

# Search Engine Dictionary™

version 1.0 335 terms

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

Published & distributed by [Pandecta Magazine](#)



*The Most Complete Guide to Search Engine Terminology in the World – by far.*

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

## Important links to the web:

SED download page: <http://pandecta.com/sed.html>

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

Black = normal text

Green and Red = highlight colors

Blue = Links to the web

Orange = Internal links (to other sections in the book)

If you have ideas, suggestions or corrections for future versions of SED, please [tell us](#).

If you have a question, please contact [Pandecta support](#).



# 1. Foreword

([CLICK HERE TO JUMP STRAIGHT TO THE DICTIONARY](#))

## The Most Complete Search Engine Dictionary

Calling this glossary “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. We also added many general web marketing terms that are often used in the context of search engines.

We are confident that this is **the most complete** glossary of search engine terms ever created.

## Continued Research

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

**But we’ve thought of that...**

On our 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**)

We 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 us. **If we 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 this book.

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

[Click here to suggest a better definition for an existing term](#)

## Update Cycle

Every **January** the entire [SearchEngineDictionary.com](http://SearchEngineDictionary.com) web site is compiled into a new Search Engine Dictionary – just like this book 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 us 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

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

...by redistributing this book freely.

**Yes, really.**

Give this book 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.

## About The Search Engine Yearbook

You'll find mentions of the **Search Engine Yearbook** throughout this book. It's a fantastic and really thorough look at **EVERY** aspect of the search engine world. More about the Search Engine Yearbook later in this book.

**(click here to go there now)**

---

A special thanks to Charnell Grobler of [DotLens.com](http://DotLens.com) for proofreading this e-book. If you ever need contact lenses, sunglasses – or **anything** else that has a lens in it, chances are that the DotLens site has it.

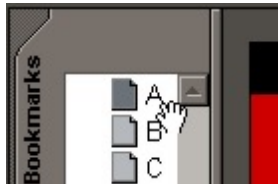


## 2. How To Use This Book

**You're welcome to print this entire book**, but consider reading it on-screen. We've added many great features like **bookmarks**, **internal links** and **links to the web** that makes the on-screen format fast & easy to use.

Here's how:

### Bookmarks



See the “**Bookmarks**” over to the left of your screen?

Simply click on any letter to jump straight to that letter in the dictionary. I told you this was easy. ☺

### Internal Links

If you see orange text, it designates an internal link – in other words a link that takes you to somewhere else within the book. We use internal links extensively in the definitions.

Like so:

## **bridge page / bridging page**

See [doorway page](#).

An entry like this means that a “bridge page” and a “doorway page” is exactly the same thing. Instead of giving you the same definition over and over, we just link to the variation most commonly used.

So in the above example you’d simply click on “doorway page” to find out what it is.

**Here’s another one:**

## **Googlewhacking**

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.

In this definition of the term “Googlewhacking”, there are a couple of words in orange. That means they’re also explained in this dictionary. Simply click on one to jump to that word in the dictionary

This just keeps getting easier doesn’t it? ☺

Let’s look at **external links**...

## External Links

These are links (**the blue ones**) to pages on the web. Depending on your Acrobat Reader settings, clicking on a blue link will either open the site right here in Acrobat Reader or in your web browser.

Here's how they look:

### **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.

In this definition of the term “directory” we included 2 examples of directories for you to look at.

Clicking on either will

... you guessed it ...

**open** those sites for you.

That's it.

That was the quick introduction.

**If you're new to Acrobat Reader, we have a couple of useful hints that'll help you navigate this book (and other Acrobat files) like a pro.**

We've placed that little tutorial on the Pandecta web site.

Here it is:

<http://www.pandecta.com/acrohelp.html>

If you really get stuck using the book and need a **human answer** to your question, drop us a line. We can't promise that we will get to **every** question, but we do promise to try.

Send us an e-mail:

[sed-help@pandecta.com](mailto:sed-help@pandecta.com)





## 3. Table of Contents

(Click a page number to jump to that section of the book)

1. **PAGE 2** Foreword
2. **PAGE 5** How To Use This Book
3. **PAGE 9** Table of Contents
4. **PAGE 10** **The Dictionary**
5. **PAGE 111** Contact Information
6. **PAGE 112** About The Search Engine Yearbook
7. **PAGE 113** More Free Stuff From Pandecta Magazine
8. **PAGE 114** Copyright Notice & Disclaimer

## 4. The Search Engine Dictionary



### 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 **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 the **Search Engine Yearbook**.

### **Interested in learning more about search engines?**

Pandecta Magazine’s Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It’s NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

## 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>d</sup> after Google and **Wisenut**. 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 the **Search Engine Yearbook**.

### AltaVista

[www.altavista.com](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 the **Search Engine Yearbook**.

## **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

[www.askjeeves.com](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.



## 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 the **Search Engine Yearbook**.

### Interested in learning more about search engines?

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



### **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](#). For more on this technique (and why you shouldn't bother), please refer to the [Search Engine Yearbook](#).

### **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 a more detailed comparison between different Internet advertising techniques, please refer to the [Search Engine Yearbook](#).

### **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**.

### Interested in learning more about search engines?

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



## **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% ( $10 / 1000 \times 100\%$ ).

## **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 the **Search Engine Yearbook**.

## **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 the **Search Engine Yearbook**.

### Interested in learning more about search engines?

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

## 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 the **Search Engine Yearbook**.

### **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](#). We have a more detailed discussion of the SEO uses of CSS in our [Search Engine Yearbook](#).

### **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”.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

**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**. The **Search Engine Yearbook** 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 **SERPs**. 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 the **Search Engine Yearbook**.

## 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

[www.dogpile.com](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. The **Search Engine Yearbook** 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 **entry page**, the 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 **Search Engine Yearbook**.

### **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 the **Search Engine Yearbook**.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

### **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. **SERP**s 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



[www.excite.com](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 the **Search Engine Yearbook**.

## 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.



### **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 the **Search Engine Yearbook**.

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 **Search Engine Yearbook**.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

**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.

### Interested in learning more about search engines?

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



### **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 the **Search Engine Yearbook**.

## Googlebot / Google Bot

**Google's spider**.

## Googlewhacking

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 the **Northern Light search engine**.



### 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**. For more SEO techniques and the complete do's and don'ts of SEO, please refer to the **Search Engine Yearbook**.

### 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 [Search Engine Yearbook](#).

## 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 study by Pandecta Magazine (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 Pandecta Magazine's comparative study, a more detailed look at Hotbot and all the other search engines worth knowing about, please refer to the [Search Engine Yearbook](#).

### 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.

#### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

## **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 the **Search Engine Yearbook**.  
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 it's importance in SEO, please refer to the Search Engine Yearbook.



### **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 the **Search Engine Yearbook**.

### **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 the **Search Engine Yearbook**.

### **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**.

#### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



## 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 the **Search Engine Yearbook**.

### **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 the **Search Engine Yearbook** 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 **Search Engine Yearbook**.

**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 **Search Engine Yearbook**.

## **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.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

**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 the **Search Engine Yearbook**.

**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 **Search Engine Yearbook**.

### **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 and analyzers, please refer to the **Search Engine Yearbook**.

### 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 the **Search Engine Yearbook**.

### 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 the **Search Engine Yearbook** for a more detailed look at Lycos, how it works and its importance in **SEO**.





## 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 **Search Engine Yearbook**.

## 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 [Search Engine Yearbook](#)). 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. The **Search Engine Yearbook** discusses meta search engines in more detail and lists some of the more popular ones.

### **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 **Search Engine Yearbook**.

## **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**.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



### **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.



### Interested in learning more about search engines?

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



### **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**

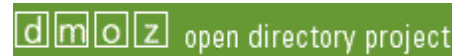
See **Open Directory Project**

### **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 **Search Engine Yearbook**.

## 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 an in-depth look at Overture and different PPC strategies, please refer to the **Search Engine Yearbook**.



**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.

**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.

**Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

**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. We take a look at some of the important PPC search engines (like **Overture**) and reveal some top PPC strategies in the **Search Engine Yearbook**.

### **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”.



### **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](#). In most cases it has 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 [Search Engine Yearbook](#). 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](#)

### Interested in learning more about search engines?

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

## 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 **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 the **Search Engine Yearbook**.

### **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 the **Search Engine Yearbook**.

### **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.

#### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.





### **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. The details of how search engines work are beyond the scope of this book but discussed in more detail in the **Search Engine Yearbook**.

### **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 the [Search Engine Yearbook](#).

## **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**.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

**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 it's **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**. For a detailed discussion of what to do and what not to do, please refer to the **Search Engine Yearbook**.

## **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 the **Search Engine Yearbook**.

**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. For links to the submission rules pages of the major search engines, please refer to the **Search Engine Yearbook**.

**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 our [Search Engine Yearbook](#).

**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 [Search Engine Yearbook](#).

**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. For a details on what **spiders** are looking for when **indexing** pages and the varying importance of different elements, please refer to the **Search Engine Yearbook**.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.

## 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)

The logo for Turbo10, featuring the word "Turbo" in a stylized, italicized font with a blue-to-white gradient, and the number "10" in a bold, black font with a blue outline.

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).

**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](#)

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



**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 [Search Engine Yearbook](#) for a more detailed discussion on 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.



## 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.

## Wisenut

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



A fairly large **search engine**. Wisenut was at one stage (about 2001) considered a credible threat to **Google's** dominance, but has failed to deliver on that early promise.

## 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.

### **Interested in learning more about search engines?**

Pandecta Magazine's Search Engine Yearbook is arguably the most complete guide to search engines available on the web. It's NOT for everyone

– but then –

**not everyone can win in the search engine game.**

Visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.



**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**.



## 5. Contact Information

### DOWNLOADS & UPDATES

Updates of this book are regularly made available through the Search Engine Dictionary web site. To download your **free** copy of the current version, visit [www.searchenginedictionary.com](http://www.searchenginedictionary.com).

### SUPPORT

Please direct support inquiries to [sed-help@pandecta.com](mailto:sed-help@pandecta.com). SED-related support inquiries directed elsewhere can unfortunately not be processed.

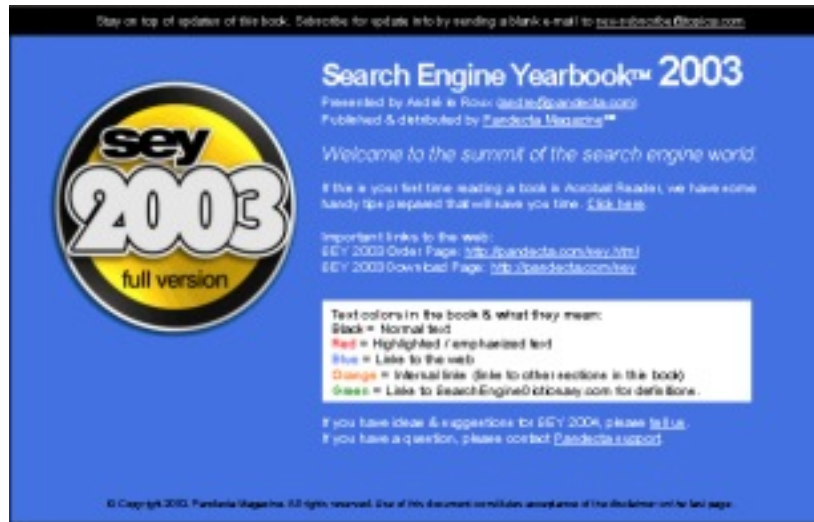
### COPYRIGHT

No need to ask permission... **You may freely redistribute this book** on the condition that it is **not changed**, that it is **not sold** and that you redistribute the **latest version only**. If you have additional concerns or wish to report a violation of these terms, please send us an e-mail: [legaldesk@pandecta.com](mailto:legaldesk@pandecta.com).

### PANDECTA MAGAZINE

Web: [www.pandecta.com](http://www.pandecta.com)  
e-mail: [inbox@pandecta.com](mailto:inbox@pandecta.com)  
André le Roux: [andre@pandecta.com](mailto:andre@pandecta.com)

## 6. About The Search Engine Yearbook



EVERYTHING you wanted to know about search engines – in **one** place and from search engine professionals.

This e-book carries our **unconditional 1-year guarantee**. Buy a copy and decide if it really delivers what **you** need. If you are not **convinced** that it's worth far **more** than you paid, please feel free to claim your **immediate, full refund**.

### Some reader feedback (on SEY 2002):

*"...I will definitely be recommending you to my friends and colleagues in this field... ...Thank you for opening a new door to me."* Steve Haire

*"A great resource for webmasters who do a lot of the work themselves and want to learn more."* Larry Sullivan

*"It's definitely worth the money"* Druggan Svetic

Click [here](#) to view the complete Table of Contents of SEY 2003.

Click [here](#) to order your copy of the Search Engine Yearbook 2003 (reduced price for SED readers).

Or visit [www.searchengineyearbook.com](http://www.searchengineyearbook.com) for more information.





## 7. More Free Stuff From Pandecta Magazine

### Another **free** book

It's called the **"Mother of all Search Engine Reference Books"** and is essentially a scaled down version of our flagship product, the **"Search Engine Yearbook"**. **Get your free copy of "Mother" [here](#).**

### Regularly updated search engine info

Subscribe to our popular **"EnginePaper" Newsletter**. All the search engine news you need. When you need it. And **NOTHING** else.

**Subscribe by sending a blank e-mail to:**

[send-ep-subscribe@topica.com](mailto:send-ep-subscribe@topica.com)

You may also like our **"ElectronicLight" Newsletter** where we tell you what we learn about making money online – as we learn it.

**Subscribe by sending a blank e-mail to:**

[electronic\\_light-subscribe@topica.com](mailto:electronic_light-subscribe@topica.com)

REMEMBER TO VISIT THE PANDECTA SITE FOR **FREE** TUTORIALS  
ON SEO AND E-BUSINESS. [www.pandecta.com](http://www.pandecta.com)

## 8. Ó Copyright Notice & Disclaimer

All logos are copyrights and trademarks of their respective owners and are used as "fair use" under 17 U.S.C. Section 107 for news reportage purposes only. None of these owners has authorized, sponsored, endorsed or approved this publication.

**ÓCOPYRIGHT 2003, Pandecta MagazineÔ. All rights reserved.**

This document may be freely redistributed provided that it is in no way changed, that it is not sold and that only the most current version is redistributed. Failure to comply with these provisions may result in termination of copyright permission and/or legal action.

All copyright correspondence should be sent to [legaldesk@pandecta.com](mailto:legaldesk@pandecta.com).

Pandecta, Search Engine Dictionary, 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, its owners and associated companies / associated individuals / 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 regarding this product, whether express or implied.

**Thanks for supporting this publication**

