internal links.  
 Click orange links to flip to that section in the book.

**Want to print this TOC?**  
 Try this **printer-friendly HTML version** available on the Pandecta web site.

# Section 1: The Search Engines



# 1

## The Search Engines

### Important Note

When I made this free version, I simply took the full version and chopped it down to about half the original size - so there are big chunks of the book left out. You'll probably find instances of orange links (internal links) in this free version that seems broken. They're not really broken. They point to content only available in the full version. So if you click a link and nothing happens, that's why :-)

Remember, the full version comes with a full money back guarantee. It's a risk free purchase. Details [here](#).

# 1.1 How Search Engines Work

Let's start by distinguishing between **search engines** and **directories**.

---

## Search Engines

(like [www.google.com](http://www.google.com))



The main characteristic of search engines is that they rely on **spiders** to crawl the web, indexing pages as they go. Spiders are browser-like programs that follow links from page to page and from site to site, indexing everything it finds.

When you submit a web page to a search engine, all you really do is tell the spider about the page.

Your page does not get added to the search engine's database immediately – that only happens once the spider gets around to visiting and indexing the page.

---

## Directories

(like [dmoz.org](http://dmoz.org))



Directories do not use spiders.

Instead, they use **real people** (editors) who visit and **evaluate** sites – and add them **only** if they meet the directory's minimum quality requirement.

This is an important difference:

- **Search engine spiders can index thousands of pages a day.**
- **Directory editors cannot.**

**So why do we have directories if they can't compete?** The answer is **quality**.

Editors are considerably harder to impress than spiders. The page has to offer **unique information** or a **unique product**. When you submit a page to a specific category in a directory, the editor of that category will visit your page and decide if it's good enough to add to the directory.

Editors usually reject pages with typos, broken links, unclear navigation etc.

---

## The Components

**Search engines and directories all consist of 5 major components:**

1. The spider (or editor in the case of directories)
2. The indexer (again the editor in the case of directories)
3. The database
4. The search software
5. The interface

### 1. The spider

Sometimes called a robot, this is a browser-like program who's job it is to retrieve a web page, read it, send it to the indexer, follow a link to the next page, read it... and so on. Important to remember is that the spider does not "see" the page. It looks at the page source. To see what the spider sees, simply open a site and from the browser (IE) menu, select "View" and then select "Source".

### 2. The indexer

It's the indexer's job to analyze the data received from the spider before dumping it into the database. It analyzes the various elements of each page, looking at things like the title, headings, body text, links etc.

### 3. The database

Search engine databases are massive "copies" of the web. It does not contain replicas of web pages, but information on each web page the indexer analyzed. Most search engines store only key information on each page. Only full-text search engines store every single word.

### 4. The search software

This is the part that matters. It is here where decisions are made (based on the search engine's algorithm) about which pages to list in response to a query and also, very importantly, in which order to list them. Search engine optimization (SEO) specialists spend a lot of time trying to understand how each search engine ranks web pages.

## 5. The interface

This is the part that you and I see. The web page, search box, advertisements etc. This is where a search starts. The text entered in the search box (the query) is sent to the search software which in turn “pages through” the database, finds all the relevant documents, sorts them from most relevant to least relevant and sends it back to the user in the form of search results. All in a fraction of one second. Not bad.

--- S I D E B A R ---

### **Confused by the terminology?**

**Learn some search engine lingo...**

Most of the search engine terms used in this book are explained in the **Search Engine Dictionary section**.

You can also download the dictionary as a separate, **free** e-book. Visit [www.searchenginedictionary.com](http://www.searchenginedictionary.com) for details.

## 1.2 Shortcut Page to the Major Search Engines

<b>Google</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	<a href="#">Submit Your Site Here</a> (Free)
<b>AltaVista</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	<a href="#">Submit Your Site Here</a> (Free or “Express”)
<b>Yahoo</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	Suggest in appropriate category (Pay for review: \$299 annually)
<b>Overture</b>	<a href="#">Main Search Page</a>	--	<a href="#">Submit Your Site Here</a> (Pay-per-click)
<b>DMOZ</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	Suggest in appropriate category (Free)
<b>Excite</b>	<a href="#">Main Search Page</a>	--	Submit to <a href="#">Google</a> , <a href="#">LookSmart</a> , <a href="#">Inktomi</a> , <a href="#">Ask Jeeves</a> , <a href="#">About</a> , <a href="#">Overture</a> , <a href="#">FindWhat</a> or <a href="#">AllTheWeb</a> . Paid inclusion also available.
<b>Lycos</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	<a href="#">Submit Your Site Here</a> (Pay-per-click / Paid inclusion)
<b>AlltheWeb</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	<a href="#">Submit Your Site Here</a> (Paid Inclusion via Lycos / Free)
<b>Teoma</b>	<a href="#">Main Search Page</a>	<a href="#">Advanced Search</a>	<a href="#">Submit Your Site Here</a> (Pay for review via <a href="#">Ask Jeeves</a> : \$30 first URL, \$18 per URL thereafter) or submit to <a href="#">DMOZ</a> )

Note added for the free version: □

If you click an orange link and nothing happens, it means that link □ points to something only available in the full version.

## 1.3 The Major Search Engines



### 1.3 THE MAJOR SEARCH ENGINES

#### 1.3.1 Google ([www.google.com](http://www.google.com))



### URLs

Main:	<a href="http://www.google.com/">http://www.google.com/</a>
Advanced search:	<a href="http://www.google.com/advanced_search.html">http://www.google.com/advanced_search.html</a>
Submission page:	<a href="http://www.google.com/addurl.html">http://www.google.com/addurl.html</a>
Contact page:	<a href="http://www.google.com/contact/">http://www.google.com/contact/</a>
Physical address:	2400 Bayshore Parkway, Mountain View, CA 94043
Phone number:	650 330 0100 (8:30 a.m. - 6:00 p.m. PST)
Google images:	<a href="http://images.google.com">http://images.google.com</a>
Google groups:	<a href="http://groups.google.com">http://groups.google.com</a>
Google directory:	<a href="http://directory.google.com">http://directory.google.com</a>
Google preferences:	<a href="http://www.google.com/preferences/">http://www.google.com/preferences/</a>
About Google:	<a href="http://www.google.com/about.html">http://www.google.com/about.html</a>



Google Toolbar: <http://toolbar.google.com/> (Highly recommended)  
Google AdWords (Paid listings): <https://adwords.google.com/select/?hl=en>

## The Company

Google has only been around since September 1998 – surprising when you consider how far they are ahead of the other search engines today. The company was founded by Larry Page and Sergey Brin.

Google is a privately held company with (at the time of writing) just over 400 employees.

## Google & YAHOO!

Google still supplies web results to compliment Yahoo directory results – only now the Google results are shown first.

This has **HUGE** implications. **For one thing, a Google listing now reaches almost twice as many eyeballs.** Also see the [discussion on Yahoo](#) for a more detailed look at what this change means to us in terms of SEO.

## Google & AOL

Google powers [AOL Search](#). Below is an extract from a Google Press Release:

*Under the agreement, Google's search technology will begin powering the search areas of AOL, CompuServe, AOL.COM and Netscape this summer. By joining Google's industry-leading platform with America Online's extensive consumer audience and popular online brands, the companies plan to create an even better search experience for AOL's **more than 34 million members** and **tens of millions of visitors** to America Online's Web-based properties, both domestically and internationally.*

**To summarize: Getting Google right is crucial**, because your Google listing reaches not only Google and Yahoo users but also everyone using AOL Search.

(Not many other search engines left, are there?)...

It's worth noting that the **paid listings** at AOL (previously supplied by [Overture](#)) are now supplied by Google AdWords.

## Google & Search Engine Optimization

We estimate that Google results now reach **75 to 80% of all search engine users**.



**Yes, that's 75 to 80% !!!**

Those that don't use Google directly see results **supplied** by Google – either

at AOL Search, Yahoo or one of the smaller search engines powered by Google.

This immense reach means that Google absolutely **HAS** to be the focus of your search engine optimization efforts. **Fortunately for us, Google is fairly easy.**

For starters, submitting your site to Google is free.

There is a rumor floating around SEO forums that the site submission service at <http://www.google.com/addurl.html> is only there to humor us. That Googlebot (Google's spider) has more than enough URLs in its "to-do" list. Besides, Google only lists web sites that has at least some inbound links – and if it has inbound links, Googlebot will pick it up on its own.

This theory seems fairly credible, but unlikely. At Pandecta, we still submit all our new sites – just to be sure. There's no harm.

A popular misconception is that Google penalizes sites for regular resubmission. Most other search engines do, but Google clearly states on their site that they do not. There is however no point to regular resubmission as it will not improve your site's rank.

For **more on Google**, please refer to the **"Google Spotlight"** section of this book. Jill Whalen's article, **"Who Will Be The Next Google?"** is also a **must-read**.



**Only in the full version:**



**Reviews of AltaVista, Yahoo!, Overture, DMOZ (ODP), Excite, Lycos, Teoma, AlltheWeb, Ask Jeeves.**



-----

**Not in the free version: p21 to p50**

-----



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## 1.4 Google Spotlight

### Contents at a glance:

[Google Today](#) || [Features](#) || [Power Player](#) (an interview) || [AdWords](#) || [PageRank](#) ||  
[Do's & Don'ts](#) || [The Google Dance](#) || [Freshness & Everflux](#) || [More Google Resources](#)



### 1.4 GOOGLE SPOTLIGHT

#### 1.4.1 Google Today

When I first published the “Mother of All Search Engine Reference Books” in 2000, I went out on a limb calling Google the number one search engine. If you’ll allow a soapbox moment, today I can say I told you so. **Google ‘kicks butt’** – even in China.

So for both search engine users and site owners, getting Google right has become more important than ever before. Google consistently returns more relevant results than any other search engine – at a speed that makes searching frustration-free.

Understandably, it's the first place Web surfers look for information.

This 'close-up' is intended to help you get maximum value from Google. Let's look at some features...



**Only in the full version:**



**Google Features, Google Power Player (interview with Sergey Brin), Google AdWords.**



**Not in the free version: p54 to p61**



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**



## Acknowledgement

This explanation of Google's PageRank system is based in part on the explanations offered by [Phil Craven](#) and [Ian Rogers](#).

## What PageRank (PR) Is

Google's measure of the number & quality of inbound links to a web site. The PageRank (PR) of each page is one of the about 100 criteria Google uses to rank web pages.

## How Much PageRank Matters

It is only one of 100 criteria Google uses, but from experience I'm convinced that it weighs quite heavily in Google's ranking algorithm. Pages with a high PR value usually outrank pages with a low PR value.

Some search engine experts feel that webmasters in general assign too much value to PR – and they are probably right. A high PR is only valuable if the page is properly optimized for the keywords it targets. The Google homepage has a perfect PR of 10, but it does not rank first for every keyword search.



## How PageRank is calculated

Google measures the number and quality of links to a page – both links from outside the site and links from other pages in the same site.

**The PR formula is:**

$$PR(A) = (1-d) + d(PR(t1)/C(t1) + \dots + PR(tn)/C(tn))$$

Don't be discouraged. It's not as difficult as it looks.

Before I explain how it works, I should mention that this is the **original formula** used by Larry Page and Sergey Brin when they developed the PageRank system. It is likely that the formula has been tweaked since then.

## The Formula Made Easy

A	= The page for which we want to calculate PR
t1 to tn	= All the pages linking to page A
C	= The number of outbound links each page has
d	= A damping factor (set to 0.85)

**Let's do an example:**

## EXAMPLE: Calculating PR

Page A has inbound links from pages X, Y and Z. Pages X and Y each have only one outbound link: The one to page A. But page Z has three outbound links of which only one points to page A. Page X has a PR of 1, page Y has a PR 2 and page Z has a PR 3.

Here's this example's formula:

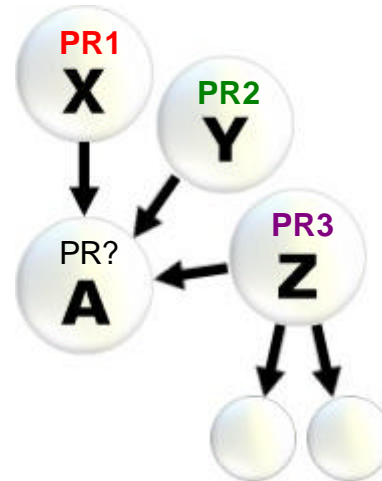
$$PR(A) = (0.15) + 0.85(1/1) + 0.85(2/1) + 0.85(3/3)$$

$$PR(A) = 0.15 + 0.85 + 1.7 + 0.85$$

$$PR(A) = 3.55$$

So page A has a PR of 3.55.

Black shows page A, red shows page X, green shows page Y and purple shows page Z.



Each page has a PR1 to start out with. When it links to another site, it has 0.85 (the damping factor) worth of muscle to vote with. But that 0.85 has to be distributed between all outbound links, so if there are 2 outbound links, each receiving site gets only 0.425 worth of PR added to their existing 1 PR point.

## PageRank 11?

Did you spot that? In the example, if we had a couple more inbound links to A, the PR would increase above 10 (10 is supposed to be the maximum). Well, 10 isn't really the maximum. It is a **symbolic value** assigned by Google to pages with the highest PR.

It could be that PR 1-10 is shown as 1, PR 11-100 is shown as 2 etc. or Google can assign 10 to the highest scoring site and assign the other 9 values proportionately to that.

No-one outside Google knows for sure.

## Once Isn't Enough

**Here's something to wrap your brain around...**

In the example above, we assumed that X had PR1, Y had PR2 and Z had PR3. But how does Google know that? **What if A linked to Z?** Then Z's PR might jump to 4 – which means A's PR might jump to 4 – which means Z's PR increases again etc.

We need A's PR to get Z's PR, but we can't get A's PR until we have Z's PR.

The solution is to repeat the calculation a couple of times. No matter how many times the calculation is repeated, the values will never be 100% accurate, but after about 50

iterations it starts settling down to the point where there's no significant change in PR with new iterations.

## Total PageRank

**Ok, get a fresh cup of coffee, let the cat out and put the kids to bed. This is where it really begins to matter...**

### Ready?

In Google's eyes, every page on the web starts out with a PR of 1. So if you have a 20-page site, your site's total PR is 20, distributed evenly between the 20 pages (provided that there are no inbound or outbound links).

By linking poorly, **it is possible to lose some of that 20 PR points.**

Remember the damping factor (0.85)? That is how much of its 1 PR point each page can give away. **The important thing is that, according to the original PR formula, that 0.85 is always subtracted – even if there are no outbound links.** So a page with no outbound or inbound links has a PR of only 0.15.

The lesson is that **every page on your site should link to another page on your site,** even if they all link only to the homepage. That way each page gives its 0.85 to the homepage. If the homepage links back to each of the internal pages, that PR is redistributed to the internal pages.

## Channeling PageRank To Important Pages

You don't necessarily want all our pages to have an equal share of the site's total PR. It would be ideal if you could channel some of that to pages optimized for competitive keywords.

Well, you can. **This is where kicking the butts of the big players becomes reality...**

### EXAMPLE: Channeling PR

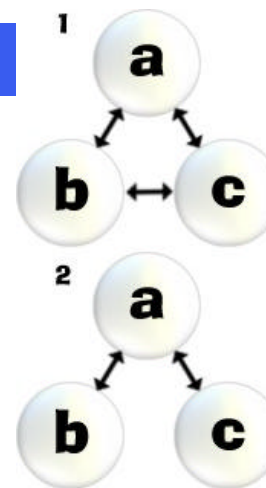
In the illustrations of internal site structures to the right, the first shows a site where all pages link to all pages. No PR is wasted and all pages have an equal share (PR1 each).

In the second illustration, the link between b and c is dropped. Every page still links to at least one other page, so no PR is wasted, but the distribution of the site's total PR of 3 (one PR point per page) is not even. Here's what happens when we run this second structure through the PR formula:

**Page a = 1.85**

**Page b = 0.575**

**Page c = 0.575**



But remember, once isn't enough. After 100 iterations it's clear that page 'a' comes out of this one the winner.

**Page a = 1.459459**

**Page b = 0.7702703**

**Page c = 0.7702703**

The total is still 3, so no PR is wasted.

## Dangling Links

In the original research paper, Brin and Page define dangling links as ***“links that point to any page with no outgoing links.”*** These present a problem for the PR formula since it isn't clear where their weight should be distributed. The solution is to remove them at the start of the calculation and add them back in at the end. That way they do not influence the PR calculation for other pages.

**Having dangling links in your sites will hurt your site's total PR.**

Any page that has no outbound links contributes only 0.15 to the site's total PR (1-d). They don't hurt other pages since Google drops them from the calculations, but consider adding at least one link from every page on your site to anywhere else in the site.

## Further Reading

### That's about as much of that as my brain can process...

If you're just getting warmed up, I suggest you head to Phil Craven's PageRank paper. It's called "**Google PageRank And How To Make The Most Of It**". Here's the URL:

<http://www.webworkshop.net/pagerank.html>

Phil even built a **fantastic** PageRank calculator that lets you quickly evaluate different linking structures:

[http://www.webworkshop.net/pagerank\\_calculator.php3](http://www.webworkshop.net/pagerank_calculator.php3)

And if you want more when you're done with Phil's paper, here's a similar one by Ian Rogers:

<http://www.iprcom.com/papers/pagerank/>

Google's (short) explanation of PageRank:

<http://www.google.com/technology/index.html>

The original paper by Larry Page & Sergey Brin:

<http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>

*--- A special word of thanks to Phil Craven for his input ---*



**Only in the full version:**



**Google Do's & Don'ts, The Google "Dance", Google Freshness & Everflux, More Google Resources, About Inktomi, About AOL Search, About MSN Search, About Looksmart, About HotBot, About Wisenut, The 117 Search Engines & Directories Worth Knowing About, Topical Search Engines & Directories, 252 Country-Specific Search Engines,**



-----

**Not in the free version: p71 to p127**

-----



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/se.html>**



## 1.14 Important New Search Engines

[www.wondir.org](http://www.wondir.org)



The Wondir Foundation is a new, nonprofit, 501(c)(3), organization. Their mission is simple: **eliminate the barriers between questions and answers.**

The most exciting thing about this new search engine is that it's nonprofit. Without the pressure of having to make money, they have a real advantage over other search engines in that they can **FOCUS on relevance of search results.**

But that's not the only promising thing about this search engine...

They say they want to "connect people with information needs with the people and information that can help them". **In short, if your search results are unsatisfactory, you can ask an expert.** And "the service will be **free** to all and open to all."

### You Can Help Wondir

**Donations to the Wondir Foundation are tax-deductible.**

They also need people to help with the open-source development of the technology and they need experts to help answer searcher questions (a great way to establish yourself as an expert in your field).

 [www.turbo10.com](http://www.turbo10.com)

Here is another ambitious & very promising project.

The UK-based Turbo 10 search engine provides **access to both the “surface web” and the “invisible web”** (or DeepNet as they call it). “Surface web” refers to those documents that normal search engines can index – things like html, pdf, doc etc.

The “invisible web” is that part of the web that normal search engines can’t index – files that are publicly available but “invisible” to most of us. These are typically contained in specialist databases from business associations, universities, libraries and government departments.

The “Turbo 10 Trawler” connects to these specialist databases – and it does so dynamically the moment you hit “Search”. Your query is also passed to surface web search engines.

An interesting twist is that Turbo 10 serves results as fast as they become available. Results from the fastest search engine are displayed first.

For a list of invisible web resources that Turbo 10 searches, take a look at: <http://turbo10.com/collections.html>

## 1.15 Other Noteworthy Search Engines



A priceless addition to the search engine world. The Wayback Machine's database has cached versions of pages from 1996 onwards.

<http://web.archive.org/>

### Newspapers.Com

A great specialty search engine. It finds not only newspapers but all kinds of publications. (Not limited to the U.S.)

<http://www.newspapers.com>



The American government search engine.

<http://www.firstgov.gov/>



This search engine is listed in this category for one reason: It claims to have 3.5 billion web pages in its index, putting it right up there with [Google](#). Personally I'm skeptical. The site is in beta testing but messy even for a beta test. I'll keep an eye on this one and report on it in the EnginePaper Newsletter. Subscribe with a blank e-mail to [send-ep-subscribe@topica.com](mailto:send-ep-subscribe@topica.com).

<http://www.openfind.com>

## 1.16 Spiders & Robots

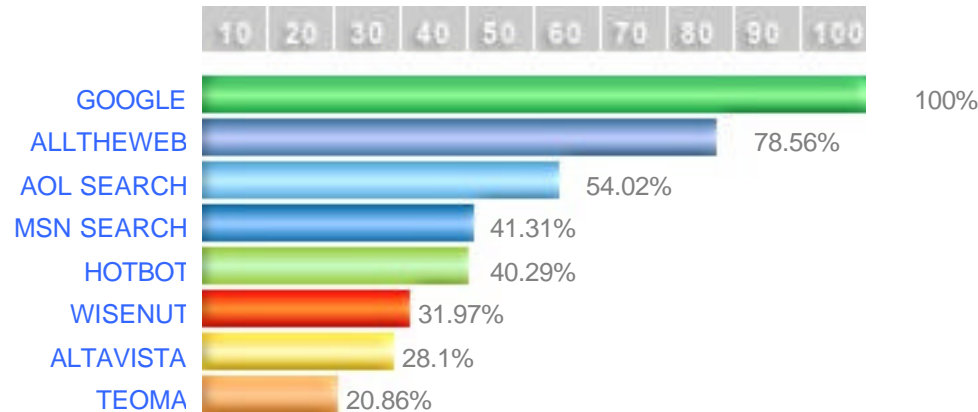
Spiders are **browser-like programs** that automatically surf and index the web. Spiders follow links from one page to the next and from one site to the next. The term robots is sometimes used to refer to spiders, but it is in fact a collective name for a group of programs of which the spider program is one.

Here are some spider names you might see in your log files and the search engine they're from. **If I missed any that you know of, please suggest them.** If I use your suggestion, your name is added to the list of people who will **get SEY 2004 for free.**

Search engine	Spider name
Abacho	AbachoBOT
Aesop	AESOP_com_SpiderMan
Ah-ha	ah-ha.com crawler
Alexa	ia_archiver
AltaVista	Scooter
AlltheWeb	FAST-WebCrawler
Atomz	Atomz
Excite	ArchitextSpider
Euroseek	Arachnoidea
EZResults	EZResult
Google	Googlebot

Inktomi	Slurp.so/1.0
	Slurp/2.0j
	Slurp/2.0
	Slurp/3.0
Lexis-Nexis	LNSpiderguy
LookSmart	MantraAgent
Lycos	Lycos_Spider_(T-Rex)
Mirago	HenryTheMiragoRobot
Northernlight	Gulliver
National Directory	NationalDirectory-SuperSpider
Openfind	Openfind piranha, Shark
SearchHippo	Fluffy the spider
Teoma	teoma_agent1
Ttravel Finder	ESISmartSpider
UKSearcher	UK Searcher Spider
Walhello	appie
Websmostlinked	Nazilla
Wisenuit	ZyBorg

## 1.17 Stats: Relative Database Sizes



### NOTES

1. The study was conducted in the 4<sup>th</sup> quarter of 2002.
2. The values above are not indicative of actual database sizes. Rather, they indicate database sizes of some of the major search engines **relative to the size of the Google database**. The Teoma database, for example, is about 5 times smaller than the Google database.
3. The values were arrived at by conducting 30 single-word searches, adding up the total number of results returned by each search engine and translating that number to a percentage of the total number of results returned by Google.

4. The search terms were not chosen randomly. They were mostly English and mostly without any geographic connotation. On average, the number of results returned per search engine per word had to be 1000 or less. This was to ensure that one term could not dominate the results.

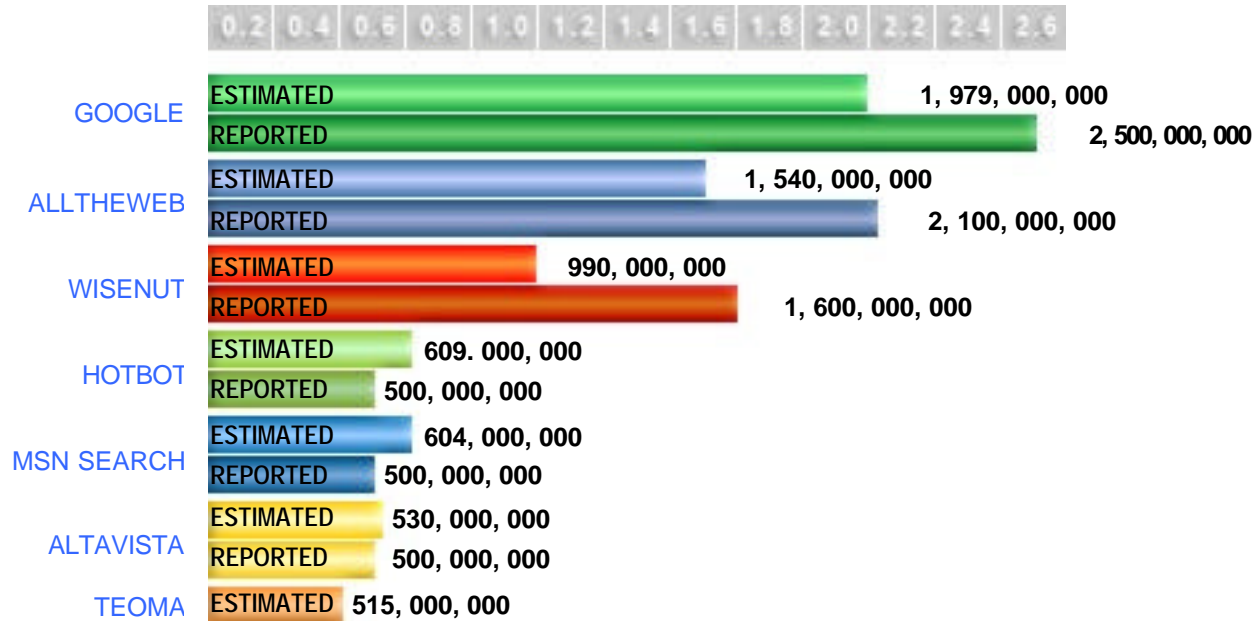
## REMARKS

- Google includes sites in its database that it only “knows about” (through links from other sites), but that Googlebot has not actually spidered. Google’s database also includes file types (like PDF) not usually indexed by other search engines.
- AOL did pretty well, but it should be noted that this is mainly due to their partnership with Google, whereby Google supplies results to the “matching sites” category of their results. They have their own database maintained by AOL editors, but it is fairly small.
- Wisenut and Teoma faired poorly, considering early claims that they where both capable of displacing Google from the #1 spot. Teoma’s paid inclusion program is probably a major contributor to its comparatively small database.





# 1.18 Stats: Estimated Total Database Sizes



## NOTES

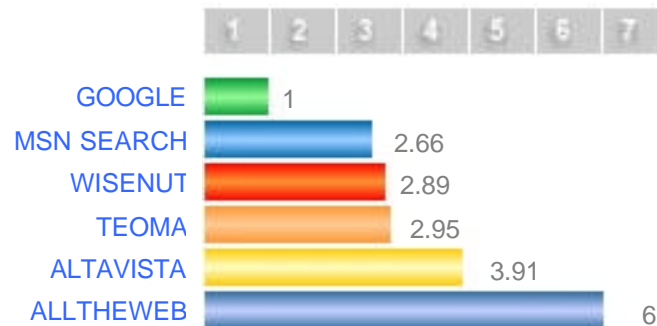
1. This study was conducted in the 4<sup>th</sup> quarter of 2002.
2. The results are our own findings and, although we consider them to be fairly accurate, they were not confirmed by the search engines and they should therefore not be regarded as official.

3. The estimated values are the average of the reported database size at the time, the estimated database size reported on SearchEngineShowdown.com and our own estimate based on the relative search engine database size reported in the previous graph.
4. Discrepancies between estimated values and reported values are due to many factors. Our study of relative database sizes was fairly small (30 search terms) and therefore cannot be regarded as 100% accurate. Search engine also typically spread their databases over several servers, any number of which may have been unreachable or down for maintenance at the time the study was conducted.
5. No reported database size for Teoma was available at the time of this study, nor would they give any specifics when asked. Teoma was also not included in SearchEngineShowdown.com's study. The estimate displayed above reflect only our own estimate.
6. AOL receives results from Google and was therefore not included in this study.

## 1.19 Stats: Average Speed

From time to time I compare search engine speeds for my own reference. The study is far from comprehensive, but it gives a general idea of how the search engines measure up. I thought I'd share it with you. **Please take note that these numbers are based on a fairly small study over a short time span.**

Each search engine's response time was divided by that of the fastest search engine (Google). **The numbers you see are therefore not response times in seconds, but response times relative to that of Google.**



Surprises here are MSN Search claiming second spot and AlltheWeb being on average 6 times slower than Google in the searches I did. But even that is **FAST!** In the end I think these figures mean very little. These days the level of competition leaves no room for a slower engine – and the ones in this test all still exist because they are all very fast.



## 1.20 More Search Engine Statistics

### Statistics from [Searchengineshowdown.com](http://www.searchengineshowdown.com):

#### **Relative Size Showdown:**

Updated August 14, 2001.

<http://www.searchengineshowdown.com/stats/size.shtml>

#### **Total Size Estimate:**

Updated August 14, 2001.

<http://www.searchengineshowdown.com/stats/sizeest.shtml>

#### **Change Over Time:**

Updated August 14, 2001.

<http://www.searchengineshowdown.com/stats/change.shtml>

#### **Database Overlap:**

Updated Feb. 21, 2000.

<http://www.searchengineshowdown.com/stats/overlap.shtml>

#### **Unique Hits Report:**

Updated March 9, 2000. (Data from Feb. 21, 2000)

<http://www.searchengineshowdown.com/stats/unique.shtml>

#### **Dead Links Report:**

Updated Feb. 21, 2000.

<http://www.searchengineshowdown.com/stats/dead.shtml>

**Statistics from [Searchenginewatch.com](http://searchenginewatch.com):****Search Engines Size:**

Graphical look at how large each search engine is, with trends over time. Links to information on whether size matters.

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

**Directory Sizes:**

Directories are usually human-compiled web guides that list sites by category. This compares prominent directories.

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

**Searches Per Day:**

Shows how many searches per day are performed on some search engine

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

**Search Engine Index:**

Interesting stats about search engines, at a glance.

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

**NPD Search and Portal Site Study:**

This quarterly survey measures satisfaction with search engines.

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

**GVU Survey:**

This twice-per-year survey shows how people locate web sites.

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

**Search Engine Reviews Chart:**

At-a-glance guide to search engines with the best reviews.

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

# 1.21 Search Engine Relationships

Search Engine	Receives results from	Sends results to
Google	<b>OWN DATABASE</b> Directory listings from DMOZ.	<b>Main results</b> to Yahoo, Netscape, iWon and AOL Search (and many smaller search engines). <b>Paid listings</b> (from AdWords) to Teoma, Netscape, Ask Jeeves and AOL Search.
Yahoo	<b>OWN DATABASE</b> Main results from Google. Paid listings from Overture.	None
AltaVista	<b>OWN DATABASE</b> Directory listings from LookSmart. Paid listings from Overture.	None
DMOZ	<b>OWN DATABASE</b>	<b>Main Results</b> to Lycos. <b>Directory listings</b> to Google & HotBot <b>Some results</b> to AlltheWeb & Teoma.
Overture	<b>OWN DATABASE</b> Some results from Inktomi.	<b>Main results</b> to Go.com <b>Paid listings</b> to Yahoo, MSN Search, Lycos, AltaVista, InfoSpace.
AlltheWeb	<b>OWN DATABASE</b>	None
Excite	<b>Meta search.</b> Receives results from Google, LookSmart, Inktomi, Ask Jeeves, About, Overture, FindWhat, Fast.	None

CONTINUED ON THE NEXT PAGE

Search Engine	Receives results from	Sends results to
<a href="#">Lycos</a>	<b>OWN SMALL DATABASE ("LYCOS NETWORK")</b> Main results from Fast. Paid listings from Overture	None
<a href="#">Teoma</a>	<b>OWN DATABASE</b> Paid listings from Google AdWords Some results from DMOZ	Some results to Ask Jeeves
<a href="#">LookSmart</a>	<b>OWN DATABASE</b> Some results from Inktomi	Main results to MSN Search Some results to AltaVista

## What To Do With This Info

Use it to **focus** your SEO efforts. For example: Being listed at [Google](#) & [DMOZ](#) is very important, because they both "feed" a couple of other major engines (and many smaller ones). Once your site is in Google & DMOZ, it will eventually start popping up all over the place.

## Get Free Updates

I will report changes to these relationships in my *EnginePaper Newsletter*. Subscribe (free) with a blank e-mail to [send-ep-subscribe@topica.com](mailto:send-ep-subscribe@topica.com).

## 1.22 Search Engine News

In SEY 2002, I reported page after page of news – all outdated by the time the book launched.

Last year we also introduced the **EnginePaper Newsletter** to keep you informed of important search engine news throughout the year. That newsletter has taken off better than expected and proved a far more effective way of reporting news.

### Subscribe (Free)



To subscribe, simply send a blank e-mail to [send-ep-subscribe@topica.com](mailto:send-ep-subscribe@topica.com)

For those who prefer news directly from the search engines themselves, here are...

### The News Pages Of Some Of The Top Search Engines:

Google Press Room: <http://www.google.com/press/index.html>

AltaVista Press Room: [http://www.altavista.com/sites/about/press\\_welcome](http://www.altavista.com/sites/about/press_welcome)

Yahoo! Press Releases: <http://docs.yahoo.com/info/pr/releases.html>



DMOZ Press (2002): [http://dmoz.org/Computers/Internet/Searching/Directories/Open\\_Directory\\_Project/Press/2002/](http://dmoz.org/Computers/Internet/Searching/Directories/Open_Directory_Project/Press/2002/)

Excite Media Relations: <http://corp.excite.com/News/>

Lycos Press Room: <http://www.terralycos.com/press/index.html>

Fast Press Releases: <http://www.fastsearch.com/index.php?d=press>

LookSmart Press Room: <http://aboutus.looksmart.com/about.jhtml> (Click "Press Room")

## 1.23 Telephone Directories

Most online phone number directories are derived from one of two major databases: “**infoUSA**”, formerly known as American Business Information Inc., and from **Acxiom**. So to keep it short and to the point, I’ll give you a major online directory for each database:



**(Uses Acxiom)**

The SuperPages homepage offers a yellow pages search (businesses). For a white pages search, select the “People Search” link from the menu. SuperPages allows you to search by US state or the entire country. Notably, the Acxiom database returned slightly more results in a test search than infoUSA.

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

The logo for Switchboard.com features the text "Switchboard.com" in a blue, sans-serif font. The text is set against a yellow rectangular background. The letter 'i' in "Switchboard" has a red dot above it, and the letter 'd' in ".com" has a red dot to its right.

**(Uses infoUSA)**

A slightly cleaner looking homepage that offers a choice of white or yellow pages right from the start.

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



## 1.24 Meta Searching

### What Is A Meta Search Engine?

A meta search engine looks a lot like a regular search engine when you arrive at the main search page.

But there is a **BIG** difference **below the surface**.

A meta search engine typically **does not have its own database** of indexed web sites. It takes your search query, runs off to a number of “real” search engines and queries those search engines’ databases. The results returned to the user are therefore a collection of results from different search engines.

That could be great – more search results from more sources – great for finding obscure information, right?

Wrong.

### The problem with meta search engines

They represent a commendable effort, but very seldom does a search on a meta engine provide better results.

Apart from major limitations like the absence of advanced search and the real possibility of timeouts, they often retrieve only the top 10, top 50 or top 100 results from each search engine. **You end up with fewer results** than you would if you searched directly at one of the search engines it queries. Phrase and Boolean searching are rarely processed correctly, because the search engines being queried implement it differently.

That said, meta search engines can be useful. The revamped **HotBot** search engine, although not a meta search engine in the strictest sense, is a great tool for power searching and for comparing databases.

## Some Of The More Popular Meta Search Engines



Dogpile searches an impressive list of sources:

LookSmart, Overture, Thunderstone, Yahoo, Open Directory, About.com, Lycos' Top 5%, Direct Hit, and AltaVista. It offers other searches for Usenet, FTP, News Wires, Business News, Stock Quotes, Weather, Yellow Pages, White Pages, and maps. The wide reach and ability to customize results makes Dogpile one of the most popular meta search engines.

<http://www.dogpile.com>



*"Mamma.com is the **largest independently owned metasearch engine on the Internet.**  
Mamma.com:*

*is a Nielsen/NetRatings Top 10 Search Engine.*

*is a Media Metrix 500 Company.*

*reaches over 7,000,000 unique users per month.*

*returns results for over 30,000,000 searches per month.*

*provides its search functionality to over 13,000 third party websites.*

*further increases its reach with over 100 major strategic alliances."*

Mamma also has its own "Mamma Collection" – a quality, human reviewed collection of web sites. Once your site is added to this collection, it receives a ranking boost in normal search results at Mamma. Submitting your site to the Mamma collection is **not** free. You have a choice of "Velocity Submit" and "Standard Submit"

### **Velocity Submit**

Your site is reviewed within 2 business days. The last time we checked, the price was \$59.99 with a \$19.99 annual subscription.

### Standard Submit

Your site is reviewed within 8 weeks. The price is \$29.99 – again with a \$19.99 annual subscription.

NOTE: paying to have your site reviewed does not guarantee that it will be included in the Mamma Collection – only that it will be considered for inclusion. If you have a quality site with no dead links or images, your chances of getting in are good.

<http://www.mamma.com>

--- S I D E B A R ---

### Confused by the terminology?

Learn some search engine lingo...

Most of the search engine terms used in this book are explained in the **Search Engine Dictionary section**.

You can also download the dictionary as a separate, **free** e-book. Visit [www.searchenginedictionary.com](http://www.searchenginedictionary.com) for details.



# 1.25 The Future of Search

Contributed by I-Search moderator Detlev Johnson



**Only in the full version:**



**The Future Of Search (by Detlev Johnson),**  
**Who Will Be The Next Google (by Jill Whalen)**



-----

**Not in the free version: p151 to p159**

-----



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## Section 2: Resources For Search Engine Users



# Resources For Search Engine Users



## 2.1 Internet Search Strategies

The Internet is without any doubt the largest source of information on just about any topic you can think of. The problem is that you can easily waste many hours sifting through irrelevant sites.

This little tutorial is about cutting **down your search time** by searching **smarter**.

There are thousands of search engines and directories on the Net, **so the first thing you have to do is decide which one to use...** No, the answer is not always “Google”.

You may end up using a directory instead – especially if you are researching a fairly broad topic.

### When And How To Use A Directory

Directories like **DMOZ** (<http://dmoz.org>) are usually human-created indexes of web sites neatly organized into topical categories. Because they are created by hand, they are usually **much smaller** than search engines. You might be thinking that search engine are therefore far better at finding relevant info, but...



Small can be good. Let's say we're looking for something very general – educational PC games.

There must be **thousands** of sites mentioning “educational PC games”. Sifting through all that will take hours.

But when you use a directory, someone else has already done the sifting. That’s what makes directories useful. There is almost always some kind of **editorial selection process** where sites are measured against a standard set by the directory. At one stage, the Yahoo editors were rumored to reject as many as 9 out of 10 site submissions.

**Because of this, directories will have only a few sites per category, but they are very likely the **best** sites on the topic.**

Let’s see if we can find educational PC games. I think I’ll head to



EXAMPLE: “Educational PC games”

When you use the Yahoo search feature, the results you see are from Google.

That’s not what we want, so we instead go to their category listings looking for something like “Computers”, “Software” or maybe even “Shopping”.

Yes, there it is. “Software”...



Under the main category, “Computers & Internet”, there’s a sub-category called “Software”. Now it’s just a matter of drilling down.

When you click “Software” it shows its sub-categories. Under “Software” there is “Education”, under that there’s “Teaching & Learning Aids” and under that there’s “Games”.

In this case the “Games” sub-directory is as far down as you can go. It shows only sites listed in that category – no further sub-categories.

Here are the two sites listed there:

#### **Alphabetical**

- [LearningWare, Inc.](#) - develops tools for teachers or trainers to create gameshows and quizzes for fun learning.
- [Solete Software](#) - free downloads of educational software and computer games.



## About Using Search Engines

This is where it gets more complicated, but **stay with me**. I'll make you a **super searcher** if you do... 😊

How much time do you spend searching during an average day? I probably use search engines a bit more than most people. I discovered that I spend about 2 hours a day finding information via search engines – correction... **looking** for information. Actually finding it is another thing altogether.

I decided to read up on search techniques and with some nifty new tricks chopped my search time (almost) in half. Unfortunately being good at searching costs me more time than it saves. Friends now phone me up – “André, hi! I need something on the diet of the Malaysian hunting spider for Billy’s science project. Any ideas?” Uh, yeah Bob, buy my book.

Seriously though, **here’s what I learned about searching the web...**

**The first and most important thing in web searching is to use the *RIGHT* search engine.** Contrary to popular belief, they don’t all index the entire web – even though they have billions of documents in their databases.

Ok, we know that when looking for something fairly broad, directories are great. Now, here’s...

## When To Use Which Search Engine

For **broad, general searches**, try <http://www.google.com> or <http://www.teoma.com>

For **quality academic resources**, try <http://www.ii.org> or <http://www.academicinfo.net>

For **shopping**, try <http://www.yahoo.com> or <http://www.overture.com>

For **natural language questions**, try <http://www.ask.com>

For **expert links**, try <http://www.about.com> or <http://vlib.org>

For **news**, try <http://news.google.com>

For **government info** (U.S.), try <http://www.firstgov.gov>

For **images**, try <http://images.google.com> or <http://images.altavista.com> or <http://ditto.com>

For **multimedia**, try <http://www.alltheweb.com/advanced>

For **kids' sites**, try <http://www.yahooligans.com>

For **queries containing stop words**, e.g. "To be or not to be", try <http://altavista.com>

## Boolean Searching

Most search engines allow you to use Boolean operators like AND, OR etc.

Imagine you're ordering a ham sandwich. You want cheese but no tomato or unions. To a search engine you'd say:

*“ham sandwich” AND cheese AND NOT tomato AND NOT union*

**No, it's not that easy.**

**It would be if all search engines used the same Boolean operators, but they don't. Here's what they do use:**

Search Engine	Boolean Operators	Other Characters
Google	<b>AND</b> (default) <b>OR</b> <b>+</b> (to include <b>stop</b> words) <b>-</b> (to exclude words)	<b>“ ”</b> (quotes for phrase searches) <b>*</b> (wildcard to replace words in a phrase) <b>Other fields:</b> <b>allintitle:</b> , <b>allinurl:</b> , <b>link:</b> and <b>site:</b>
Yahoo	<b>AND</b> (default) <b>OR</b> <b>+</b> (to include words) <b>-</b> (to exclude words)	<b>“ ”</b> (quotes for phrase searches) <b>*</b> (wildcard to replace words in a phrase) <b>Other fields:</b> <b>t:</b> (title) and <b>u:</b> (URL)
AlltheWeb	<b>AND</b> (default) <b>+</b> (to include words) <b>-</b> (to exclude words)	<b>“ ”</b> (quotes for phrase searches)
AltaVista	<b>AND</b> (default) <b>+</b> (to include words) <b>-</b> (to exclude words)	<b>“ ”</b> (quotes for phrase searches) <b>*</b> (wildcard to replace words in a phrase) <b>Other fields:</b> <b>domain:</b> , <b>host:</b> , <b>image:</b> , <b>title:</b> , <b>url:</b> , <b>link:</b> , <b>like:</b> , <b>anchor:</b> and <b>applet:</b>
Teoma	<b>AND</b> (default) <b>+</b> (to include <b>stop</b> words) <b>-</b> (to exclude words)	<b>“ ”</b> (quotes for phrase searches)



## 2.2 More Tutorials On Internet Searching

□

**Only in the full version:**□

□

□

**More Tutorials On Internet Searching, Articles On Internet Searching, □  
General resources For Search Engine Users □**

□

-----□

**Not in the free version: p169 to p174**□

-----□

□

□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**



## Section 3: Search Engine Optimization (SEO)



# Search Engine Optimization

## 3.1 Overview Of The SEO Industry

Search engine optimization continues to be the most cost effective of online marketing techniques. **But there is a catch:** The search engine optimization industry has become saturated.

As competition between SEO providers increase, achieving decent rankings will become more and more difficult. For ordinary folks like us, competing for top keywords like “business” or “e-commerce” is a complete waste of time.

**Let’s get to the bottom line right away:** SEO has become a **specialized business**.

Fortunately for us, the web is still a fairly level playing field – **and armed with this book, you have a fighting chance.**

EXAMPLE: David & Goliath

Here’s a little (true) story of how we at [Pandecta Magazine](#) outperformed a much larger company on some tough keywords...

Our “[Electronic Light](#)” [web site](#) was built as an experiment. I wanted to see for myself if there is really any money in affiliate programs. So I signed Pandecta up as an affiliate for distributors of all kinds of lamps and built a lamp-shopping site.

### The next step: Pulling traffic off the search engines.

**Only problem:** The top spots on Google for all the keywords I wanted to target were taken – most by the **same**, (very) large lamp distributor.

I got top 10 placement for about 80% of my top keywords – but no number ones.

I knew that the pages were as optimized as I could make them without cheating, so I shifted my focus to the site's **PageRank**. **A decent link building campaign saw Electronic Light's PageRank increase from 1 to 5** (as reported by the Google toolbar) – and sure enough, we moved into the number 1 slot on **3** of our biggest keywords. Woohaa!

**PS:** If you're interested, I share what we learn from the [Electronic Light affiliate site](#) in my *Electronic Light* newsletter. You can subscribe for free by sending a blank email to [electronic\\_light-subscribe@topica.com](mailto:electronic_light-subscribe@topica.com).

### To further illustrate this point:

A couple of days ago a spoke to a guy who operates a gambling site. He wanted to know why search engines are so **bad** at listing bigger companies at the top. My response was “That's SEO in action”. Some of the little guys know how!

My aim with this section is to make **you** one of those little guys/gals that **know how** and consistently **beat the bigger players**.

## 3.2 Overview of Web Marketing Techniques

There are many ways to get your web site noticed. Some techniques work very well, some don't, and some are simply a huge waste of your time.

**Here's a rundown of the most popular / most hyped Internet marketing techniques, each with an explanation.**



### 3.2 OVERVIEW OF WEB MARKETING TECHNIQUES

#### 3.2.1 Web Marketing Techniques: Search Engines

This is **by far** the most effective (and most cost effective) way to attract visitors to your web site.

**By very far.**

Many people feel that marketing your site on the search engines is not all it's cranked up to be. These are usually people who tried their hand at it and had limited success. Research has shown that **more than 70% of your first-time visitors will have found you on one of the major search engines**, so there is no real argument against SEO.

Your site **has** to be found on the search engines.

There are many factors impacting this “70%” figure, so it'll vary from site to site. In my own experience, small business sites that are properly optimized for the search engines can see that figure climb to as much as **90%**.

On a scale of 1 to 10, SEO scores a perfect 10.

--- S I D E B A R ---

### SEY 2004: Your 25% discount

As an owner of SEY 2003, you qualify to receive SEY 2004 at 25% below the regular price. **BUT:** I need your permission to e-mail you the link to the special order page. All you have to do is subscribe to the SEY updates list. I promise to send you only 1 e-mail a year: when the new SEY is ready. Subscribe with a blank e-mail to [sey-subscribe@topica.com](mailto:sey-subscribe@topica.com).



## 3.2 OVERVIEW OF WEB MARKETING TECHNIQUES

### 3.2.2 Web Marketing Techniques: **Link Building**

Links from other web sites to yours will probably **not** send a lot of new traffic your way. It depends on the link itself.

If the link is just a few words or a small graphic, don't expect much. If the link is **preceded by a review / introduction of your service**, the clickthrough rate rockets, but still seems small next to search engine traffic.

Inbound links are becoming an important factor in search engine optimization (**more later**) and because of this it does matter. Getting people to link to your site is doable, but not all link building strategies are equally effective. **Some could even hurt your site.**

**We'll take a more detailed look at link building further down. [Click here to jump to that section now.](#)**



## 3.2 OVERVIEW OF WEB MARKETING TECHNIQUES

### 3.2.3 Web Marketing Techniques: **Word Of Mouth**

Word of mouth is fairly **difficult to create, but extremely powerful**. It has more to do with product development than with marketing. A great product at a great price **earns** word of mouth.

**If this is true offline, it is especially true online.** Discussion forums, newsrooms, chat rooms, e-mail and newsletters all combine to form a medium that spreads "the word" like nothing before. Easy, fast and effective information exchange is after all what the Internet is all about.

Of course, your customers will share **negative** experiences just as effectively.

**While we are on the topic, here's something else to keep in mind:**

Techniques like spam marketing give unethical Internet businesses high (if ineffective) visibility. The perception created is that "the web is full of scammers". Consumers are generally more careful when shopping online, so **any hint of deception will loose sales**. **Soft selling** works really well for me. I don't use hyped phrases like "Get it now!". Simply talking to the customer as if in an e-mail gets results. Keep in mind that this will not necessarily be as effective for you unless you're also targeting web savvy entrepreneurs.



## 3.2 OVERVIEW OF WEB MARKETING TECHNIQUES

### 3.2.4 Web Marketing Techniques: **Online Advertising**

You have many options when it comes to buying online advertising. You're no longer limited to standard, horizontal banners and many offers may seem tempting.

But be warned:

**Effective online advertising is extremely difficult.**

**Less than 0.4% of people who see your ad will click on it.**

**That's if you have a very appealing ad.**

**Most advertisers struggle to reach a click-through rate of 0.1%.**

In the early, wild wild web days, advertisements worked. Some banners commanded clickthrough rates as high as 10%. But web surfers quickly became desensitized to advertising, learning that the sites behind the ads often do not deliver what the ad promises. This phenomenon is now so generally accepted that a new word, “banner blindness”, was created to describe it.

That said, the online advertising industry is slowly getting back on its feet after the dotcom boom left it in tatters.

If you decide to try online advertising, invest in a system that can track results precisely. Measure the ROI and branding value of each ad separately.





## 3.2 OVERVIEW OF WEB MARKETING TECHNIQUES

### 3.2.5 Web Marketing Techniques: **Offline Advertising**

You already have your Internet address (URL) on your letter head & business card, right?

Add it to everything.

Every promotional item you send out. Every advertisement. Even work it into your radio ads. Print your URL on stickers for use on the company car and on free samples of your products.

**Your URL should be just as easy to find as your company's telephone number.**

Advertisements on television, radio, newspapers and magazines can be effective, but an offline ad reaches a lot of people who have no chance of visiting your site - either because they don't have access to the web or don't know how.

This is of course changing as the Internet continues to worm its way into everyday life.

## 3.3 SEO Facts

It is quite common to find to SEO “experts” **contradicting each other.**

There are almost as many opinions as there are experts. Below are what I consider ground rules - SEO principles that are generally accepted as fact and rarely questioned.



### 3.3 SEO FACTS

#### 3.3.1 SEO Facts: **Content Is (Still) King**

Way back in 1997, one of the original search engine gurus, Jim Rhodes, said “Content is King”. Well done Jim. You were right then and you are even more right now.

**Good content creates word of mouth. It sells itself.**

One year later, in September 1998, **Google** and its revolutionary **PageRank** system took Jim’s idea to the next level. PageRank effectively rewards good content by factoring incoming links into its algorithm. All the major search engines now measure link popularity and use it to improve the accuracy of their results.

The rule is that good sites will get more visitors. Always. **Concentrate on building true value first. It’s the hardest but the most important principle in SEO.**



### 3.3 SEO FACTS

#### 3.3.2 SEO Facts: **Keyword Targeting**

How can you **double** your site traffic without doubling your effort?

Yes, proper keyword targeting.

Which keywords will your customers enter into the search box when looking for your product? If you've been guessing up to now, you no longer have to.

**Here's a strategy that works well for me:**

## **STEP 1**

Type the root form of your best keyword into "**GoodKeywords**", a little application you can download for free. It then shows you how many people use that keyword and it also shows **99 variations** of the word listed from most used to least used. Study that list closely for variations or synonyms you didn't think of. Use the GoodKeywords list to make your own list of possible keywords to target.

## STEP 2

**Next, take your list to Google.** Type in the words you want to target and look at a couple of the sites listed in the top 10. Can you beat them? Remember to look at their **PageRank** too. Scrap from your list the ones for which you can't compete. If you can't compete on any of your words, go back to GoodKeywords and aim lower.

## STEP 3

Take a close look at the site listed in the number 1 slot for each of your keywords. Remember that **keywords in links pointing to that site also count**, so look at those too by doing a search for link:www.your-competitor's-domain-here.com on Google.

All that's left now is to "out-optimize" that site. Yes, not that easy, but the rest of this section of SEY will give you a fighting chance.

As a general rule you should not target bigger, more competitive keywords. If you can rank well for them, then go for it, but usually they are a waste of time.

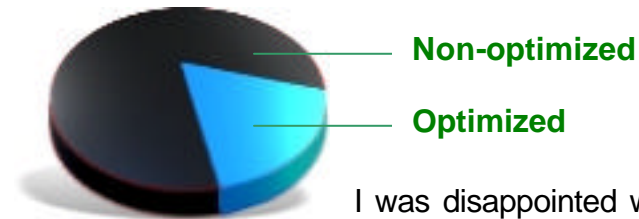
**You should focus on efforts on keywords that will bring top 10 rankings.**

I'm currently experimenting with a more blanket strategy (versus a targeted one).

**Here's what I learned from the Pandecta site:**

I noticed that my best (very competitive) keyword delivers 4% of my total search engine traffic. The second best 2.5% and so on.

In total my top 20 keywords are responsible for almost **19%** of my search engine traffic.



I was disappointed when I saw that. It means that all my efforts to optimize for those 20 keywords only bring **less than one fifth** of my search engine traffic. The other 81% type in words I didn't think of or combinations of words or they include keywords in phrases.

So right now I'm experimenting with ways to include **more variations** of keywords. Although most experts will tell you to focus each page narrowly on one keyword, I think it might pay off to optimize for groups of related keywords – especially on **Google**.

Anyway, I'm still playing with that. I'll report on my findings in my *Electronic Light* newsletter. (You can subscribe for Electronic Light by sending a blank e-mail to [electronic\\_light-subscribe@topica.com](mailto:electronic_light-subscribe@topica.com))

□

**Only in the full version:**□

□

□

**Invisible Text, Resubmission, Search Engines That Matter, Domain Names, Cross-Linking, Dedicated IP Addresses, Robots.txt & The Robots Meta Tag, Link Building** □

□

-----□

**Not in the free version: p189 to p205**□

-----□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## 3.4 SEO “Maybes”

Here are some more do’s and don’ts of SEO. I call them “maybes” because they are somewhat more controversial than techniques discussed above as “SEO Facts”.

**You’ll find lots of conflicting advice about these on the web. In the 5 years I’ve been playing the search engine game, this is what has worked for me:**



3.4 SEO “MAYBES”

3.4.1 SEO “Maybes”: **Getting Doorway Pages Right**

**Doorway pages are keyword focused pages that link to your main web site.**

They are designed to score well on search engines, and then act as a bridge between traffic from the engines and your main site / order page.

Doorway pages still work, but search engines have, over the last year or two, changed their attitude towards doorway pages and are devising means of weeding out doorway pages from their indexes.

**Why?**

Because dishonest webmasters create basic, template pages and fill them with **keyword gibberish**, redirecting visitors from them to the main site. Even worse, you can now buy software that automatically churns out doorway pages.

This technique used to work, but search engines (and notably Google) have stated that they're aware of the problem and that **they will penalize sites that use automatically generated doorway pages.**

The solution?

**Create doorway pages that search engines love.** Here's how...

These two techniques have worked well for me. Both require some effort, but the rewards are long-term.

## Search Engine Friendly Doorway Pages: Technique 1

Write an article for each keyword.

It does not have to be very long – about 200 words work well. The thing is to make those 200 words count in 2 ways:

1. You **HAVE** to deliver unique value. It's not that hard. Share some knowledge. No-one expects you to reveal trade secrets for free, but **if you don't give your**



**visitor something on page one, she's gone.** She'll arrive at your site with her trigger finger on the back button. You have to convince her to stay.

2. Those 200 words have to be **keyword rich** to impress the search engines. Be careful though. There is such a thing as "keyword stuffing". Excessive use of keywords will get your site penalized. Besides, you don't want your visitor to read. "Welcome to Acme Lawnmowers, the lawnmower shop. We sell lawnmowers and also lawnmowers." If it does **sound** right, it isn't.

Next, create links between your articles, so that your collection of doorway pages becomes like an article archive. No search engine will ever exclude valuable, on-topic content.

**The downside to this is that you no longer have just one path from your doorway page to your order page.** Web surfers get distracted easily, so make sure the button that leads to the main site / order page is more prominent than the links to your other doorway pages.

## Search Engine Friendly Doorway Pages: Technique 2

**Optimize your product pages themselves.** This one works well for me because I have a small number of products, so I can create and optimize product pages **by hand**.

Again, don't overdo it. Compare these two sales pitches for a tiffany lamp:

- A: "Tiffany Lamp: Tiffany-style lamps. Buy this Venetian Tiffany Lamp from "Tiffany-Lamps-R-Us". This tiffany lamp..."
- B: "Tiffany Lamp #123: The Venetian Tiffany Lamp. This unique tiffany lamp will transform any room..."

A is clearly overdoing it. B is also pushing it, but notice how much easier it reads.

Getting doorway pages right is critical. Here's a book that has, in my opinion, the most eye-opening and comprehensive discussion on doorway pages.



Ken Evoy's "**Make Your Site Sell 2002**" is probably the **most complete guide to getting entry pages right**.

He calls them "Keyword Focused Content Pages" (KFCP). Yes, really. He's Canadian you see. ☺

It's the same thing though – and Ken definitely **knows** his stuff.

By the way, this book covers Keyword Focused Content Pages **and everything else** a Netpreneur could possibly want to know. If you haven't read it, you should. No other complete guide to e-commerce comes close. It sells at about \$30 if I remember correctly.

<http://www.sitesell.com/book6.html>

NOTE: Pandecta is an affiliate for SiteSell.com. If you buy "Make Your Site Sell", we get a cut for referring you. I do however really believe in this book. It gave me a massive head-start. I signed up as an affiliate because this is a product I feel comfortable promoting. Try it for yourself.



**Only in the full version:**



**Updated Thinking On Meta Tags, Submission Software, Cloaking**



**Not in the free version: p211 to p223**



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## 3.5 Getting Listed at DMOZ (ODP)

### Why DMOZ Matters So Much

*(If you already know why, click here to skip to the how-to section)*

Getting your pages listed at **DMOZ** (a.k.a. Open Directory Project (ODP)) is **extremely** important.

**Here's why:**

EXAMPLE: DMOZ & PageRank

One of my sites had a **PageRank** (PR) of 4 (as reported by the **Google** Toolbar). At that point, most of its inbound links came from one of my other sites. I submitted the site to DMOZ.

I heard that DMOZ will sometimes allow a page to be listed in 2 different categories, provided that it is appropriate for both. I tried it and it worked. The site's homepage was accepted at both categories I submitted to. **The PageRank for the homepage jumped to 6.** I checked, there were no new inbound links except the 2 from DMOZ.

I should mention that some search engine experts believe a listing at DMOZ is **not** that important. From my experience above, I very much disagree.

Apart from the PageRank boost, a DMOZ listing might have other advantages.

I say *might* because this is purely guessing based on what I would do if I were Google, but **it would make sense for Google to check for keywords in the**

- **DMOZ title,**
- **DMOZ description and**
- **DMOZ category.**

Rumor has it that Google gives a **special PageRank boost** to sites listed at **DMOZ** and **Yahoo**. The thing is that a Yahoo listing will cost you \$299 per year. It's debatable whether that's worth it.

**Submitting to DMOZ is free, so it's a no-brainer.**



## 3.5 GETTING LISTED AT DMOZ (ODP)

3.5.1 DMOZ Submission Tips: **Before You Submit****What are DMOZ editors looking for above all else?**

**Unique, valuable content – and lots of it.** If your site has little or none, create some. Write a number of informative how-to articles, safety tips for your industry, list some related resources etc. Use your experience in your field to make your site unique & valuable.

It is more doable than you think.

- Are you selling household cleaners? Tell me when and where I should use which type.
- Are you selling baby products? Give me some tips on making baby sleep. (PLEASE!)
- Are you selling a book? Put a sample chapter right there on the site.
- Are you selling furniture? Share some of your ideas on interior decorating.

Everyone has experience locked away in their brains. Experience other people would pay for. **Get some of that on paper** and give it away from your site. Without it getting into DMOZ will be much harder.



**Only in the full version:**



**Finding The Right Category, About Regional Sites, About Adult Sites, About Affiliate Sites, Your Submission, Getting Pay-Per-Click Marketing Right**



-----

**Not in the free version: p228 to p237**

-----



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## 3.7 Why Can't I Get My Site Listed?

### Frustrated?

Make sure your site is not guilty of any of these:



### 3.7 WHY CAN'T I GET MY SITE LISTED?

#### 3.7.1 Mistakes: **Browser Requirements**

I've lost count, but I must've built 50 sites over the last 5 years. **The ones that get traffic and make money are ALWAYS the simplest, text-rich ones.** The only exception is for a large site where we used Java to build a monster of a shopping cart system. It actually worked! (Well done Marius and Richard!).

#### **Back to search engines:**

Search engines use spiders to index pages. These little machines look at text. They love text. Most of them don't love (disregard) the fancy stuff – Java, Flash, DHTML etc. To my knowledge, only **Google**, **WiseNut** & **Inktomi** can spider dynamic content.

**Effective web design means cutting back on the gimmicks.** Your site should have the minimum. Just enough to make it user-friendly.



If your site really **(really really)** does require gimmicks to work, consider creating text-rich, gimmickless landing pages to submit to the search engines. Note that there is a right and a wrong way to build landing pages.

More about the differences [here](#).

I should mention that search engines, notably [Google](#), are improving their ability to spider dynamic content.

I should also mention – just in case you haven't thought of it – that sites that requires passwords cannot be spidered. I told you this is a *complete* search engine book ;-)



### 3.7 WHY CAN'T I GET MY SITE LISTED?

#### 3.7.2 Mistakes: **Frames**

**Frames, when used correctly, are fantastic, but only if you're building an intranet or a site that don't need / want search engine traffic.**

You downloaded this book though, so **you want traffic** – and lots of it. Don't use frames.

Most search engines cannot index framed pages. They see only the frameset page, not the (keyword-rich) source pages of individual frames.

There is a way to get search engines to index your framed site correctly, but I strongly advise that you **avoid frames altogether**. As great as they are, they're not worth the mountain of additional time and effort.

#### **If you must, here's how:**

Inside your `<noframes>` tag, write a complete, keyword-rich description of your site. Feed your spider. Also drop some links in there so it can hop through the rest of the site.

There are potential problems (and fixes) to this, but we're moving into technical web design territory here. If you're interested, I recommend the frames tutorial at Webreference.com: <http://www.webreference.com/dev/frames/>



## 3.7 WHY CAN'T I GET MY SITE LISTED?

3.7.3 Mistakes: **Automatic Redirects**

There are different ways to automatically redirect visitors from the page they land on to your main page. It is however a no-no that'll get your site penalized or dropped.

**If you have automatic redirects, remove them.**

Your site won't get anywhere as long as you use them.

--- S I D E B A R ---

### **Confused by the terminology?**

**Learn some search engine lingo...**

Most of the search engine terms used in this book are explained in the **Search Engine Dictionary section**.

You can also download the dictionary as a separate, **free** e-book. Visit [www.searchenginedictionary.com](http://www.searchenginedictionary.com) for details.



### 3.7 WHY CAN'T I GET MY SITE LISTED?

#### 3.7.4 Mistakes: [Google Minimum PageRank](#)

This isn't really a mistake but a shortcoming of many sites – and it's one that can cause extreme frustration.

[Google](#) relies heavily on [PageRank](#) to rank sites. **According to the Google site, they won't index sites that have no inbound links** because the PageRank for those sites “can not be calculated in a meaningful way”.

To check your inbound links, do a search on Google for `link:www.your-domain-here.com`

If you know of sites that link to you that don't show up here, **submit them** to Google and wait for the next [Google Dance](#). If you haven't yet, submit your site to [DMOZ](#). A link from there to your site is usually enough to get you over this hurdle.

Consider paying the \$299 annual fee to get your site listed at [Yahoo](#).

Also look at the [discussion of link building](#) earlier in this section.

By the way, [PPC marketing](#) is a fast a reliable way to get traffic to your site while you're still building your site's link popularity.



## 3.7 WHY CAN'T I GET MY SITE LISTED?

### 3.7.5 Mistakes: **Free Space**

Free (banner-supported) hosting is a bargain, but **only** if you run a hobby site.

If you're trying to sell something, **free hosting looks amateurish** and it can be a disadvantage in SEO. Sites on free servers often share the same IP address. It is possible that your site's IP is blocked because someone sharing your IP misbehaved.

Also see the [discussion of IP sharing](#) above.



## 3.7 WHY CAN'T I GET MY SITE LISTED?

3.7.6 Mistakes: **Blocking Spiders**

You may accidentally be telling the search engine spiders to **NOT** index your site. If you have a “robots.txt” file in your root folder, check it.

If it says

```
User-agent: *  
Disallow: /
```

then that is why you can't get listed. You're telling all search engine spiders (\*) to ignore everything on your site (/).

Fortunately this one is easy to fix. Refer to the [discussion of the robots.txt file](#) earlier in this section.

## 3.8 If You Can't Beat'em, Delete'em

During the course of this book I'll try to convince you of the value of honest SEO, but what do you do when you discover a site listed above yours that does not play within the rules?

That's right. Delete'em

Don't feel too bad about it. They don't deserve to be highly ranked.

Search engines fight a never-ending battle against spam (Spam, in the context of search engines, is sometimes also referred to as "spamdexing"). Most search engines have a wall of spam-catching measures, but these cannot catch every "SEO trick". To the contrary, spamdexing is fairly easy.

Rather than show you how, this section shows you how to **report** spammers. **Once the search engine knows about him/her, it's a matter of time before their sites are deleted from the index** (and your site moves up a notch).

**First, here's what the search engines usually consider spam techniques:**

Any technique that aims to deceive in order to gain search engine placement, specifically:

**Cloaking**, discussed in more detail above, offers a way of delivering an optimized page to search engines spiders and your "real" page to human visitors. All search engines discourage cloaking. Cloaked sites run the risk of receiving a life ban. One

way to detect cloaked pages is to compare the actual page with **Google's** cached version.

- ❑ **Doorway pages**, also discussed earlier, is considered spam when it consists of keyword gibberish that automatically redirects to another page. Automatic redirection can be detected by comparing the URL shown in the search results to the actual URL.
- ❑ **The bait & switch technique** involves creating 2 pages – one filled with keywords and the other with the real content you want your visitors to see. The second page is uploaded into the place of the first as soon as the first is indexed. This is not very effective though. It's extremely time-consuming and almost impossible to predict when the spiders will revisit. Spammers using this technique shoot themselves in the foot.
- ❑ **Cybersquatting** refers to the practice of registering domains that resemble popular domains. Domains like [www.altavidta.com](http://www.altavidta.com), [www.gogle.com](http://www.gogle.com) etc. are designed to get traffic through typos.
- ❑ **Invisible or hidden text** is text of the same color as the background
- ❑ **Overused keywords** and irrelevant keywords in the title, **meta tags** and body.
- ❑ Submitting sites to **inappropriate categories** at directories like **DMOZ**.

If you find a site guilty of any of the above, report them to the search engine where you found the offending site. Here's how:

**Google:** Fill out the form at <http://www.google.com/contact/spamreport.html> or email to [spamreport@google.com](mailto:spamreport@google.com) .



**AltaVista:** Fill out the form at <http://help.altavista.com/contact/search> and select “Spam Reporting” in the subject field.

**AlltheWeb:** Send an e-mail to [spam@fastsearch.com](mailto:spam@fastsearch.com) with the subject “Spam report”.

**Overture:** If you find a site not conforming to Overture’s terms of use (<http://www.overture.com/d/USm/about/company/terms.jhtml>), you can report it to [termsofuse@overture.com](mailto:termsofuse@overture.com) .

**DMOZ (ODP):** Because they’re built by human editors, directories usually contain fewer spammy sites than search engines. If you find one at DMOZ, e-mail [staff@dmoz.org](mailto:staff@dmoz.org) .

**Lycos:** Fill out the form at [http://help.lycos.com/LycosHelp/help/watchdog/htdocs/lycos\\_watchdog\\_form.htm](http://help.lycos.com/LycosHelp/help/watchdog/htdocs/lycos_watchdog_form.htm)

## Section 4: SEO Resources

□

**Only in the full version: □**

□

□

**Sorry, the entire Section 4 is only available in the full version. □**

□

----- □

**Not in the free version: p249 to p269 □**

----- □

□

□

□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It □  
comes with an unconditional money-back guarantee, so it's a completely risk-free □  
purchase. <http://www.pandecta.com/sey.html>**

## Section 5: Outsourcing SEO



# Outsourcing SEO

- ❑ **This section is not** a DIY guide to search engine optimization.
- ❑ **This section is about** knowing what to look for in a search engine optimization (SEO) provider.
- ❑ **This section is about** knowing what questions to ask the SEO providers **before** you pay them.
- ❑ **This section is about** understanding what separates professionals from scammers.
- ❑ **This section is about** saving time and money.

**This section is ultimately about finding the right SEO provider that will get you the results you want.**

## 5.1 Introduction: Importance Of Proper SEO

The search engine optimization industry has more than its share of scammers. **Armed with this book, you will be able to find a reputable company that's right for your business, your web site and your budget.**

---

*Location, location, location* – so we're taught – is the **key** to selling **offline**.

*Search engine placement* – research shows – is the **key** to selling **online**.

The success of your web site will not be measured by how good it looks, how great the sales copy is or how fast it loads. **The success of your site will be measured by the bottom line: How much money it makes.** And for that to happen, you need customers. Your site has to be found.

**Offline:** Location is important because a great location means you're easier to find.

**Online:** Good search engine placement is important for exactly the same reason.

It is more than likely that as many as **75% of your first-time visitors will have found your site on one of the major search engines.** The problem is that there are **millions** of sites clamoring for position on those search engines.

Let's say you sell street maps. You probably have:

- more than 500 sites competing directly with you,
- a couple of thousand competing indirectly, offering free but limited street maps,
- and a million other sites that only mention "street maps" but don't offer them directly.

A search for "street maps" on a search engine like [Google](#) will produce thousands if not millions of matches. The problem that you and I face is that the average search engine user **does not look further than the first 20 matches** for his search.

**Only the first 20 sites will attract visitors.**

**Only** those 20 sites have a chance to **convert a site visitor into a new customer.**

**The rest of those sites die.**



## 5.2 The Basics Of Search Engine Optimization (SEO)



**Only in the full version:**



**The Basics Of Search Engine Optimization: Types Of Search Engines, How Search Engines Work, Keyword Targeting, Submitting Your Site, Tracking & Improving Results**



-----

**Not in the free version: p274 to p286**

-----



**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## 5.3 Should You Outsource SEO?

**According to a recent study in the U.S., only about 20% of businesses outsource search engine optimization.** The other 80% either do not know that there is such a thing as search engine optimization or they believe that they have the skills to do it in-house.

Perhaps that is why so many companies are hard to find on the search engines.

The problem is that your in-house expert probably does not know enough. Search engine optimization used to be fairly easy, but today the search engine industry is

- ❑ extremely complex
- ❑ extremely competitive and it
- ❑ changes daily.

**Your in-house expert could make mistakes like using "free for all" pages or resubmitting your site too often.** He could end up getting your site dropped from the search engines. If he uses practices such as **cloaking**, he could get your site permanently banned from the search engines.

**This costs you money in lost sales. Nine out of ten times you'll do better if you outsource.**

One of the drawbacks of outsourcing search engine optimization is that the expense is a recurring one. Having your site optimized every time it changes significantly can become

expensive. Whether or not it's worth it will depend on your site and sales copy. If your site consistently converts visitors into customers, you can afford to spend money on acquisition.

This is important.

**If your site is a [sales getter](#), you can afford to pay for traffic, because you know that a percentage of your visitors will become customers.**

If you'd like to learn more about creating a site that consistently gets the sale, I strongly recommend getting your hands on Ken Evoy's popular ebook called "[Make Your Site Sell](#)" (recently updated). It is the definitive work on selling online. Nothing else comes close.



## 5.4 The Truth About SEO Providers

Let me start off by saying that I'm **not** against the idea of hiring an SEO provider – even though it may sound that way sometimes. There are many reputable SEO providers who know more about search engines than I do.

Ok, that said, here's the reality:

**On the Internet, almost anyone can learn almost anything.**

It's a small step from there to selling that new knowledge - either as an e-book (like this one), on a subscription basis or on a consultation basis.

That's part of the beauty of the Internet, but it's also part of the problem.

There are many SEO providers that really know what they're doing, **but for every reputable, serious search engine optimization company, there are 3 that don't know enough to be selling it.**

**Most people who hire SEO companies cannot tell the difference.**

- ❑ On face value, the basement operator's site looks professional.
- ❑ On further investigation, it often sounds like he knows what he's talking about.
- ❑ Some of these "companies" even charge ridiculously high prices to add perceived value to their services.

They are not always out to mislead their customers. Some of them really believe that they know how to maximize visitors to your site, **but they make mistakes that will cost you visitors & money.**

So how do you distinguish?

The rest of this section takes the guesswork out of choosing your SEO company. On the next page we'll start off by looking at 4 warning signs.

Read this entire section - from here to the end. When you get there in 20 minutes or so, you'll know **exactly** what to look for.

## 5.5 Four Warning Signs

The “warning signs” I list here are my own.

Obviously lists like these irritate many reputable SEO providers, because it makes their customers apprehensive – sometimes **more apprehensive than necessary**. So take this for what it is: **Only my objective opinion**.

Most of the warning signs listed here have to do with ethics. If you’re not particularly concerned with **how** your SEO provider gets traffic – only that they do – then read this carefully: Unethical optimization can get your site de-listed or even banned from the search engines. When that happens, **the cost to you is enormous** while they get away with only another slight dent in their reputation. **You’re not just trusting them with getting traffic; You’re trusting them with your brand name.**

*(If you’re an SEO provider and disagree with any of these or would like to add to it, please [share your thoughts](#).)*

### 1. Spam marketing

As a general rule, **don’t do business with SEO providers (or anyone) that uses spam as a marketing tool**. Using spam is simply unethical – **not** the type of people you want to trust your site with. If you receive spam saying something like “I noticed you’re not listed in some of the search engines... bla bla bla”, write the company’s name on your “bad guy” list.

## 2. Mass submit

If they offer to submit your site to “**thousands of search engines**”, they’re trying to impress you with something **you do not need**. There are only a **handful of search engines that really matter**.

## 3. Lack of transparency

If they are unwilling to explain how they will get traffic to your site it usually means that they use techniques that are not within the rules. Some SEO providers may argue that secrecy is necessary in order to protect trade secrets. I disagree. **The kind of SEO that gets long-term results is simply about doing it right. There are no “tricks” and no “secrets” in serious SEO.**

## 4. Not listed at Google

Being listed at Google is (at the moment) the most important thing in SEO. If your SEO provider’s site is not listed at Google, they are either completely **clueless** or their site was **dropped** from the Google database because they tried to cheat.



## 5.6 Question To Ask SEO Providers

**This is where it gets interesting.**

**Armed with this book, you are able to actually test SEO providers.** You do not have to rely only on the sales copy you found on their web sites. Here are a couple of **tough questions** to ask.

**Before we look at the questions, read this paragraph carefully:**

There are many SEO providers. There are so many that you can afford to show **100** of them the door if they do not convince you that they know what they're doing. There are always more where they came from.

These questions are not difficult and they're about crucial elements of SEO, so there's no compromise. If they stumble over these, walk away.

Let's begin. Here are **questions every SEO should be able to answer:**



### ***What is link popularity and why do I need it?***

#### **ANSWER:**

A site's "link popularity" refers to its **number of incoming links** - in other words the number of links to it from other web sites.

You need it because search engines measure it (and the quality of the links) and use that info when ranking sites. Without it your site probably won't rank well.

**Link popularity is crucial.** More and more search engines measure link popularity when determining how relevant your site is for a certain search. The thinking is that, if many sites link to yours, you probably have a good site with lots of useful information.

Any SEO worth his salt should be able to suggest ways to improve your site's link popularity. There are right ways and wrong ways to do this that we looked at in [Section 3](#).

Here's a quick recap:

- **Links from FFA pages:** This one doesn't work. It could HURT your good standing with the search engines. If your SEO provider suggests using them, he does not know enough.
- **Link-share services:** This one used to work. The idea is that you join a "club" where everyone links to everyone. Most search engines now penalizing sites that use this technique.
- **Reciprocal links:** This is a bit of a gray area. Search engines are still deciding how they feel about these. The important thing at the moment is that you only exchange links with sites that are on a related topic.
- **Editorial links:** This is the most effective long-term strategy. It involves creating unique, valuable content for your site so that other webmasters will **want** to link to you.

Armed with this answer, judge whether he knows what link popularity is, how important it is and how to improve it. **There's no compromise here.** Link popularity is **vital** - that's why it's question number one. If he "will come back to you on this one", thank him for his time.



## 5.6 QUESTIONS TO ASK SEO PROVIDERS

5.6.2 Questions For SEOs: **Keyword Targeting**

*How does **keyword targeting** work? What words will my prospective customers enter in the search box?*

**ANSWER:**

Web sites can be optimized for specific keywords. The trick is in targeting the right keywords. There are ways to see what words people use when searching (referred to as "**keyword usage**"). This can then be weighed against the number of sites competing for that keyword. For more on this, refer to the **Basics of SEO** earlier in this section and **SEO Facts** in Section 3.

You could use "sex" as a keyword. Just make your site title something like "Mario's Bookkeeping Services SEX SEX SEX". After all, it's the **number 1** search term.

Right?

Yes, it's the number 1 search term, but

- it's probably difficult to sell your bookkeeping services to horny teenagers and
- there are too many sites competing for those top keywords.



**What you really want is targeted traffic.** People who are actually looking for what you offer. Selling bookkeeping services becomes so much more doable when you're selling it to people who typed "bookkeeping services".

A small amount of targeted traffic will result in more sales than huge amounts of untargeted traffic. You'll also save on hosting fees because you won't need so much bandwidth.

All your SEO provider needs to find out is whether they type "bookkeeping services" or "bookkeeping companies". **If your SEO provider cannot suggest some kind of scientific method of keyword research, he's wasting your time.**

**This is important.**

**I learned the hard way that proper keyword selection gets you twice the traffic for half the effort / money.**

Get him to explain how he collects information on actual search term usage.



## 5.7 About Guarantees

□

**Only in the full version: □**

□

□

**About Guarantees, About The Contract, Finding SEO Providers □**

□

-----□

**Not in the free version: p299 to p300 □**

-----□

□

□

□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**



## 5.10 How To Report Dishonest SEO Providers

In the US, the Federal Trade Commission handles complaints about dishonest business practices. If you feel deceived by your SEO provider, consider filing a complaint. There are three ways:

### 1. Online

Visit [www.ftc.gov](http://www.ftc.gov) and click the “File a Complaint Online” link.

### 2. Phone

Call 1-877-FTC-HELP

### 3. Regular Mail

Write to:

Federal Trade Commission  
CRC-240  
Washington, D.C. 20580

If you're outside the US, try [www.econsumer.gov](http://www.econsumer.gov)



# Section 6: The Search Engine Dictionary



## SE Dictionary

### SECTION 6: CONTENTS AT A GLANCE

- 6.1 About The Search Engine Dictionary
- 6.2 The Search Engine Dictionary

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

## 6.1 About The Search Engine Dictionary [www.searchenginedictionary.com](http://www.searchenginedictionary.com)

(CLICK HERE TO JUMP STRAIGHT TO THE DICTIONARY)

### A Separate Book & Web Site

I initially planned to explain some search engine terminology at the end of this book. That section kept growing – to over 100 pages – so we decided to split it off into a separate book called the Search Engine Dictionary.

It's still included in the Yearbook (below), but also available as a **free PDF download** from [www.searchenginedictionary.com](http://www.searchenginedictionary.com).

### The Most Complete Search Engine Dictionary

Calling this dictionary “complete” is probably a bit arrogant. It is however **based on a combination of the five biggest search engine glossaries on the Web** – with many new entries added and old definitions updated and expanded. I also added a couple of general web marketing terms that are often used in the context of search engines.

I'm confident that this is **the most complete** glossary of search engine terms available anywhere.

## Continued Research

No matter how complete the dictionary is now, I realize that new words are constantly being created to describe new concepts.

**But I've thought of that...** On my web site ([SearchEngineDictionary.com](http://SearchEngineDictionary.com)) anyone can suggest new additions or corrections. In return...

**You get some free exposure (and a link to your site)**

I invite you to become part of this project. If you can think of a search engine related term not listed on the web site or you can improve on our definition of a term already listed, send your suggestion to me. **If I use it, your name (and a link to your site) will be added below the new entry.** Your new entry / correction plus the link will be published on the [SearchEngineDictionary.com](http://SearchEngineDictionary.com) site and in the Search Engine Dictionary PDF book.

[Click here to suggest a new term.](#)

[Click here to suggest a better definition of a term already listed.](#)

## Update Cycle

Every **January** the entire [SearchEngineYearbook.com](http://SearchEngineYearbook.com) web site is compiled into a new Search Engine Dictionary – just like the current Search Engine Dictionary was compiled from the current site.

**So be sure to check back every January.**

You can either slap a sticky note on your computer or you can **let me remind you**. Just send a blank e-mail to [sed-subscribe@topica.com](mailto:sed-subscribe@topica.com) to be notified when we update.

## About the Price

The dictionary is **free** – and we'd like to keep it that way. Please help us by simply linking to <http://www.searchengineyearbook.com> and...

**...by redistributing the dictionary freely.**

**Yes, really. Give the dictionary away from your site.** Your visitors will LOVE you for it. **As long as you don't change the contents or sell it** and as long as you're giving away the **most recent edition**, we get extra readers and you add real value to your site.

A win-win if there ever was one.



## 6.2 The Search Engine Dictionary

Note: □

The [www.searchenginedictionary.com](http://www.searchenginedictionary.com) web site is constantly updated. If you can't find the term you're looking for in this version, consider visiting the web site. You'll probably find it there. Click anywhere in this block to open the site in your browser.



### About

[www.about.com](http://www.about.com)

Formerly known as The Mining Company, About is a large Internet **directory**.



### above the fold

With reference to the top part of a newspaper, the term is used on the Net to describe the top part of the page that the user can see without scrolling down.

### acquisition

A term used in Internet marketing to describe the point at which a visitor becomes a qualified lead / customer. Generally this is the point where the visitor

- buys a product or
- provides contact details and indicates an interest in the product or
- subscribes to a newsletter.

### acquisition cost

Total cost of an advertising / marketing campaign divided by the number of visitors (visitor acquisition cost) or divided by the number of customers (customer



acquisition cost). Monitoring of acquisition cost is an important factor in effective **PPC** advertising.

--- SIDEBAR ---

**Remember, orange text indicates internal links. Clicking on an internal link takes you directly to that word in the dictionary.**

### **adjacency**

Referring to the relationship between words, particularly words used in a **search engine query**. Search engines typically assign higher value to pages where the search terms appear next to one another (as in the query) than to pages where the search terms are separated by other words.

### **adjacent searching**

see **proximity**

### **ad broker**

An Internet advertising specialist. Ad brokers act as middlemen between web site owners with advertising space to sell and advertisers.

### **ad inventory**

The number of potential **page views** a site has available for advertising.

### advanced search

An option at most of the major [search engines](#) that allow users to specify certain search criteria. For example, users can elect to see only documents added to the [database](#) after a certain date, documents in specific languages etc.

### AdWords

[Google's PPC](#) program.

### affiliate program / affiliate link

Affiliate programs allow other people to sell your products on a commission basis. All your affiliates really do is place a [link](#) to your site. When a [visitor](#) arrives at your site, your affiliate program "makes a note" of the site that referred him. If a visitor buys something and the referring site belongs to one of your affiliates, you pay that affiliate either a percentage of the sale or a fixed amount - according to your agreement.

### agent name delivery

Different pages can be presented at the same URL. Different pages are delivered based on the agent name requesting the page. Typically, agent names starting with "Mozilla" indicate regular browsers while search engine [spiders](#) use names like [Googlebot](#), [Scooter](#) etc. Agent Name Delivery is not a very effective form of cloaking though. Search engines can (and do) disguise spiders as "Mozilla" agents. Also see [cloaking](#), [IP delivery](#).

## algorithm

Algorithms are sets of rules according to which **search engines** rank web pages. Figuring out the algorithms is a major part of search engine optimization. The thinking is that if you understand how they calculate **relevance**, you can make specific pages on your site super relevant for specific **search terms**. For more on algorithms and **SEO** in general, please refer to **Section 3**.

## algorithm-based software

Data mining software typically used for statistical analysis.

## AliWeb

[www.aliweb.com](http://www.aliweb.com)

An Internet **directory**.

## AlltheWeb

[www.alltheweb.com](http://www.alltheweb.com)

A very large **search engine**, gaining in stature and popularity. At this stage (2002) it seems to be the top contender for **Google**'s throne. In a study by Pandecta Magazine, conducted in the 4<sup>th</sup> quarter of 2002, AlltheWeb was estimated to have the second largest database (after Google). It also did well in relevancy test: 3<sup>rd</sup> after Google and **Wisnut**. It lost out in the speed test though. It came in last. For more details on that study, AlltheWeb and the other search engines worth knowing about, please refer to **Section 1**.

Note added to the free version of SEY 2003: □  
□  
You'll see links like this one that says "Section 3" (at the end of the "algorithm" definition). □  
These links take you back into the book where the topic is discussed in more detail. If you □  
click one of these links and nothing happens, □  
it means that that part of SEY 2003 has been □  
left out of the free version.

**alltheweb**...  
all the web. all the time

## AltaVista

<http://www.altavista.com>

A very popular **search engine**, once reported to have the biggest index of them all. According to recent estimates, it's now the 4<sup>th</sup> largest. For a detailed look at AltaVista and the other major search engines, refer to **Section 1**.



## alt attribute

More commonly known as the “alt tag”. The alt attribute is an **HTML** element specified within an image tag. The syntax is:

```
<IMG SRC="main-logo.gif" ALT="Pandecta Logo">
```

The text in the alt attribute, “Pandecta Logo” in this example, will be displayed in the place of the image “main-logo.gif” while the image loads or if the user has images turned off. In most **browsers** the text also appears as a “tool tip” when the user hovers the mouse pointer over the image after it has loaded.

Creating an alt attribute for images is not required, but recommended since the alt text is factored into the **algorithms** of most **search engines**.

## alt tag

Common name (erroneous) for the **alt attribute**.

## alt text

Text specified in the **alt attribute**.

## applet

A small application, usually in Java, usually for use on the Web.

### ArchitextSpider

The name of the **Excite search engine's spider**.

### Ask Jeeves

<http://www.askjeeves.com>

A fairly popular **search engine**. Its claim to fame is that it lets you to enter plain text questions as opposed to only keywords.

Ask Jeeves receives search results from **Teoma**, **Overture** and **ODP**.



### ASP

Active Server Pages. A server-side scripting language used to deliver **dynamic content**.

### attribute

A term used in the **HTML** language to refer to display settings. For example, the "bgcolor" attribute inside the <body> tag specifies the background color of a page.

### audience reach

In the context of **search engines**, the term refers to the percentage of the total Internet population that use a particular search engine during a given month. Together with **search hours**, audience reach is an important measure when calculating the popularity of the different search engines.

This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)

**automated submission**

The practice of machine-based, automatic submission of URLs to [search engines](#), usually with the use of [submission software](#) or [submission services](#).

Also see [mass submission](#). For more on automated submission, mass submission and [submission software](#) (and their dangers), refer to [Section 3](#).



### **bait-and-switch**

A technique (considered [spam](#)) used in [SEO](#). It involves creating an optimized page and a regular page. The optimized page is submitted to the search engines and replaced with the regular page as soon as the optimized page has been [indexed](#).

### **banner blindness**

Refers to a “condition” amongst experienced web users who tend to automatically ignore banner ads. Banner blindness is arguably the main cause of low [click-through rates](#) in banner advertising. For more on Internet advertising, please refer to [Section 3](#).

### **begins-with partial word matching**

Some search engines will match indexed words that contain a search term at the beginning. For example, if you're searching for "guns", documents containing the following variations of the term will show up in your search results:

Guns (exact match)

Gunsmith (Begins-with partial word matching)

Gunslinger (Begins-with partial word matching) etc.

Also see [partial word matching](#).

**bells-and-whistles**

Advanced features. A web site is said to have too many bells-and-whistles when it contains unnecessary animations etc.

**beta**

A testing stage / testing version of a product. For example, when a beta version of a [search engine](#) is released, users can access it online and are encouraged to report [bugs](#) and give general feedback.

**Boolean search**

A Boolean combination of terms allowing the inclusion or exclusion from search results of documents containing certain words. This is achieved through the use of [operators](#) such as AND, NOT and OR.

**bibliometric analysis**

see [link tracking](#)

**blog**

The name originates from “Blogger”, which was the name of a content management program. The term “blog” is today used to describe sites that can best be described as mini-directories, often populated with the site owner’s personal favorites and his/her comments. Blogs often contain message boards / chat rooms etc.



**bridge page / bridging page**

See [doorway page](#).

**broadband**

short for: broad bandwidth

A high-capacity data transmission channel. Broadband access to the Internet allows users to send and receive data at a much higher speed than is possible with a regular phone line. Broadband utilizes the same frequency division multiplexing technique used in cable TV, allowing for the simultaneous transmission of different types of signals.

**broken link**

See [dead link](#)

**browser**

a.k.a. Web browser

A program used to display Internet content. Two of the best-known and most widely used browsers are Netscape Navigator and Microsoft Internet Explorer. Browsers read coded ([HTML](#), [JavaScript](#) etc.) pages and display them as web pages. Browsers typically include features such as bookmarks, back & forward buttons etc.

**browser compatibility**

Referring to the different ways different [browsers](#) display the same page. A key consideration in web design (and [SEO](#)) is to create pages that are browser

independent – in other words pages that work as they are supposed to regardless of the user's choice of browser.

**bug**

An error or glitch in a program / [search engine](#).

This dictionary is also available as a separate PDF book. Get it **(free)** from  
[www.searchenginedictionary.com](http://www.searchenginedictionary.com)



### **Cascading Style Sheets**

See [CSS](#).

### **categorization**

The practice of grouping web pages by topic to form a [directory](#).

Also see [Classification](#)

### **category**

In the context of Web [directories](#), categories refer to [collections](#) of links to sites of a similar topic.

### **CGI**

Common Gateway Interface - a popular interface between web server software and other programs.

### **channels**

See [Directory](#); [Category](#)

**classification**

The process of organizing documents available online into topical **categories** to form **directories**. These are normally hierarchical tree structures with “Main Categories” and a number of “Sub Categories” which often go several levels deep.

**click tracking**

Search engines can track user clicks in order to “learn” from users which pages are most **relevant** to a **query**. The best-known example is that of “Direct Hit”, a discontinued search engine that not only tracked clicks but also logged the amount of time users spent on pages returned in order to improve **relevance**.

**client**

A computer, program or process requesting information from a server. Email programs are sometimes called e-mail clients. They request e-mail messages from pop3 servers. **Spiders** (like **Googlebot**) and **browsers** (like Internet Explorer and Netscape) are also clients.

**click through (click-through; clickthrough)**

Referring to the action of clicking through from, for example, a search engine’s results page to a web site. **Click through rates** become especially important in Internet advertising where it is an important factor in determining the success of an advertisement.

### click through rate (CTR)

a.k.a. click rate

Often used in Internet marketing to describe the percentage of users who click on a link or advertisement. The CTR is used as a measure to determine the effectiveness of a link / advertisement. It is most effective if used in conjunction with other measurements like [conversion rate](#) (CR).

For example, if an advertisement is displayed 1000 times (1000 [impressions](#)) and generates 10 [click throughs](#), the CTR is 1% (10 / 1000 x 100%).

### cloaking

The practice of delivering content based on the [IP address](#) of the [client](#). The practice is sometimes defended by saying it's a way of protecting code from theft. It should be noted that the practice of cloaking can get your site banned from the search engines. For a detailed discussion on cloaking and links to cloaking resources, please refer to [Section 3](#).

### cluster

Search results grouped together (to save space on the [SERP](#)), usually based on a shared top-level [domain](#).

### clustering

A technique the search engines use to group different pages from the same [domain](#) in their search results pages. Without clustering, the top spots for certain search terms are often completely dominated by one site. [Clusters](#) usually consist of one

or two pages from one domain with a link that says something like “More results from pandecta.com”.

### **collaborative filtering**

Also known as “social filtering”. A technique used to improve **relevance**, it returns documents other users with similar **queries** found **relevant**. This technique is also very effective in cross selling, as seen at Amazon.com (“People who bought ‘*Mary’s Guide to Fast Food*’ also bought ‘*Jane’s Recipes*’”)

### **collection**

A group of documents queried.

### **collection fusion**

The practice of combining search results from multiple **collections**. **Meta search engines** are faced with the problem of effectively combining & re-ranking results that have already been ranked by different **algorithms**.

### **combined log file**

A log file that tracks visitors on a web site. A combined log file typically includes additional information on user agents, **referrers** etc.

Also see **log file** and **common log file**.

For more on log file analysis and downloadable tools that make it easier, please refer to **Section 4**.

**comment**

Comment tags (in [HTML](#)) allow the site designer to enter comments explaining the code, making it more understandable for human readers. Comments are not displayed by the [browser](#). Comments are enclosed by the comments tag: `<!-- like this -->`. The comment tag is also used to enclose [scripts](#), ensuring that the raw code is not displayed on non-compliant [browsers](#). Comment tags are sometimes loaded with [keywords](#) to artificially inflate a page's [ranking](#). Loose that sparkle in your eye though... most search engines ignore comment tags completely.

**common log file**

A standard log file with no additional information.

Also see [log file](#) and [combined log file](#).

For more on log file analysis and tools that help you read log files, please refer to [Section 4](#).

**concept search**

A [search](#) for documents related conceptually to a [search term](#), rather than for documents that actually contain the search term itself.

**conversion cost**

Total cost per sale, calculated by dividing the total cost of an advertising campaign by the number of resulting sales. For example, if \$1000 is spent on an advertising campaign and that campaign results in 20 sales, the conversion cost per sale is \$50 ( $\$1000 / 20$ ). That means it costs \$50 to generate one sale.

**conversion rate (CR)**

The percentage of site visitors that deliver the **most wanted response (MWR)**. The CR is an important measure of the effectiveness of the online sales effort. For example, if 4 out of every 100 visitors to a site deliver the MWR, the CR for that site is 4%.

**cosine similarity**

See **Similarity**.

**CPA**

Cost per action. Similar to **CPS**. Also see **conversion cost**.

**CPC**

Cost per click. The total cost of an advertising campaign divided by the resulting number of **unique visitors**.

**CPL**

Cost per lead. The total cost of an advertising campaign divided by the resulting number of new **leads**.

**CPM**

Cost per thousand **impressions** (M= Roman numeral for 1000). A pricing system often used in the banner advertising industry. Typically a fixed price is offered for 1000 impressions of a banner. The price is usually influenced by the topic of the site (how targeted the audience is) rather than the popularity of the site.



**CPS**

Cost per sale. Similar to [CPA](#). Also see [conversion cost](#).

**crawl**

What [spiders](#) do. It refers to the action of following links to navigate from page to page and site to site.

**crawler**

See [Spider](#).

**cross linking**

Referring to links between a family of [domains](#) – for example your business site, your personal homepage and your cat's homepage. Cross linking is sometimes used to inflate [link popularity](#) and excessive cross linking is (rumored to) be penalized by the [search engines](#).

**CSS (Cascading Style Sheets)**

An add-on to HTML that allows for more accurate control over the way a web page is rendered. CSS allows designers to create custom styles that are then applied to the web site in one of a variety of ways. The main benefit is that something like text colors for an entire site can be changed by editing only the CSS file. CSS can also be used in [SEO](#), but most SEO techniques that involve CSS are considered [spam](#).

**counter / page counter**

Typically accompanied by something like “You are visitor number \_\_\_ since Oct 2001”. Counters count **page views**, not visitors. The difference is that one visitor can generate many page views by opening many pages on the site. Counters offer a relatively inaccurate way to measure site traffic and are generally considered amateurish. **Log files** offer far more accurate and comprehensive visitor data.

**cybersquatting**

The practice of buying **domains** that contain popular trade names (for example fordmotors.com) or are common misspellings of popular trade names (for example gogle.com). The intent is usually to either resell the domain or to pull traffic through misspellings, rather than to develop a serious, unique site. Traffic gained through misspellings is often automatically redirected to another domain.

Also see **DNS parking**.

**cybrarian**

Referring to professional online researchers. Sometimes also referred to as “super searchers”.

This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)

**data traffic**

Refers to the number of packets traversing a network.

**database**

An electronic filing system containing information that is usually highly organized and categorized. The benefit of electronic filing by means of a database is that specific information can easily be extracted according to given parameters. **Search engines** are essentially very large, searchable databases. **Dynamic web pages** typically rely on databases.

**date range / date limit**

Most of the major **search engines** allow users to limit **search results** to documents created / modified on / before / after a specified date.

**dead link**

A **link** to a page that no longer exists or has been moved to a different **URL**. Search engine **spiders** regularly respider pages in its **index** and removes dead links. Most search engines also offer ways for users to report dead links.

### deep linking

The practice of linking to the inner pages of another web site – as opposed to linking to the homepage. Although the vast majority of site owners don't mind deep links to their sites, it should be noted that deep linking has potential legal ramifications.

### de-listing

Referring to the removal of pages from a search engine [index](#). De-listing can occur at the request of the site owner or a variety of other reasons. Most often, de-listing occurs when a page breaks one of a search engine's [submission rules](#), making itself guilty of some sort of [spamdexing](#). [Section 3](#) contains comprehensive guidelines to help you avoid spamdexing and de-listing.

### description

In the context of the search engines, the description refers to the descriptive text accompanied by a title and [URL](#) in the search results page. Some search engines take this description from the meta description while most generate their own from the page content. [Directories](#) often ask for a description when you [submit](#) your page.

### description tag

An [HTML](#) tag that gives a general description of the contents of the page. This description is not displayed on the page itself, but is largely intended to help the [search engines](#) index the page correctly. Some search engines use the description found in the description tag on their [SERP](#)s. A growing number of search engines

are completely ignoring the description tag. For a more detailed look at the description tag and other types of [meta tags](#), please refer to [Section 3](#).

### DHTML

Dynamic [HTML](#). DHTML is sometimes referred to as the next generation HTML. It gives site designers increased control over the appearance of a site.

### Direct Hit

Discontinued [search engine](#). It was acquired by [Ask Jeeves](#), who, in my opinion, failed to capitalize on its tremendous promise. What made it special was that it tracked user behavior and “learned” from it, constantly improving the [relevance](#) of [search results](#). Direct Hit has been assimilated into [Teoma](#), Ask Jeeves’ other acquisition.



### directory

A categorized collection of links to the web, usually compiled manually. Directories can either be general (to the entire web) like [ODP](#) or Topical like the [Dotcom Directory](#). Although they cannot rival [search engines](#) for [index](#) size, they generally do offer higher quality [search results](#), arrived at through some editorial selection process.

### DMOZ

See [ODP](#).

## DNS parking

A **domain** is set to be “parked” when it has been registered but not developed into a web site. The registrant pays the annual renewal fees to prevent the domain from falling into someone else’s hands. DNS parking is typically done to protect trademarks. Domains registered for resale are usually also parked.

## Dogpile

<http://www.dogpile.com>

A popular **meta search engine**.



## domain / domain name

A sub-set of internet addresses. Top-level domains are divided into .com, .net, .org, .biz, .info, .gov and .edu. Apart from these there are also country-specific domain extensions like .ca, .com.au, .co.za, .fr etc. In **SEO** it is generally accepted that having a keyword-rich domain is beneficial. **Section 3** contains a more detailed discussion of the importance of domain name selection in SEO, as well as what to look for when choosing a domain.

## doorway domain

A keyword-rich **domain** name used to achieve high **search engine** ranking for a particular keyword / key phrase. Similar to an **doorway page**, a doorway domain serves only as a point of entry that leads search engine **traffic** through to the “real” content of the page. This technique is not advisable. Domains containing only a page or two don’t normally rank well on the search engines and **spiders** typically ignore pages that automatically redirect to other pages. For a detailed discussion

on multiple domains and automatic redirection, please refer to the discussion of domain names in [Section 3](#).

### **doorway page**

Also known as bridge pages, bridging pages, entry pages and landing pages. Referring to a page designed to rank well for a selected keyword and redirect visitors to another, “real” page. Important here is that there are two kinds of doorway pages: those generated automatically based on a template and manually created keyword focused content pages ([KFCPs](#)). The first kind is considered [spam](#) and penalized by most search engines. The second is an important and usually very effective [SEO](#) technique. For a detailed discussion of doorway pages and all the do’s and don’ts, please refer to [Section 3](#).

### **drill down**

The action of clicking on links within a web site or [directory](#), working through [categories](#) and [sub-categories](#), in order to find specific information.

### **dynamic content**

Web site content generated automatically, usually from a [database](#) and based on user actions / selections. Dynamic content typically changes at regular intervals, for example daily or each time the users reloads the page. [SERPs](#) are dynamically generated pages, changing depending on user input.



### **electronic library**

The term normally refers to web sites that provide access to public information like catalogs, e-books, [databases](#), audio files etc.

Also see [cybrarian](#).

### **entry page**

See [doorway page](#)

### **EPC**

Earnings Per Click. A unit of measure used to determine a site's ability to convert visitors into customers. Calculated by dividing total sales amount by total [page views](#).

Also see [EPV](#), [ROI](#), [conversion rate](#)

### **EPV**

Earnings Per Visitor. A unit of measure used to determine a site's ability to convert visitors into customers. Calculated by dividing total sales amount by total number of visitors to the site.

Also see [EPC](#), [ROI](#), [conversion rate](#)





## Excite

<http://www.excite.com>

A major **search engine**. For a detailed look at Excite and the other major search engines, please refer to our detailed discussion of Excite in **Section 1**.

## exact match

If not for **partial matching**, **fuzzy matching**, **collaborative filtering** and **stemming**, **search engines** would only return exact matches. A search for “power” would only return documents containing the exact term, not documents containing variations or related terms like powerful, strength etc.

## eye candy

Aesthetically pleasing web sites are said to provide eye-candy. The term is used to describe sites both positively and negatively. In the context of search engines and SEO, eye candy is generally perceived as unnecessary, not contributing to the marketing effort.

This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)

**faceted search**

The combination of **Boolean operators** and **parenthesis**. Faceted search allows for very specific, powerful searches.

**fake copy listings**

The practice of stealing content from another web site, republishing it and **submitting** the duplicate page to the **search engines** in a hope to steal **traffic** from the original site. Apart from the obvious ethical problem, copyright legislation is slowly adapting itself to the Internet, making it increasingly difficult for thieves to steal content. The copyright holder may also appeal to the search engine(s) that listed the duplicate page(s) and to the thief's hosting company. It is advisable to display a clear copyright notice (or a link to one) on every page of a web site.

**false drop**

A web page displayed in the **SERP** that is not clearly **relevant** to the **query**. The most common cause of false drops is words with multiple meanings. If the query gives no indication of context, the **search engine** has no way of predicting which of the possible meanings the user has in mind. The term "argument", for example, has

different meanings in general use and in programming jargon. Other possible causes of false drops include [spamdexing](#) and [bugs](#).

## FFA

Free For All. Referring to web pages that contain links to other pages and very little (or nothing) else. The difference between FFA pages and [directories](#) is that directories contain links to sites selected through some editorial process, while FFA pages allow anyone to add a link to any page. For a more detailed look at FFA pages and their dangers, please refer to [Section 3](#).

Also see [link farm](#)

## Flash

Short for “Macromedia Flash”

A vector graphic animation technology that requires a plug-in but is browser-independent.

## flash page

See [splash page](#).

## FindWhat

[www.findwhat.com](http://www.findwhat.com)

A popular [PPC search engine](#).



**frames**

An **HTML** tag construct that allows designers to display two or more web pages simultaneously. The general perception is that frames can greatly improve site navigation, but they are browser-dependant and not **search engine** friendly. Most search engines do not index framed pages correctly. For a more detailed look at the problems with frames and possible solutions, please refer to the **Section 3**.

**frequency cap**

A limit used in Internet advertising. It refers to the maximum length of time or number of times a user will be exposed to a specific type of advertisement.

**FUD**

Fear, Uncertainty and Doubt.

The action of spreading fear, uncertainty or doubt. It is a fairly straight forward but malicious technique that is typically used to negatively influence the public perception of a competitor or his/her product.

**full-text search engine / full-text index**

A full-text **search engine** indexes every word on every document it **spiders**.

**fuzzy search**

A type of search made possible by **fuzzy matching**. The **search engine** returns results that it predicts will be relevant, even when the terms used in the **query** does not appear anywhere in the matched document.

### fuzzy matching

As opposed to exact matching.

Fuzzy matching attempts to improve **recall** by being less strict but without sacrificing **relevance**. With fuzzy matching the **algorithm** is designed to find documents containing terms related to the terms used in the **query**. The assumption is that related words (in the English language) are likely to have the same core and differ at the beginning and/or end. A search for “matching”, for example, would also return documents containing match, matched etc. Unfortunately it will also return documents containing unrelated words like catching, matchbox etc.

This dictionary is also available as a separate PDF book. Get it **(free)** from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)



### gateway page

See [doorway page](#)

### ghost site

A site that remains available online but is no longer updated. Ghost sites are not simply abandoned sites. They typically contain some statement explaining that it is no longer being updated.

### Go.com

[www.go.com](http://www.go.com)

Used to be a top [search engine](#), then named “[Infoseek](#)”. Acquired by Disney, Go.com now simply displays search results from [Overture](#).

### Go Guides

[www.goguides.org](http://www.goguides.org)

A web [directory](#) started by former editors of the [Go](#) directory. Also see [JoeAnt](#).



## Google

[www.google.com](http://www.google.com)

Arguably the biggest, fastest and most accurate **search engine**.

Google is famous for its **PageRank** system. For a detailed look at Google, how important it is, how to rank well at Google and how Google compares to other search engines, please refer to **Section 1**.



## Googlebot / Google Bot

**Google's spider**.

## Googlehacking

The name of a "**Google game**". Google has an immense **database**. The aim is to enter a **query** that returns only one result from the database. Yes, that's it. If you see "Results 1-1 of 1", you win.

## Goto / GoTo

A **PPC search engine** now known as **Overture**.

## Gulliver

The name of the **spider** used by **Northern Light**.



### heading / heading tag

An **HTML** tag of 6 sizes. The syntax is `<H1></H1>`, `<H2></H2>` etc., with H1 being the largest. Heading tags have significance in **SEO**. **Search engines** normally assign more weight to documents where the keywords used in the **query** are found inside heading tags. Pages that use heading tags generally rank higher, but excessive use might get the page **de-listed**.

### hidden text

Text on a web page designed to be visible to **spiders** but not to human visitors. The aim is to load the page with keywords without deterring from the visitor's experience. Of the various techniques of hiding text, the most common is to set the text color to exactly or nearly the background color. Most **search engines** can now detect hidden text and consider it **spamdexing**. Pages that contain hidden text are penalized or even **de-listed**. For more on hidden text and the dangers of using hidden text, please refer to the **Section 3**.

### hit

One hit is one request for a file on a web server. A visitor opening a page with 5 images will in the process generate 6 hits (1 each for the images and one for the



**HTML** page itself). The term is sometimes also used with reference to the number of results (hits) a **search engine** returns for a specific **query**. Hits are often confused with **page views** and **unique visitors**. Also see **log file**

### homepage / home page / home

The main “index” page or navigation hub of a web site. The homepage is not necessarily the first page. Many sites use **splash pages** to welcome visitors and lead them from there to the homepage. At most **search engines** you can simply submit your homepage and leave it to the **spider** to **crawl** the rest of the site from there.

### Hotbot

[www.hotbot.com](http://www.hotbot.com)

A fairly popular **search engine**, although its popularity has declined sharply as **Google** rose to dominance. Hotbot was once reported to have the largest **database** of them all. In a our comparison of search engine database sizes (4<sup>th</sup> quarter of 2002) it was estimated to have the 4<sup>th</sup> largest database after Google, **AlltheWeb** and **Wisenut**. HotBot exploits NOW (Network Of Workstations) parallel computing technology in order to achieve both speed and size. NOW is basically interconnected workstations and LANs. When you add up the combined computing power of those smaller components, you get supercomputer-class performance. For more on Hotbot, please refer to **Section 1**.



**hot linking**

The practice of displaying images files, video files etc. on a web site when those files are on another (usually someone else's) server. Effectively the site displays content that uses up someone else's bandwidth. Hot linking is generally considered unethical unless prior permission is obtained.

**HTML**

Hypertext Markup Language. HTML is the primary language used to create web sites.

**HTTP**

Hypertext Transfer Protocol. HTTP is the most common transfer protocol used to facilitate communication between servers and browsers.

**hyperlink / link**

Clickable content on a web page usually leads to another page, another site or another part of the same page. The clickable content therefore is said to link to the other page / site / part of the same page. **Spiders** use links to **crawl** from one page to the next as they index web sites.



### image map

An image that has different clickable areas linked to different pages. Image maps can either be imbedded in the [HTML](#) code or called as an external file. [Search engines](#) usually have difficulty [spidering](#) image maps when they are included from external files.

### impression

One display of an image or advertisement.  
Also see [CPM](#)

### inbound link

When site A links to site B, site A has an [outbound link](#) and site B has an inbound link. Inbound links are counted to determine [link popularity](#), an important factor in [SEO](#). For more on link popularity, link building and the importance of inbound links in SEO, please refer to [Section 3](#).  
Also see [reciprocal link](#)

**index**

Plural: indices / indexes.

Referring to the searchable **database** of documents stored by a **search engine** – often simply referred to as a search engine’s database. When used as a verb, it describes the process of adding sites to a searchable database. The term is sometimes also used to refer to **directories** like **ODP**.

**index file**

A file created by a search indexer program, designed to store information in a format that makes fast retrieval possible.

**information extraction / information filtering**

A field of study related to **information retrieval** that attempts to identify semantic structures in order to extract relevant data.

**information retrieval**

A field of study related to **information extraction**. Information retrieval is about developing systems to effectively **index** and **search** vast amounts of data.

**Infoseek**

Infoseek is the old name for the Go.com **search engine**. Go.com was acquired by Disney and started displaying results from **Overture**, a **PPC** search engine. Today it is little more than a mirror of the Overture search engine.



## Inktomi

A large **database** of web sites, started in 1996, that feeds results to some **search engines**. Inktomi also provides a range of other services, including content networking solutions, search solutions and wireless solutions. For a more detailed look at Inktomi and its importance in **SEO**, please refer to **Section 1**.



## intranet

Essentially a web site or group of (usually interlinked) web sites that is only accessible to people within a specific group or organization. Most large companies have intranets. Intranets offer a safe place for employees to publish information that improves workflow. Intranets typically house shared applications, internal telephone and e-mail directories, rules and regulations, help files etc. Many large intranets have a **search** facility that allows users to find specific information more easily.

## inverse document frequency

A measure of how rare a term is in a **collection**.  
Also see **term frequency**.

## inverted file

A file that represents a collection of documents or **database**. The inverted file lists all words that appear in all documents in the database, as well as a reference to the document where the word appears.

### **invisible web**

A popular collective name for documents of types that **search engines** do not typically index. Because they are not in any search engine **database**, they can be very difficult to find and are in a sense invisible. Recently a couple of specialized search engines have begun an attempt to make the invisible web more accessible.

### **IP**

Internet Protocol. Essentially a set of standards that are necessary to ensure that data sent between networks are readable on both sides. IP provides the standard for the way data is scrambled and sent over the Internet, while TCP (transmission control protocol) provides a standard for the way data is unscrambled. These two standards are essential to the working of the Internet.

### **IP address**

Every Internet user and every server has a numeric address. Something like 123.45.67.890. IP addresses provide essential identification online. **Domain names** can be set up to have a unique IP address, something that is useful in **SEO**. For more on the role of IP addresses in SEO, please refer to **Section 3**.

### **IP delivery**

Similar to **cloaking**. A technique for automatically delivering different pages to different users based on the user's **IP address**. Although IP delivery has legitimate uses (like delivering different content to people from different geographical areas), it has been applied extensively in cloaking, causing IP based delivery to be banned

by most **search engines**. For more on IP delivery and the potential dangers, please refer to **Section 3**.

### **IP spoofing**

A controversial technique for reporting a false **IP address**. In the context of **search engines**, IP spoofing is sometimes used to refer to the practice of **cloaking**.

This dictionary is also available as a separate PDF book. Get it **(free)** from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)



## Java

A powerful, platform-independent programming language. In other words, Java can be used to create advanced programs that can be run on different computers with different operating systems. Java is also used extensively to create **applets** for use on the web.

## JavaScript

A comparatively simple scripting language used extensively on the web to, amongst other things, make web pages interactive. JavaScript shares characteristics of **Java**, but it is less complex and less powerful. One of the main benefits of JavaScript is that it can seamlessly integrate with **HTML**.

## JoeAnt

[www.joeant.com](http://www.joeant.com)

A **directory** started by former editors of the **Go** directory.  
Also see **Go Guides**.







## Kanoodle

[www.kanoodle.com](http://www.kanoodle.com)

A comparatively small **search engine** that uses the **PPC** model.

## keyword

A word used in a **query**. In **SEO**, pages are typically **optimized** for specific keywords. Keywords are targeted based on what users looking for the specific information or product are most likely to use as part of a query. Accurate **keyword targeting** is considered by most to be essential to effective SEO. For more on keyword targeting and ways to obtain statistics on actual keyword usage, please refer to **Section 3**.

## keyword density

A measure of the percentage of words on a page that are specifically chosen **keywords**. When a user enters a **query**, **search engines** display a list of pages containing the **search terms**. These are ranked based on (amongst many things) the percentage of words on a page that are similar to the words used in the query (keyword density). When keyword density is inflated artificially, it is often referred to as **keyword stuffing**.

**keyword domain name**

A **domain name** that contains **keywords**. Please refer to **Section 3** for a more detailed look at the importance of keywords in **SEO**.

**keyword phrase / key phrase**

Two or more words that form a “**keyword**”. In **SEO** the term keyword is usually used to refer to both keywords and key phrases. It simply refers to words entered in a **query** / words a page has been **optimized** for.

**keyword purchasing**

Not to be confused with **PPC**, keyword purchasing refers to the practice of buying advertising space on specific **SERPs**. It offers a fairly high level of targeted advertising, because the ad is only displayed to users who enter specific **keywords** in a **query**.

**keyword search**

Basically the same as **search**, it refers to a search for documents containing specific **keywords**.

**keyword stuffing**

Excessive repetition of **keywords** in an attempt to artificially inflate **keyword density** and improve a page’s **ranking**. Keyword stuffing is easily detected by **search engines** and pages that use this technique are penalized.

**keyword tag / keywords tag**

A **meta tag** listing **keywords** associated with the page.

**keyword targeting**

The practice of optimizing certain pages of a web site to rank well in a search for specific keywords. Keyword targeting is generally considered vital to effective SEO. For more on keyword targeting and ways to obtain statistics on actual keyword usage, please refer to the **Section 3**.

**KFCP**

Keyword Focused Content Page. The term was coined by e-selling guru Ken Evoy and refers to a “**search engine** friendly” **doorway page**. Sometimes simply called honest doorway pages. For more on KFCPs and doorway pages, the differences and the dangers, refer to our discussion of doorway pages in the **Section 3**.

**kickback marketing**

A collective name for post-dotcom-bust Internet marketing techniques that focus on revenue sharing. Examples of kickback marketing include **affiliate programs**, pay-for-performance programs, bartering etc. The success of kickback marketing lies in its utilization of the nature of the Internet to effortlessly pass customers back and forth between affiliated sites.

**KISS**

Keep It Simple Stupid. Generally considered one of the golden rules of web design and online business.

**legacy data**

Referring to information contained in old file types. Usually legacy data can only be viewed with special reader programs.

**lead**

A typical **MWR**, mostly referring to a potential customer's contact details. Many companies don't sell online but rather use their sites to generate leads that are then followed up. Many **affiliate programs** also reward affiliates on a per-lead basis rather than a per-sale basis.

**link**

See **hyperlink**

**linkage**

See **link popularity**

**link checker / link validator**

A program that scans web sites for **dead links**. Most link checkers generate reports that list all dead links on a site.

**link farm**

Similar to [FFA](#) pages, it refers to a page where anyone can list a web site to be linked to. Link farms are used to artificially boost [link popularity](#). Most [search engines](#) penalize sites associated with link farms.

Also see [FFA](#)

**link popularity / linkage**

A measure of the quantity and quality of [inbound links](#). Link popularity is an important factor in [SEO](#). For more on its role in SEO as well as legitimate ways to improve a site's link popularity, please refer to [Section 3](#).

**linkrot**

Similar to [dead links](#), but more specifically referring to the general problem of dead links on the web. Linkrot is a major headache for the [search engines](#) who has to return relevant and up-to-date results.

**link swop / link swap**

Similar to [reciprocal links](#), referring to the practice of two or more sites exchanging [links](#) in an effort to boost [link popularity](#). For more on this and other ways to boost link popularity, please refer to the [Section 3](#).

**link tracking**

A type of indexing designed to track inbound links to a document. Many search engines offer ways to easily track inbound links. At Google, for example, simply

type “link:www.your-domain-here.com” (without the quotation marks) for a list of sites linking to www.your-domain-here.com.

### log file

Each web site has a log file (stored on the server), which records details every time a visitor to the site requests a file. Log files store data such as the **IP address** of the visitor, the visitor’s nationality, operating system, **browser** etc. The log file can be analyzed to obtain statistics on **unique visitors**, **page views**, **hits** etc., which are often used as measures in **SEO**.

Also see **log file analysis**.

### log file analysis

Referring to the analysis of records stored in the **log file**. In its raw format, the data in the log files can be hard to read and overwhelming. There are numerous log file analyzers that convert log file data into user-friendly charts and graphs. A good analyzer is generally considered an essential tool in **SEO** because it can show **search engine** statistics such as the number of visitors received from each search engine, the **keywords** each visitors used to find the site, visits by search engine **spiders** etc. For more on log file analysis, please refer to the **Section 4**.

### LookSmart

[www.looksmart.com](http://www.looksmart.com)

A comparatively small **directory**. For a complete review of LookSmart and its **PPC** model, please refer to **Section 1**.

The logo for LookSmart, featuring the word "looksmart" in a bold, lowercase, sans-serif font. The "o"s in "look" are stylized with a dot above them, and the "a" in "smart" has a dot above it as well.

## Lycos

[www.lycos.com](http://www.lycos.com)

Lycos started out as a **search engine** and was very highly rated in the late 90's. Today, web search remains one of its features, but there has been a shift of focus to become a more general **portal** site with features like e-mail, personalization etc. Please refer to **Section 1** for a more detailed look at Lycos, how it works and its importance in **SEO**.



This dictionary is also available as a separate PDF book. Get it **(free)** from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)



## Magellan

MAGELLAN

A discontinued [directory](#). Once listing only the very best of the best web sites, it was considered the “holy grail” of [SEO](#).

## manual submission

The process of manually [submitting](#) a web page to a [search engine](#) or [directory](#) as opposed to using [submission software](#) or a [submission service](#). Manual submission is considered by many to be the only reliable form of submission, although some programs and services have begun distinguishing themselves as viable options. We discuss the two programs worth your money in the [Section 3](#).

## mass submission

A service offered by [submission services](#) whereby a page is [submitted](#) to “thousands of [search engines](#)”. Most [SEO](#) specialists agree that mass submission is not worth the time or money. In truth, there simply are not thousands of search engines. There are about 5 that really matter and another 100-or-so worth knowing about (listed in the [Section 1](#)). The rest of the “1000s” are usually obscure [directories](#) or [FFA](#) pages.



## match

A match occurs when a document in the **search engine's index** contains terms entered as part of the **query**. The matching documents, simply called matches, are then displayed on the **SERP**. It's worth noting that search engines have different criteria for deciding when a document is a match. Most search engines only require that one word in the query match one word in the document. Some search engines (like **Google**), require all words to appear in the document before that document is considered a match.

Also see **begins-with partial word matching** and **Boolean search**

## Metacrawler

[www.metacrawler.com](http://www.metacrawler.com)

A popular **meta search engine**.



## meta refresh

An **HTML** tag that is used to reload or refresh the page after a specified interval, often use to automatically redirect visitors to another page. Most **search engines** penalize pages that use meta refresh or any other type of automatic redirection.

## meta search

A **search** performed on a **meta search engine**. MetaSearch is also the name of a meta search engine found at [www.metasearch.com](http://www.metasearch.com).

### meta search engine

A type of [search engine](#). Meta search engines usually do not maintain databases. Instead, they query other search engines' databases and return results from all of them – usually with a mention of the search engine next to the each result. Refer to [Section 1](#) for more on meta search engines.

### meta tag

An [HTML](#) tag placed in the head section of a web page. The tag provides additional information that is not displayed on the page itself. The initial idea was that webmasters should use these tags to help [search engines](#) index the page correctly by providing an accurate description of the page content and a list of [keywords](#) associated with the page. Unfortunately this left the door open to abuse. Many webmasters used these tags to gain an unfair advantage, forcing search engines to begin disregarding meta tags. For a detailed how-to on meta tags and an updated discussion on their importance (or unimportance) in [SEO](#), please refer to the [Section 3](#).

### Mining Company

Former name of the [About.com](#) web [directory](#).

### mirror sites

Referring to sites that offer authorized duplicates of content also found on other sites. The initial motivation was to ease bandwidth load and increase availability by distributing popular files to many servers. In the context of [SEO](#), the term is mostly used to refer to sites that attempt to deceive [search engines](#) into [indexing](#) more

than one instance of a site by duplicating it on another server and [domain](#). Most search engines now have filters in place to detect mirror sites and many of them penalize these sites by [de-listing](#) both the original site and the mirror site.

### **Mosaic / NCSA Mosaic**

An early web browser developed by the National Center for Supercomputing Applications (NCSA). It was the first cross-platform browser, building on work done by Tim Berners-Lee. Mosaic became the precursor to [Netscape](#).

### **most wanted response (MWR)**

A term coined by Ken Evoy, referring to the aim of a web site, for example, to generate a sale or to get the visitor to subscribe to a newsletter.

### **mousetrapping / circle jerking**

The practice of using scripts to prevent a user from leaving a web site. Typically these involve disabling the back button and the close button or using pop-ups that seem to multiply each time the visitor closes one.

### **Mozilla**

An early, open-source web browser.

### **MWR**

See [most wanted response](#).



### Natural Language Processing (NLP)

A system that allows **search engine** users to type a question rather than keywords. There are a couple of ways to do this kind of processing. At the simplest level, the search engine simply removes the **stop words** in the question to leave **keywords** that are then processed as if it was a regular **query**. At the other end of the scale are very advanced systems that use statistics and linguistic analysis to accurately **match** documents to the user's question. The best-known example of this kind of approach is the **AskJeeves** ([www.askjeeves.com](http://www.askjeeves.com)) search engine.

### Netscape

An early Internet company, since acquired by AOL. The company is famous for its **Netscape Navigator** browser that dominated the browser scene from 1994 to about 1997.

### Netscape Navigator

An early web browser, based on the **Mosaic** model and developed by the Netscape company – as they were then known. The browser is still around today, available from [www.netscape.com](http://www.netscape.com). It's popularity declined rapidly after Microsoft

steamrolled the browser scene (about 1997) by starting to bundle their Internet Explorer browser with Windows.

### **NewHoo**

Former name of **ODP**.

### **newsgroup**

A discussion forum where users can post messages and reply to other users.

### **Northern Light**

[www.northernlight.com](http://www.northernlight.com)

Used to be a popular **search engine**. Although it still has a searchable **database**, it is a “special collection” of articles that only paying customers may access.



This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)



### **obfuscation**

A seldom-used term, more often called [spamdexing](#). It refers to the misrepresentation of [meta tags](#) and page content in order to gain an unfair advantage in the [search engines](#). The term is sometimes differentiated from spamdexing in that it is used to refer to pages that, through [stealth](#), rank highly although they are poorly [optimized](#). The idea is to deliberately mislead others who might steal the page.

### **ODP**

See [Open Directory Project](#)

### **ontology**

In the context of [search engines](#) it refers specifically to a file that defines relationships between words.

Also see [fuzzy matching](#).

## Open Directory Project (ODP)



[dmoz.org](http://dmoz.org)

A massive **directory** continually expanded by volunteers. What sets this directory apart is that it makes its **database** of indexed documents available to other directories & **search engines**. The end result is that a listing here often results in the page automatically being listed in many other directories and search engines. The model of using volunteer editors is fairly ambitious – and surprisingly successful. There are of course certain difficulties like slow processing of **submissions** and occasional dishonesty in the review process, but in the end it is a mammoth achievement and an asset to the online world. Getting a site indexed at ODP can be a daunting task, so we've included comprehensive guidelines and a full review of this directory in the **Section 1**.

## Open Text

[www.opentext.com](http://www.opentext.com)

A fairly large directory listing only business sites.

## operators

“AND”, “NOT” and “OR” as used in **Boolean Searching**.

## optimize / optimization

A page is said to be optimized when it has been structured in such a way that it ranks well (on the **SERP**s) for those terms it targets. It is a fairly subjective concept. What some see as optimization might be termed **spamdexing** by others. In the strictest sense, optimization means simply making a page **spider**-friendly by, for

example, using text **links** rather than image links. In the **SEO** industry the term is more often used as a collective name for all the “tricks” webmasters use to improve a page’s **ranking**.

### **outbound link**

When site A links to site B, site A has an outbound link and site B has an **inbound link**.

### **Overture**

[www.overture.com](http://www.overture.com)

The largest and most popular of the PPC (pay-per-click) search engines. Formerly known as **Goto**. For a more detailed look at Overture, please refer to **Section 1**.





**packet sniffing**

The practice of monitoring pieces of data (called packets) as they move over the Internet.

**page impression**

See [page view](#)

**page jacking / pagejacking**

The act of duplicating a (usually high ranking) web page and presenting the duplicate as the original. This kind of blatant theft is fairly uncommon. In most cases the legitimate author / owner can easily prove ownership of the material.

**page popularity**

See [link popularity](#)

**PageRank**

Google's measure of the [link popularity](#) of a page. [Section 1](#) has more on PageRank.

**page view / page impression / page request**

Often confused with a [hit](#), the term refers to the actual number of pages (not files) viewed by all visitors to a site in a given time period. The number of page views (and other statistics) can be obtained through [log file analysis](#).

**parentheses**

Some [search engines](#) allow users to use parenthesis ( ) to group words. This is especially useful in [Boolean](#) searchers.

**partial word matching**

Some search engines will consider not only exact matches, but also partial matches. This means that if the search term is contained within a word in a document in its index, the search engine considers the document a match. It's not as complicated as it sounds though. If the user enters "word" as the query, the search engine will consider a document a match if it contains word or wordiness or foreword or MSWord etc. So the search term should be contained in the word.

Also see [begins-with partial word matching](#).

**pay per click**

See [PPC](#)

**pay-per-click search engine**

See [PPC search engine](#)

**pay per lead**

See [PPL](#)

**personally identifiable information**

Referring to information collected by a web site that can be used to identify a user. It does not refer to usernames or nicknames, but rather to information like real names, telephone numbers, physical addresses etc.

**phrase search**

A [search](#) for documents containing an entire phrase – as opposed to one or more [keywords](#). The important distinction here is that in a phrase search, the words has to appear side by side in the document (exactly as in the [query](#)) for that document to be considered a [match](#). If the words appear scattered or they appear side by side but in the wrong sequence, it is not considered a match. Phrase searching can be done on most search engines by simply enclosing the phrase in quotation marks.

**placement**

See [positioning](#)

**politeness window**

Most [spiders](#) will not [crawl](#) an entire site in one session. Instead, they crawl a couple of pages and return after a day or two to crawl a couple more and so on until they have [indexed](#) the entire site. This is a self-imposed limit in order not to overburden a server. These gaps between sessions are collectively known as the politeness window. Nice spiders.

**pop-under / popunder / pop under**

A supposedly less annoying variation of the [pop-up](#). It creates a new browser window, usually containing an advertisement that is displayed behind the current window. The user then only sees the pop-under when the current window is closed or minimized. In truth, many users find pop-under as annoying as pop-ups, with the added irritation of feeling tricked into not closing the new window immediately.

**pop-up / popup / pop up**

A new browser window (usually containing an advertisement) automatically opened when the users performs a specified action – like opening a page, clicking a [link](#), closing a page etc.

Also see [pop-under](#).

**portal**

A web site that functions as a kind of starting page or entry point to the web. Portals typically have a wide variety of features such as [search](#), free web-based e-mail, news etc. Well-known examples include [Excite](#) and [Yahoo](#).

**portal page**

See [doorway page](#)

**portal site**

See [portal](#)

**positioning**

Often used as a synonym for [optimization](#).

**PPC**

Pay-Per-Click. An advertising payment model where the advertiser pays only when the advertisement is actually clicked. In other words, the advertiser literally pays only for visitors rather than per advertisement [impression](#). The term PPCs is sometimes used to refer to [PPC search engines](#).

**PPC search engine / PPCSE**

A search engine that uses the [PPC](#) payment model. Advertisers bid on [keywords](#) they wish to target. The search results are then ranked based on the bids with the highest bidder's site ranked first. Advertisers only pay when their [links](#) are clicked – not every time their sites appear in the results. PPCSE marketing has become a fairly important and potentially effective online marketing technique. Please refer to [Section 3](#) for more on effective PPC marketing.

**PPL**

A system where the receiving site pays a certain amount to the referring site for every new [lead](#).

Also see [PPC](#).

**precision**

[Search engines](#) will often consider a document a [match](#) to a [query](#) when that document is not [relevant](#). These mistakes happen because search engines, to a

certain extent, have to “guess” what the user is looking for – especially when words used in the query have double meanings. Search engines must find a balance between **recall** (it’s ability to find all relevant documents) and precision (it’s ability to find only relevant documents). The aim in **information retrieval** is to get both recall and precision spot-on. In other words to return all relevant documents and nothing else. In the real search engine world however, it is often a trade-off. Precision is scored by dividing the total number of pages found by the number of relevant pages found. For example, if 1000 documents are found and 770 are relevant, the search engine’s precision is 0.77 or 77%.

### **precoordination of terms**

The use of compound terms to describe a document. A page about herbal cures for common ailments, for example, could be **indexed** under “herbal remedies”.

### **postcoordination of terms**

The use of 2 or more single words to describe a document. A page about herbal cures for common ailments, for example, could be **indexed** under “herbal”, “cures” and “remedies”. The **search engine** would then consider that document a **match** to a **query** like “alternative remedies”.

### **PR0 / PR zero**

**PageRank** zero. A penalty (rumored to be) imposed by **Google** on sites caught **spamdexing**. It’s worth noting that Google denies having such a penalty.

**probabilistic model**

Referring to any **search engine** model that determines **matches** based on the probability that a document will be **relevant** to a **query**.

**proximity**

See **adjacency**

**proximity search(ing)**

In proximity searching the user can specify a maximum distance between **keywords**. For example, in a **search** for “guns roses” with a maximum distance of 2, documents containing the following are considered matches:

- guns and roses
- guns ‘n roses
- more guns than roses

While these are not:

- ...used guns, but in the next example André used roses
- Guns blazed in the rose garden

Ok, bad example. It’s worth noting that some search engines also let you define the order, so “roses and guns” does not count as a match.



### query

A **keyword**, group of keywords or **phrase**, with or without special instructions like **Boolean operators**, used in a search. In simpler terms, it is that which the user enters into the search box. It is what the **search engine** compares documents to in order to return only **relevant** documents.

### query-by-example / find similar

Many **search engines** have a “find similar” feature that allows users to request documents the search engine considers similar to the document the user specifies.

### query expansion / search within results

The process of basing a new **query** on an old one. Many **search engines** allow users to “**search** within these results”.

This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)





### ranking

Referring to the position of a web page on the **search engine** results for a particular **query**. For example, a page that is listed third for the term “bubblegum” is said to have a ranking of 3 for that term. When used as a verb, the term is synonymous with **optimization**.

### RealNames

An alternative web site address system whereby particular words could be registered and pointed to actual URLs. The system is no longer in use. It relied heavily on support from Microsoft. When Microsoft decided to discontinue their support, the RealNames system simply did not have the reach it needed to work.

### recall

A measure of a **search engine**’s ability to return all **relevant** results. Search engines must find a balance between recall and **precision** (The measure of a search engine’s ability to return only relevant results). If there are 10 pages about “blue bananas” in a search engine’s **database** and a search for “blue bananas” returns only 8 of those pages, the recall is scored at 0.8 or 80%. It’s important to note that recall has nothing to do with database size. If another search engine has only 3

pages about blue bananas and returns all 3, its recall is 100%, even though there are other relevant documents not in its database.

### **reciprocal link**

A [link](#) placed on site A, pointing to site B, on the condition that site B returns the favor. Also called a link swap. Contrary to popular belief, reciprocal linking does not necessarily improve a site's [PageRank](#) and can have a negative effect on PageRank. For a detailed discussion on how and when to swap links as well as getting the most out of PageRank, please refer to the [Section 1](#).

Also see [deep linking](#).

### **redirect**

Users can be redirected from one page to another either by asking them to click on a [link](#) or by means of automatic redirection, most often done with the [meta refresh](#) tag. Automatic redirection has been misused to the point where most [search engines](#) now penalize sites that use it, typically by [de-listing](#) the site.

### **referrer**

When a user follows a [link](#) from page A to page B, page A is called the referrer. The referrer is identified by the URL of the referring page. Referrer information can be accessed through the [log file](#).

### **refresh / refresh tag**

See [meta refresh](#)

**registration**

See [submission](#)

**relevance / relevancy**

The measure of the accuracy of the search results – in other words it's a measure of how close the documents listed in the search results are to what the user was looking for. The ability to return relevant results is a big thing in the [search engine](#) world – and arguably the one thing that made [Google](#) stand out of the crowd and gain much popularity in a short time.

Also see [precision](#) and [recall](#).

**relevancy algorithm**

See [algorithm](#)

**re-submission**

The process of [submitting](#) a web page to a [search engine](#) and then repeating the submission process – either a couple of times or regularly over a period of time. Contrary to popular belief, regular re-submission does not improve a page's [ranking](#) and is considered [spamdexing](#) by most search engines. For more on this and other common [SEO](#) mistakes, please refer to [Section 3](#).

**results list**

See [SERP](#)

**robot**

A browser-like program that automatically request web pages in order to **index** the page content (in the case of **spiders**) or to retrieve specific information (in the case of programs like e-mail harvesters).

**robots.txt / robots text file**

A text file (with the “.txt” extension) that tells **spiders** which pages it may not **index**. Every time a spider (that complies with the Robots Exclusion Standard) visits a site it will first request a robots.txt file to see where in the site it is not allowed to go. The syntax and correct placing of the robots.txt file as well as an alternative way to declare pages “off-limits” is discussed in **Section 3**.

**ROI**

Return On Investment. In the context of **SEO**, the term refers to sales generated as the direct result of a **search engine** marketing campaign.



### **Scooter**

The name of [AltaVista's spider](#). (The name refers to the annual motorcycle races held at the famous AltaVista Raceway)

### **score**

[Search engines](#) usually order search results from the most [relevant](#) to the least relevant (as determined by the search engine's [algorithm](#)). In order to [rank](#) documents, the search engine assigns a score to each page and those with the highest scores are listed first. Most search engines simply give the maximum score to the most relevant document and score all other relevant documents relative to the perfect document. Others compare all documents to a theoretically perfect document. The score of a web page therefore refers to its relevance as perceived by a specific search engine.

### **script**

A piece of programming designed to perform a certain function on a web page – for example to create a rollover effect on buttons or to create [pop-ups](#).

**search**

The process of locating information – on the Internet typically done by searching through documents in [search engine](#) and [directory databases](#).

**search engine**

A tool for finding information on the Internet. Most search engines consist of the following main components:

1. [Spider](#)
2. Indexer
3. [Database](#)
4. Search software
5. Web interface

Documents found by the spider are processed by the indexer and stored in a database. From the database the search software extracts documents based on parameters entered by the user. Examples of search engines include [Google](#) and [AlltheWeb](#). [Directories](#) like [Yahoo](#) and [ODP](#) are often referred to as search engines although they are not. For more on how search engines work, please refer to [Section 1](#).

**search engine marketing**

See [SEO](#)

**search engine optimization**

See [SEO](#)

## search engine positioning

See [SEO](#)

## search hours

The actual amount of time (in hours) all visitors to a [search engine](#) spent there during a given month. [Audience reach](#) and search hours are the two major factors when calculating the popularity of a search engine.

## SearchKing

<http://www.searchking.com>

A comparatively small [search engine](#). It's claim to fame is that it allows users to vote on the [relevance](#) of documents it returns for [queries](#) – and it then uses that data to continually increase the accuracy of the results. In September 2002 SearchKing was (according to them) penalized by [Google](#). The rumor has it that sites that link to SearchKing were also penalized and we decided to disable the [link](#) above. You can still visit the SearchKing site by typing <http://www.searchking.com> into the address bar of your browser.



## search results

The documents returned by a [search engine](#) in response to a [query](#). Also see [SERP](#).

## search term(s)

Words entered into a [search engine](#)'s search box to form a [query](#).

**search tree**

A seldom-used synonym for a searchable **directory**.

**SEO**

Search Engine Optimization. This term is widely used in the **search engine** industry as a collective name for those activities that are directly or indirectly aimed at improving a page's search engine **ranking**. Sometimes the term SEO is also used to refer to providers of SEO services – in other words it's used in the place of terms like “SEO provider” and “SEO specialist”. For a detailed discussion of the SEO industry and SEO techniques, please refer to **Section 3**.

**SERP(S)**

Search Engine Results Pages(s). The term refers to the page listing **search results**.

**Sidewinder**

The name of **Infoseek's spider**.

**similarity**

Similar to the idea of **relevance**, similarity is the measure of the degree to which a document **matches** a **query**.

**siphoning**

A collective name for the different techniques used to steal **traffic** from another site. For example the use of another's trade name in the **title** tag etc. Also see **obfuscation** and **spamdexing**.



**site hit**

See [hit](#).

**site search**

A [search](#) utility that allows the user to search through documents on a particular site. Different from a [search engine](#) in that its [database](#) contains only documents found on that site as opposed to a wider collection of documents from all over the web.

**skewing**

A technique used by the [search engines](#). It refers to the practice of artificially altering the [search results](#) so that certain documents will [score](#) well on certain [queries](#).

**Slurp**

[Inktomi](#)'s spider.

**Sniffer**

The name of a program that [Infoseek](#) used to “sniff out” attempts at [spamdexing](#).

**sorting results**

[Search engines](#) sort results displayed on the [SERP](#) in a particular order – usually from most [relevant](#) to least relevant. Some search engines allow the user to sort results based on different criteria, for example alphabetically, arranged from newest to oldest etc.

**spam**

A collective name for those marketing techniques that are intrusive, offensive and/or unethical in some way. A major characteristic is that it aims its message at a wide (often in the millions), untargeted audience – which it can afford because electronic distribution is very cheap. The most common form of spam is unsolicited commercial e-mail. In the **search engine** world, regular **mass submission** of web pages to search engines is also referred to as spam or **spamdexing**. Spamdexing is often used to refer to all **SEO** techniques that are deceptive or unethical.

**spamdexing**

All attempts to deceive **search engines** or gain an unfair advantage in the **search results** of a search engine. Spamdexing decreases the value of a search engine's **index** by reducing the accuracy with which the search engine can return **relevant** documents. Most search engines have measures in place to detect spamdexing and guilty pages are usually either penalized or **de-listed**. Many webmasters inadvertently make themselves guilty by braking search engine **submission rules**.

**spamming**

See **spam**, **spamdexing**

**spider, spyder**

A browser-like program that forms part of a **search engine**. Its task is to “surf” the web by following **links** from one page to the next and from one site to the next. It collects information from the sites it visits and that information is stored in the

search engine's **database**. For detailed discussions on spiders, the other components of search engines, spider names etc., please refer to **Section 1**.

### **spidering**

What **spiders** do – the process of surfing the web and **indexing** documents.

### **splash page**

A page that is displayed before users enter a site. Splash pages are often comparatively empty except for a logo, welcome message and “click here to enter” type of **link**. Splash pages are often used to house introductory **Flash** animations. Splash pages are generally considered annoying since they offer very little value. Even very impressive splash pages offer only entertainment – which distracts from the sales effort and hampers **SEO**.

### **spoofing**

See **IP spoofing**, **spamdexing**

### **SSI (Server Side Include)**

A type of **HTML** command that allows webmasters to insert code from an outside HTML document. It is especially used with things like menus, headers and footers that are the same for all pages. To change the menu, for example, the webmaster changes only the external menu file and the menu changes across the entire site. SSI can also be used to insert non-HTML elements like **scripts**.

**stealth**

A collective name for techniques (like cloaking) that aim to deliver **optimized** content to **spiders** while delivering the “real” page to human visitors. Almost all **search engines** consider stealth a form of **spamdexing**.

**stemming**

The use of linguistic analysis to get to the root form of a word. **Search engines** that use stemming compare the root forms of the **search terms** to the documents in its **database**. For example, if the user enters “viewer” as the **query**, the search engine reduces the word to its root (“view”) and returns all documents containing the root – like documents containing view, viewer, viewing, preview, review etc.

**stop word(s)**

Words like conjunctions, prepositions etc. that are so commonly used that they have little or no influence on **relevancy**. Most **search engines** ignore stop words entered in a **query**.

**sub-categories**

**Directories** are typically divided into top-level **categories** that contain sub-categories or lower level categories. Directories often run several category levels deep.

**submission**

The process of manually adding a URL to a **search engine**’s list of URLs to **spider** – in effect telling a spider about a page in order to get it spidered and ultimately added to the search engine’s **database**.

**submission rules**

Most [search engines](#) have a list of rules that must be obeyed when [submitting](#) sites to be [spidered](#). Examples of submission rules include how often the page may be resubmitted (if at all), how many pages may be submitted per day etc.

**submission service**

Services exist where the user can have pages [submitted](#) to multiple [search engines](#) for a fee. The fee is normally very low, but usually not as low as the quality of the submission. We have a more detailed explanation of submission services and the dangers, as well as guidelines to choosing a reputable [SEO](#) service in [Section 5](#).

**submission software**

Programs that assist webmasters in [optimizing](#) and [submitting](#) web pages to [search engines](#). There are countless programs available, but probably only a handful that are worth getting. You can find full reviews of the top 2 programs in our [Section 3](#).

**submit**

See [submission](#)

**substring matching**

See [partial word matching](#)



### taxonomy

A set of agreed-upon principles according to which information can more logically be stored in an **information retrieval** system. The term is used in science to describe the classification of natural elements.

### Teoma

[www.teoma.com](http://www.teoma.com)

A fairly new **search engine** (compared to oldies like AltaVista).



### term frequency (TF)

A measure of how often a term is found in a collection of documents. TF is combined with **inverse document frequency** (IDF) as a means of determining which documents are most **relevant** to a **query**. TF is sometimes also used to measure how often a word appears in a specific document.

### theme engine

A **search engine** that attempts to automatically classify sites based on the **keywords** they contain.

**thesaurus**

Similar to a dictionary, but containing lists of synonyms rather than definitions. Some **search engines** use a thesaurus in addition to things like **stemming** and **fuzzy matching** in an effort to improve **recall**.

**title**

The title of a page is displayed in the title bar right at the top of the browser window. Almost all **search engines** consider the title when determining a document's **relevance** to a **query** and most search engines consider the title the most important element. In the page, the title is specified as an **HTML** element and placed in the header section of the page.

**TLD**

Top Level Domain. See **domain**.

**toolbar**

With reference to **search engines**, toolbars are browser add-ons provided by the engines. These toolbars often include a **search** box, shortcuts to the different sections of the search engine, additional page information etc.

**traffic**

Often used as a synonym for “visitors”. The term is used to describe activity on a web site – be it **hits**, **page views** or actual visits.

**T-Rex**

The name of the [Lycos spider](#).

**Turbo10**

[www.turbo10.com](http://www.turbo10.com)

A type of [meta search engine](#) that searches both the surface-web (normal documents) and the invisible web or, as they call it, the DeepNet (documents normally not indexed by search engines).

The logo for Turbo10, featuring the word "Turbo" in a stylized, italicized font with a yellow-to-orange gradient, and the number "10" in a bold, black font with a yellow-to-orange gradient, all set against a dark purple rectangular background.

This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)



**unique visitor**

Used to describe one person visiting a site. That one person may generate multiple visits over a period of time, therefore **log files** normally show more visits than unique visitors. The shortened version “uniques” is sometimes used to refer to unique visitors.

**uniques**

Short for **unique visitors**.

**unique user**

See **unique visitor**

**upload**

The process of transferring information from a local drive to a server – specifically when that information then becomes accessible via the Internet.

**URL**

Uniform Resource Locator / Universal Resource Locator. A unique Internet address (for example <http://www.pandecta.com>) that every Internet resource must have in order to be located.

**URL submission**

See [submission](#)

**vertical portal**

See [vortal](#)

**virtual domain**

A [domain](#) that is hosted on a [virtual server](#). The domain is unique, but the [IP address](#) is normally shared with other domains. This has some implications for [SEO](#). Please refer to the [Section 3](#) for a more detailed discussion of the importance of having a unique IP address.

**virtual server**

When a [domain](#) is hosted on a virtual server, it means that it shares that server with other domains. This is a very cost effective way of hosting web sites, but access speeds are not as high as for domains hosted on dedicated servers.

Also see [virtual domain](#).

**visitor**

The term is sometimes confused with [unique visitors](#). The difference is that one unique visitor visiting a site repeatedly over a period of time will show up on the

site's **log file** as many visitors. The term therefore refers to the number of times people visit a site – not the actual number of people visiting a site.

**vortal**

The term is used to describe portals that focus on one specific (vertical) topic. In other words, they target at a specific group of people – like programmers, SEO specialists etc. – by providing in-depth information on that topic.

This dictionary is also available as a separate PDF book. Get it (**free**) from [www.searchenginedictionary.com](http://www.searchenginedictionary.com)



### Wayback Machine

[web.archive.org/](http://web.archive.org/)

A very large “archive” of the web. The Wayback Machine stores “snapshots of sites”, allowing users to have a look at how sites looked “wayback” then.



### web copywriting

Copywriting specifically aimed at an online audience. It shares many of the ground rules of offline copywriting, but has quickly evolved to become a stand-alone science. Recently it has also begun taking into account how **spiders** see web pages. Although there are many who feel copywriters should focus on converting visitors to customers and not be concerned with getting visitors, there are strong arguments for **SEO** considerations to form part of web copywriting.

### Webcrawler

[www.webcrawler.com](http://www.webcrawler.com)

A fairly old **meta search engine**.



### weighting

Describing the technique **search engines** use to compare the **relevance** of different documents to a **query**. Search engines effectively “weigh” different pages based on things like the occurrence of **keywords** in the **title** in order to list documents in order from most to least relevant.

Also see **score**.

### WHOIS

A type of **search** where the **query** is a **domain** name and the result shows details of the domain, like when it was registered, by whom, when it expires etc.

### Wisnut

[www.wisenut.com](http://www.wisenut.com)

A fairly large **search engine**. Wisnut was at one stage (about 2001) considered a credible threat to **Google**'s dominance, but has failed to deliver on that early promise. Refer to **Section 1** for a more detailed look at WiseNut.



### word stuffing

See **keyword stuffing**

**Xenu**

A widely used [link-checking](#) program.

**XML**

Extensible Markup Language. A web programming language that allows web authors to define their own, custom tags. Especially useful in the creation of web-based applications.



## Yahoo!

[www.yahoo.com](http://www.yahoo.com)

One of the first and most-loved web directories, Yahoo is presently (2002) believed to be the most visited site on the Internet.



## zones

Some [search engines](#) allow users to limit a [search](#) to specific zones – better described as topic areas. A user may, for example, elect to search only documents from a certain geographic area or only documents created within a specific timeframe.

Also see [advanced search](#).



## Section 7: General Information



# General Information



## 7.1 About SEY 2004 And Your 25% Discount

□

**Only in the full version:□**

□

□

**Sorry, the 25% discount on SEY 2004 is only for owners of SEY 2003.□**

□

□

□

□

□

□

□

□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**



## 7.2 How To Earn A Free Copy Of SEY 2004

□

**Only in the full version:□**

□

□

**Sorry, this special offer is only for owners of SEY 2003.□**

□

□

□

□

□

□

□

□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**



## 7.3 Priority Customer Support

□

**Only in the full version:□**

□

□

**Sorry, priority customer support is reserved for owners of SEY 2003.□**

□

□

□

□

□

□

□

□

□

□

**Click anywhere in this block to order your full version of the Search Engine Yearbook. It comes with an unconditional money-back guarantee, so it's a completely risk-free purchase. <http://www.pandecta.com/sey.html>**

## 7.4 About The Author



André le Roux founded [Pandecta Magazine](#) in 1999. It's an online publication about "the real-world nitty-gritty of making money online".

He first began researching search engines in 1997. What started as a hobby quickly turned into a full time passion.

In 2000 he published the first of the "Mother of all Search Engine Reference Books" series. The "mother"-books continue today, but in a slightly different guise. They are now scaled-down versions of the Search Engine Yearbooks – and are still given away for free from the [Pandecta web site](#).

Previous occupations include teacher, fine art lecturer and webmaster for a large, South African insurance company.

## 7.5 About Pandecta Magazine

**Pandecta Magazine** started out (1999) as a nitty-gritty e-business guide for Internet entrepreneurs. Over the years we've added projects – the Search Engine Yearbook series being the most ambitious and most successful so far.

For now, publication of the magazine itself has been halted. To be honest, we're learning so much about e-commerce – every day – that I started feeling uncomfortable advising entrepreneurs when we are clearly not as clued-up as we initially thought.

So right now, over at Pandecta Magazine, we're playing with different ways of making money on the Net, learning as we go. Fortunately we have a couple of past experiments delivering a steady income stream to fund new experiments. ☺

### Some URLs:

Pandecta Magazine: <http://www.pandecta.com>  
Electronic Light: <http://www.electroniclight.com> (current experiment)  
ChairBay: <http://www.chairbay.com> (current experiment)  
Search Engine Dictionary: <http://www.searchenginedictionary.com>

**Contact:** [inbox@pandecta.com](mailto:inbox@pandecta.com)



All logos are copyrights and trademarks of their respective owners. None of these owners has authorized, sponsored, endorsed or approved this publication. Screenshots in this book are directly from publicly accessible file archives. They are used as "fair use" under 17 U.S.C. Section 107 for news reportage purposes only, to illustrate various points made in the book. Text and images over the Internet may be subject to copyright and intellectual property rights owned by third parties.



ÓCOPYRIGHT 2003, [Pandecta Magazine](#)Ô. All rights reserved worldwide.

This free version of the Search Engine Yearbook 2003 may be freely redistributed, on the condition that it is not sold and not changed in any way. You may also electronically redistribute this free e-book. All copyright correspondence should be sent to [legaldesk@pandecta.com](mailto:legaldesk@pandecta.com). Pandecta Magazine, Search Engine Yearbook and EnginePaper are trademarks of Pandecta Magazine. All other graphics / trade names / logos displayed are trademarks or registered trademarks of their respective owners.

#### DISCLAIMER

Although the greatest care have been taken to ensure the accuracy of information in this document, Pandecta Magazine, the author, associated companies, associated individuals and contributors accept no responsibility for direct or indirect damage or loss of any kind suffered as a result of reliance upon information contained in this document or any document / information referred to in this document. Links to the World Wide Web, both in the case of links to regular web pages and links to affiliates of Pandecta Magazine, do not constitute endorsement of any web site or product. Readers are encouraged to investigate all offers carefully.

Pandecta Magazine offers no warranties of any kind on this free document, whether express or implied.

*Thanks for supporting this publication ;-)  
If you have comments or questions, I would love  
to hear from you. [andre@pandecta.com](mailto:andre@pandecta.com)*