

# Website Findability

How to Get Traffic from Google  
and Other Search Engines

**By Michael Heraghty – Version 3.0**

<b>1. Findability.....</b>	<b>5</b>
1.1 Introduction.....	5
1.2 Google and Other Search Engines .....	6
1.2.1 Google.....	6
1.2.2 Other Major Search Engines .....	8
1.2.3 Paid Inclusion.....	8
1.3 PageRank .....	9
1.3.1 PageRank on The Google Toolbar.....	10
1.3.2 “Real” PageRank .....	11
1.3.3 “Real” PageRank Calculation .....	12
1.4 Search Engine Optimization .....	14
1.4.1 The SEO Industry .....	14
<b>2. Planning Your Website.....</b>	<b>15</b>
2.1 Aims and Objectives .....	15
2.2 Site Structure & Integration .....	16
2.2.1 Site Hierarchy .....	16
2.2.2 Getting Listed: How Google Bots Find Your Site.....	16
2.2.3 Creating Non-Searchable Areas on Your Site .....	17
2.2.4 Robots.txt.....	17
2.2.5 Dynamic Pages.....	18
2.2.6 Session IDs.....	18
2.2.7 Multiple Domain Names.....	19
2.3 Site Content.....	20
2.3.1 Think Like a Searcher .....	20
2.3.2 Develop Unique and Detailed Content .....	20
2.3.3 Manage and Update Your Content .....	21
2.4 Keyphrase Research.....	22
2.4.1 Keyphrases.....	22
2.4.2 Keyphrase Suggestion Tools.....	24
2.4.3 Common Words and the Importance of Word Proximity.....	25
2.4.4 Assessing the Competition.....	26
2.5 The Google Sandbox Effect.....	27
2.5.1 Understanding the Sandbox Effect .....	28
2.5.2 Planning for the Sandbox Effect.....	28
<b>3. Designing Your Website.....</b>	<b>30</b>
3.1 Committing to Usability .....	30
3.2 CSS .....	31
3.3. Menu Structure & Style .....	32
3.4 Site Maps .....	32
3.5 URLs and File Names.....	34
3.5.1 Choosing a Domain Name .....	34
3.5.2 File and Directory Naming Conventions .....	35
3.6 Don’t Use Frames .....	36
<b>4. Optimizing Individual Pages.....</b>	<b>37</b>
4.1 Page Titles.....	37
4.1.1 Include the Relevant Keyphrase in the Page Title .....	37
4.1.2 Keep it Short .....	38
4.1.3 Grab the Searcher’s Attention.....	38
4.1.4 Avoid Keyword Lists.....	38

4.1.5 One Keyphrase per Page Title .....	39
4.2 Meta Tags.....	40
4.2.1 Meta Description.....	40
4.2.2 Meta Keywords .....	41
4.3 Page Structure & Layout.....	42
4.3.1 Heading Tags & Keywords.....	42
4.3.2 Page Layout .....	43
4.3.3 Table-based Layouts vs. CSS Layouts.....	43
4.4 Keyword Formats & Density .....	44
4.4.1 Formatting of Keyphrase Within the Page Content.....	44
4.4.2 Outgoing Links .....	44
4.4.3 Keyphrase Density .....	45
4.5 The Art of Writing Optimized Text.....	46
4.5.1 Writing for the Web .....	46
4.5.2 Integrating Keyphrases into Your Writing .....	47
4.6 ALT and TITLE Attributes.....	49
4.6.1 ALT Attribute .....	49
4.6.2 ALT Text Combined with Anchor Tag .....	49
4.6.3 ALT Attribute and Image Search.....	49
4.6.4 The Title Attribute .....	49
4.7 Graphics vs. Text .....	51
4.8 Flash And Other Multimedia Elements .....	52
4.9 JavaScript and Other Client-Side Languages .....	53
<b>5. Developing Links with Other Sites.....</b>	<b>54</b>
5.1 Why Your Site Needs Links .....	54
5.1.1 Links from Pages with Higher PageRank.....	54
5.1.2 In-Site Links: The Pyramid Structure.....	56
5.1.3 Note on Relative Links vs. Absolute Links .....	56
5.1.4 Avoid Superstar Pages .....	56
5.2 Good Links, Bad Links .....	58
5.2.1 Relevant Links .....	58
5.2.2 Links from Expert and Hub Sites.....	58
5.2.3 Links Beyond 100 .....	59
5.2.4 Sponsored Links.....	59
5.2.3 Reciprocal Links vs. One-way Links.....	60
5.2.4 Cross-Linking Among Sites.....	60
5.2.5 Links from Links Pages .....	61
5.3 Anchor Text .....	63
5.3.1 Anchor Text and Keyphrases.....	63
5.3.2 Anchor Text Within Your Site.....	64
5.3.3 Anchor Text Variation .....	64
5.4 How to Get Links.....	65
5.4.1 Link Exchange Request .....	65
5.4.2 Directories.....	65
5.4.3 The Open Directory Project.....	67
5.4.4 Contributing to Other Sites .....	67
5.4.5 Responding to Link Suggestion Requests.....	68
5.4.6 Linkware .....	69
5.4.7 Blog Comments .....	69
5.4.8 A Note About Guestbooks .....	70

<b>6. Maintaining and Increasing Findability .....</b>	<b>71</b>
6.1 Site Statistics .....	72
6.1.1 Hit Counters Don't Count!.....	72
6.1.2 Site Statistics Software .....	72
6.2 Monitoring Keyphrase Performance.....	76
6.2.1 Listings in Flux .....	76
6.2.2 Spreadsheet Solutions .....	76
6.2.3 Monitoring Keyphrases in Your Site Statistics.....	77
6.2.4 Monitoring Competitor Backlinks .....	78
6.3 Growing Your Site.....	79
6.3.1 Adding Content.....	79
6.3.2 Experimenting with Content and Keywords.....	79
6.3.3. Content Management Systems.....	80
6.3.4 Temporarily Popular Search Queries.....	80
6.4 Search Engine Marketing and Paid-For Listings .....	82
<b>7. Search Engine Spam .....</b>	<b>83</b>
7.1 Defining Search Engine Spam.....	83
7.2 To Spam or Not to Spam .....	84
7.3 Been Caught Spamming?.....	84
7.4 Common Types of Search Engine Spam .....	84
7.4 The War Against Search Engine Spam.....	87
<b>A Final Word of Thanks.....</b>	<b>88</b>

# 1. Findability

## 1.1 Introduction

It is hard to believe that the World Wide Web came into existence as recently as 1991. In just over a decade, websites have evolved from digital approximations of print brochures into organic, self-contained “user experiences”. Website owners talk of *usability*, *sticky content*, *page impressions* – and other jargon that describes unique features of this revolutionary medium.

As users become increasingly web literate, their online behavior grows in complexity. Today, the average internet user can interact confidently with a diverse range of websites, and can quickly tell whether or not a site is likely to contain useful information. Users no longer aimlessly “surf” the internet; they purposefully *search* the internet. Web users have become adept at finding what they’re looking for.

Understanding how web users find information is crucial to any web marketing strategy. After all, no-one’s going to visit your website if they can’t find it. If you can’t make your website findable, why make your website at all?

This book will help you to increase the traffic to your website – providing you are willing to change (at least a little) its content and design. If you haven’t yet launched your site, this book will help you get it right, and build traffic, from the start.

The good news is that with a little effort and commitment you *can* achieve impressive results in search listings, particularly in Google, the world’s most popular search engine. There are billions of pages on the web, and the vast majority of them are maintained by people unfamiliar with the principles of website findability. By the end of this book, you will have a huge advantage over your unenlightened competitors.

The even better news is that *you won’t have to spend any money* buying search engine links, or taking out banner advertisements, or paying web marketers. You may have to spend a little money on your site re-design, but only if you can’t do it yourself.

The best news is that this book will not only save you money, it will *make money* for your business. The principles of website findability that I have outlined here will show you how to increase your visitors – and how to convert them into customers.

**If you read the entire book, and practice its techniques, you will know so much about website findability that you will be able to charge others for your services!**

## 1.2 Google and Other Search Engines

The ultimate aim of a findability strategy is to get visitors to your website.

Making your site findable *does not* mean you have to get it listed in hundreds of search engines across the world. Optimize your site for a single search engine – Google – and you will achieve most of your findability requirements.

### 1.2.1 Google

Consider this 2003 report of global search engine use:

<b>Google</b>	<b>56.1%</b>
<b>Yahoo!</b>	<b>21.5%</b>
<b>MSN Search</b>	<b>9.4%</b>
<b>AOL Search</b>	<b>3.7%</b>
<b>Terra Lycos</b>	<b>2.3%</b>
<b>Altavista</b>	<b>1.9%</b>
<b>Askjeeves</b>	<b>1.6%</b>
<b>Others</b>	<b>3.5%</b>

*Source: OneStat.com, May 2003.*

Google hosts over half of all searches carried out on the web. It is more important to focus on achieving results in Google than in all other search engines combined.

There are other good reasons for making Google the primary focus of your findability strategy:

#### **Google sets the standards that other search engines follow.**

Google rose rapidly to the top of the search engine heap because its results were of high quality, and were relevant to users' queries. In becoming the benchmark, Google forced other search engines to follow its lead, and emulate its technology. The techniques that get your site found in Google will also help it perform well in other engines.

#### **By optimizing your site for Google, you will get traffic in other ways.**

The principle underlying Google's method of ranking web pages is this: *pages that are popular with visitors should appear higher in search results than pages that aren't*. In optimizing your site for Google, you will create the type of web pages that tend to be popular with visitors.

#### **Google provides the search results for AOL Search.**

In optimizing your site for Google you are optimizing it for AOL Search too, since AOL uses Google's technology to provide its search results.

## You can achieve success in Google without spending money.

Google does not have a pay for inclusion (PFI) policy – it will include your web pages for free. You may pay for *Sponsored Links* and *AdWords* on Google – more about these later. But my aim in this book is to show you how to achieve good positions in the regular Google results (see screenshot below), for which Google does not accept payment.

## “Natural” Search Results

The screenshot shows a Google search for "usability consultants". The search bar is at the top with the Google logo and navigation links like "Advanced Search", "Preferences", "Language Tools", and "Search Tips". Below the search bar, there are tabs for "Web", "Images", "Groups", "Directory", and "News". The search results are displayed in a list format. On the right side, there is a section for "Sponsored Links" with three ads. Red arrows point from the text "Paid-For Results (AdWords)" to these sponsored links. The natural results include "Website Usability Consultants", "Usability First Consulting", "Consultants Directory (BayCHI)", and "ADG Resources: Usability Consultants".

*The regular Google results are also called the “free” or “natural” results. Google never mixes AdWords or Sponsored Links with regular results.*

If you are able to design web pages, or have some limited HTML skills, you will be able to make changes to your website to improve its position in Google, without spending money.

If you don't have these skills, you can make free improvements to your performance simply by getting the right type of links to your site, as I will explain.

Even if you hire a web designer to implement the changes I recommend, this will be more cost effective than paying for online advertisements and other forms of web marketing.

### 1.2.2 Other Major Search Engines

The search engine industry has undergone much conglomeration and consolidation in recent times. The majority of Google's competitors are owned and powered by Yahoo. Even MSN's results are currently powered by Yahoo-owned technology (though Microsoft plans to launch its own search technology in 2005). But Google powers another important engine – AOL Search.

By optimizing for Google and Yahoo then, you will be optimizing for all major search engines.

The optimization techniques that work for Google differ in some instances from those that work for Yahoo, though these differences are not always mutually exclusive. For example, Google ignores a part of the web page header known as the Meta Keywords section (as explained in Chapter 4), but Yahoo doesn't. Thus, the use of Meta Keywords has no effect on Google optimization, but may improve Yahoo optimization.

I will give you several Yahoo-specific tips in this book, but its guidelines are primarily geared towards success in Google.

### 1.2.3 Paid Inclusion

Yahoo has introduced various paid inclusion programs (which are explained on – and accessible via – Overture.com, one of its sibling sites). Unlike Google, Yahoo accepts payments for URLs to be submitted for inclusion in its natural results, with subsequent payments for any clicks received.

Paying for a site submission *does not guarantee* that the URLs will be given high – or any – listings in the natural search results. Yahoo points out that less than 1% of its search results are paid-for inclusions. If you want to pay for inclusion in Yahoo and its partner sites, go ahead. But I will show you how to get listed in the natural search results of Google and Yahoo *without* having to pay.

(See also Section 6.5, *Search Engine Marketing and Paid-For Listings*.)

### 1.3 PageRank

Google revolutionized internet search when it introduced a new way of matching web pages to users' search queries.

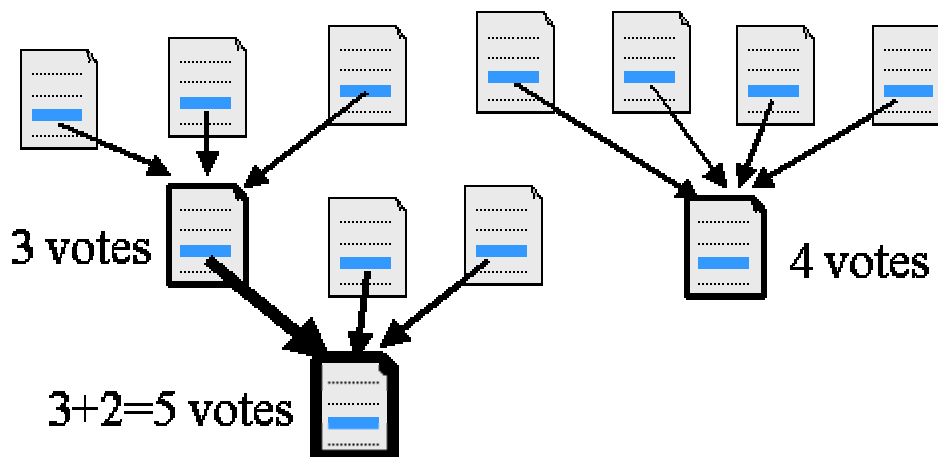
Before Google, most search engines used simple word matching and counting techniques. When a user searched for "blue widgets", for example, the search engine returned a list of results that simply contained *the most occurrences* of the phrase. This led to all sorts of manipulation of the search listings – such as pages that contained thousands of lines of keyword-stuffed text; pages with hidden or invisible text; and poor, non-relevant results.

In the late 1990s, Google's founders Larry Page and Sergey Brin realized that human nature would lead to more manipulation of search engine results over time, rendering them almost useless – unless another method for finding, evaluating and listing web pages could be developed.

Based on their research at Stanford University, Page and Brin saw that the more useful a web page was, the more links it tended to get from other sites. Each link from another site was like a vote – useful sites tended to receive more links, and more votes. The Google founders realized that a link-based measure of popularity could form the foundation of a new, improved way of providing internet search results.

In the late 1990s, Brin and Page made link analysis the basis of their new *algorithm* (a raw, mathematical computer program), which assessed the value of web pages in relation to user search queries.

Using Brin and Page's method, a search for "blue widgets" would still return pages that contained the phrase blue widgets – but the most popular pages would be returned first. Today, the concept of link-based ranking – or PageRank – remains central to the way Google orders its search results.



*Simplified interpretation of the "voting" principle that underlies the PageRank concept.*

### 1.3.1 PageRank on The Google Toolbar

To find out the PageRank of any web page, you must have the Google Toolbar installed, and you must visit that page with your web browser.

To download the Google Toolbar, got to <http://toolbar.google.com/>

Note that in exchange for using the PageRank feature, Google asks you to consent to passing it information on the URLs you visit. According to its website:

Google respects and protects the privacy of the individuals who use the Google Toolbar. Information in a URL, such as user names and IDs, passwords, and account information of any kind are not passed on to Google and are not kept or stored anywhere.

See: <http://www.google.com/privacy.html>

Once you download and install the toolbar, and consent to having it collect information about your browsing habits, you will notice that PageRank is displayed visually, in the form of a green bar on the browser.

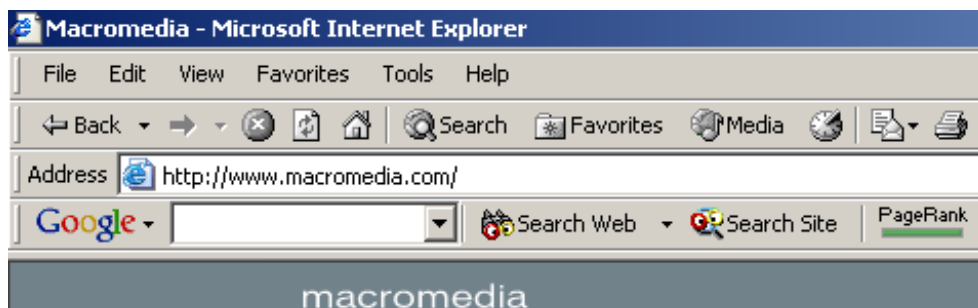


*Google Toolbar indicating a PR of 6.*

By holding your mouse over the bar, you can find out the PageRank of the page displayed in the browser, on a scale of 1 – 10. The message will say, “Page Rank is Google’s measure of the importance of this page,” followed by the PageRank out of 10, in parentheses.

Among professional search engine optimizers, Page Rank is usually referred to simply as “PR”. So a page that has a PageRank of 4 is described as being “a PR4”.

**Each web page has a PageRank value. The PR of individual pages from the same website can vary.**



*The Macromedia homepage has a Green Toolbar – a PR of 10.*

It is quite possible to have a PR of 0 – and extremely difficult to achieve a PR of 10 (among the few PR10 sites are Macromedia.com and, of course, Google.com).

A PR of 5 or even 4 is generally enough to get listed highly – even in the top 10 search results – for search phrases that aren't too competitive.

**While PageRank is important, it is only one of the factors that influence a web page's position in search engine listings for a given search query.** Don't worry if your pages seem to be “stuck” on a certain PageRank value. Many optimizers overestimate the importance of PageRank, and underestimate the importance of other optimization factors, which will be described throughout this book.

### ***Green PR Bar***

A PR10 page will display a “full” bar that is colored green. (See picture above.)



### ***White PR Bar***

A page with a PR0 will display an “empty” bar that is colored white.



### ***Gray PR Bar***

For certain pages, particularly new pages or those that are not linked from other pages, the toolbar will show an empty *gray* bar, with the message: “This page is currently not ranked by Google.”

A gray toolbar is also used to indicate that the page has been penalized by Google and is banned from its search results.

It is rare for a web page or site to be penalized; Google takes this action when it believes that the owner or creator of the page has used techniques aimed at deceiving searchers. Like unsolicited email messages, these unwanted search results are referred to as spam (more specifically, “index spam” or “search engine spam”).

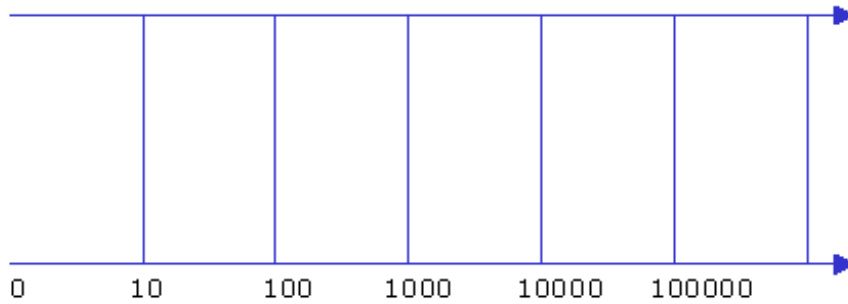
**Note: Chapter 7 contains descriptions of each of the different spam techniques that you should avoid.**

### **1.3.2 “Real” PageRank**

PageRank as displayed on the Google Toolbar is only an approximation; Google does not reveal the exact PR value of any web page.

One thing we know for sure is that the 1-10 scale shown on the toolbar is not an ordinary, linear scale. It is what mathematicians would call a *logarithmic* scale.

Think of a logarithmic scale as a ladder. On a regular ladder, each of the steps is the same distance apart. On a logarithmic ladder, however, the distance multiplies from one step to the next!



*Log scale to base 10*

So climbing the PageRank scale is a task that gets more difficult as you go along.

Google doesn't reveal what the exact order of the scale is, but the consensus among optimizers is that it's approximately a "base 6" logarithmic scale.

What this means is that getting your web page from PR1 to PR2 is *six times as difficult* as getting it from PR0 to PR1. It is a further thirty-six (six multiplied by six) times more difficult to get it to PR3. And so on...

In fact, if the base-6 speculation is correct, the distance between PR0 and PR10 is *over a million times greater* than the distance between PR0 and PR1. No wonder so few sites have PR10!

For the math fiends, the following section shows how to calculate the approximate value of your "real" PageRank.

**Note: You don't need to learn this!** I've included the following section primarily to illustrate why PageRank rises quickly with initial optimization effort, but slowly after that – so don't beat yourself up if your PageRank isn't increasing as quickly as you'd like it to!

### 1.3.3 "Real" PageRank Calculation

The toolbar PageRank uses a logarithmic scale. To calculate "Real" PageRank (RPR), i.e. PageRank expressed on a linear scale, we use the following formula:

$$\mathbf{RPR = (6 \wedge PR)}$$

*RPR is Real Page Rank, expressed on a regular scale.*

*PR is Toolbar PageRank, expressed on a log scale.*

*^ is the exponentiation operator.*

In plain English: Real PageRank equals six to the power of Toolbar PageRank.

Here are the base-six conversion values from Toolbar PageRank to Real PageRank:

<b>PR</b>	=	<b>RPR</b>
0	=	1
1	=	6
2	=	36
3	=	216
4	=	1296
5	=	7776
6	=	46,656
7	=	279,936
8	=	1,679,616
9	=	10,077,696
10	=	60,466,176

As can be seen, the values at the higher end of the scale are staggeringly high, leading some optimizers to speculate that the second half of the scale (6-10) uses a log base much lower than 6.

Google doesn't reveal what log base(s) it uses, but I hope this section will nevertheless have helped you conceptualize the difference between Toolbar PR and Real PR.

## 1.4 Search Engine Optimization

Search Engine Optimization (SEO) refers to the practice of enhancing the design, content, structure and connectivity of a website so as to maximize its performance in search engines.

While SEO is something that, armed with the theory, anyone can do – as is designing the website, or writing its content – some do it better than others. A professional SEO industry is now emerging, particularly around highly competitive search queries. If you find that you have a flair for optimizing, you may want to consider making a career out of it!

### 1.4.1 The SEO Industry

Optimization is a practice that has been around almost as long as the search engines. One popular service in the early days of SEO involved submitting URLs to a plethora of search engines – a practice that today has little value.

As Google began to transform and ultimately dominate internet search, the optimizers concentrated their efforts on discovering how to get sites listed with the upstart search engine. But Google has always been guarded about how its proprietary search technology works. SEO practitioners have had to treat it as a black box, continually trying to reverse-engineer its algorithm.

Google takes a benign attitude towards SEO, regarding it as an inevitable consequence of its own success – though it does not tolerate any attempts to deceive searchers or site visitors.

On the “Information for Webmasters” section of its site, Google points out that “many SEOs provide useful services for website owners, from writing copy to giving advice on site architecture and helping to find relevant directories to which a site can be submitted.”

See: <http://www.google.com/webmasters/seo.html>

Google warns that no optimizer can guarantee a number one listing in its search results, and not to believe any claims that it has a “special relationship” with any SEO firm. Unfortunately, there are many charlatans in the burgeoning cottage industry of SEO.

Yahoo has followed Google’s lead. It too has developed a symbiotic relationship with optimizers, and tries to root out dubious SEO practices.

By applying the guidelines contained within this book, you will be able to optimize your own website. Yet, this book may convince you of the value of hiring a good optimizer, since applying those guidelines well requires time and patience.

If you are hiring an optimizer, check their pedigree first by making sure they have an impressive portfolio of sites and results, and try to talk to some of their previous or existing clients.

## 2. Planning Your Website

If you are serious about building a website that gets high volumes of traffic every day, you must plan your site before you begin. Remember, you are not just designing an online brochure – you are creating a web-based marketing strategy.

### 2.1 Aims and Objectives

Before jumping straight into optimizing your site for certain keywords, take a step back and ask yourself what the aims and objectives of your new site are.

Website objectives generally fall into the following broad categories:

- **Communication**  
To convey information about your organization, business, products or services to the public.
- **Marketing**  
To promote your goods and services; to expand your existing market.
- **Sales**  
To sell your goods or services directly to customers worldwide.
- **Branding**  
To create or enhance the identity of your organization, and the feelings it evokes in others.

Optimizing your site for Google and other search engines will bring visitors, but optimization will do little to ensure that the site meets your objectives – such as to raise the company’s profile; increase sales; etc.

**You can bring a horse to water, but you can’t make him click on the “Buy Now” link!**

Write your content and design your content for your visitors, with search engines prominent in the back of your mind. Search engines uses automated “bots” to read web pages; if you design your pages *only* for search engines, they will be unappealing to human readers.

## 2.2 Site Structure & Integration

### 2.2.1 Site Hierarchy

Google likes sites that are simply constructed and well interlinked.

In Google's Design and Content Guidelines for webmasters

(<http://www.google.com/webmasters/guidelines.html>), the first suggestion is:

“Make a site with a clear hierarchy and text links. Every page should be reachable from at least one static text link.” This means that every page on the site should be linked to at least one other, and that you should avoid (or at least compensate for) dynamically generated pages.

We will deal with the issues of dynamic pages shortly. For now, let us examine what it means to have a well-integrated site, and why this is important.

### 2.2.2 Getting Listed: How Google Bots Find Your Site

Consider how Google finds and indexes the pages on your site. Google does not employ humans to physically click on sites and evaluate them. (This human verification process is, incidentally, used by internet *directories* such as [www.dmoz.org](http://www.dmoz.org), but not by search engines.)

Instead, Google sends out its army of *bots* – computer programs that have been designed to follow every web link they come across, and retrieve the information they find. Search engine bots are also known as “spiders” or “crawlers”.

Google's bots crawl the web regularly. There are three ways that a Google bot can arrive at your site:

1. A Google bot has previously visited your site.
2. You submitted a URL directly to Google (see: <http://www.google.com/addurl.html>).
3. A Google bot followed a link from another site to your site.

If the Google bot that searches the web for new content (called *Freshbot*) arrives at your site once, Google will return to your site with its standard spider (known as *Deepbot*) on a regular basis, to check for updates.

If you submitted a URL to Google, then Google will send *Freshbot* to the URL you submitted. *Freshbot* will start at the page you submitted, then follow each of the links on that page (menu items, etc.). On each subsequent page, *Freshbot* will continue the same process of identifying links and following them.

#### *Avoiding Orphan Pages*

If your site includes pages that can't be reached by following a link, or a series of links, from the homepage, then these pages – called “orphans” – will not be included in Google's index of your site, under normal circumstances.

*Orphans are web pages that are not linked to by any other pages.*

Google's advice is to avoid orphan pages – unless your site includes pages that you specifically do not want to appear in its index.

### 2.2.3 Creating Non-Searchable Areas on Your Site

There are instances where you may want to create pages on your site that you don't want search engines to pick up. For example, you may have some page mockups or test pages that you aren't yet ready to show to the world.

In this case, you may deliberately create orphan pages. But there is no guarantee that an orphan page will never appear in a search engine. You may not link to it, but someone else may – unlikely, but possible.

If you do wish to create pages or folders on your site that do not get indexed by the search engines, there are two options:

- (i) Password-protect the private pages or folders
- (ii) Include a robots.txt file, with instructions for search engine bots on what not to crawl, in the root directory of your site

### 2.2.4 Robots.txt

If you want to prevent Google, Yahoo! or any other search engine's bots from indexing certain web pages, or folders on your website, you can use a robots.txt file.

Robots.txt is the name you give to a file in plain text format (hence the .txt extension), that you place on the *root folder* of your web domain.

You can create the robots.txt file using a text editor such as Windows Notepad. The text file is relatively simple to create and usually very short. There are many excellent tutorials available on the web.

We won't attempt a full tutorial here, but here are a couple of simple examples of the contents of robots.txt files:

#### Example 1: To Prevent Google from Indexing a Page Named "Fun.html"

```
User-agent: googlebot
Disallow: fun.html
```

#### Example 2: To Prevent All Search Engines from Indexing the "Clients" and "Samples" Folders

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

### 2.2.5 Dynamic Pages

Dynamic pages are web pages that a software program on the web server creates. Unlike normal, static web pages, the content of dynamic pages may vary, depending with what options the site visitor selects, and/or what information she has submitted.

You will often recognize dynamic pages by their long or unusual URLs. For example:

`http://www.widget-world.com/userquery.asp?string=widgets&lang=eng`

Dynamic pages can create problems for search engines bots. For example, dynamic pages are often generated after the site visitor fills out a form, which a bot can't do.

If you cannot avoid using dynamic pages on your website, you can help Google's bots to find them by including text links from static pages to the dynamic URLs. A good way to do this is by including the links in a sitemap (see Section 3.4, *Sitemaps*).

Another way to help ensure that your dynamic URLs get indexed is to try to keep them short and manageable. As a rule of thumb, if a human user would have little difficulty in entering the URL manually, it has a better chance of being picked up by search engine bots.

### 2.2.6 Session IDs

A session ID is a particular type of dynamic page, used by sophisticated e-commerce sites, or sites that collect a lot of data about visitor behavior.

Session IDs are used to track the behavior of individual visitors as they progress through the pages of a website. Each site visitor is assigned a unique identification code, and that code is used in all the URLs that the visitor looks at in the site, during the course of her visit.

The technique is popular on large e-commerce sites such as Amazon.com, which allow the visitor to see what products she's already looked at, or what she's got in her shopping cart. Session IDs provide the technology that help create a personalized customer experience.

**Google specifically discourages the use of session IDs**, since they cause severe problems for its bots. URLs that contain session IDs also tend to be longer and uglier than most.

To ensure successful performance in search engines, don't use a session ID solution unless it is unavoidable.

Note: If you are running a sophisticated e-commerce site, that uses session IDs, it may be worth contacting Google to find out how you can get your products listed using its free "trusted feed" solution, which lets you provide real-time information about your updates on its shopping comparison site, Froogle.com, using a technology known as XML syndication.

Yahoo also offers a trusted feed service, although it charges a fee.

### 2.2.7 Multiple Domain Names

If you have multiple domain names for your site, ensure that you rely on one “master” URL, and redirect all the alternatives to that one. Otherwise Google may penalize you for having sites that contain duplicate content – which it will treat as possible plagiarism. Duplicate sites will also unnecessarily weaken your PageRank by spreading it across multiple sites instead of one.

For example, let’s say you have a master domain name called <http://www.widget-world.com>

You may, for understandable reasons, also have alternative domain names for the same site:

<http://www.widetworld.com>

<http://www.widget-world.net>

<http://www.widget-world.co.uk>

In such a case, you should create what is known as a *301 redirect* on the server for each of these backup domains, and have them all point to the master domain.

Your hosting provider will be able to set up a 301 redirect – also known as a permanent redirect. This is a straightforward request.

It is important that you do not use any other method of redirecting site visitors; otherwise the search engines may penalize you for using underhand tactics. Redirects are often used as a way of trying to trick search engine bots and “cloak” the true content of a site.

#### **Redirect methods you should *not* use:**

- a 302, or temporary redirect
- an on-page redirect, using the <meta-refresh> tag
- a Flash file redirect
- any other method of redirecting that is not a 301 or permanent redirect

**Tip:** *Avoid using alternate domain names in your marketing material, as any links that point to those alternates will not boost your PageRank.*

## 2.3 Site Content

Many websites are little more than electronic marketing brochures – a few pages of basic marketing information about the company and its services.

These slim, *brochureware* sites – however colorful or slick – are ineffective at attracting searchers. Such sites are designed according to print media principles. Yet people interact with websites in ways that are qualitatively different to their interactions with print brochures. Web users are active, focused on getting the information they want, and intolerant of interruptions (which is why pop-up advertisements are so irritating).

**To have a successful website, you must embrace the web as a medium.** The web has many unique features that have no correlations in print.

### 2.3.1 Think Like a Searcher

Imagine you are searching Google for suppliers of a specific product – a type of widget.

What happens if you begin your search by simply typing “widgets” into Google?

You are confronted with millions of matches for your search query. Each result looks like the next – generic descriptions of companies, each claiming to be the best, most reputable widget provider. Clicking on the links would take you to the sites of companies that sell widgets, with pages about who they are; the services they provide; their staff profiles; how to contact them; etc.

The process leaves you feeling overwhelmed, and no wiser about which widget provider to choose.

Regular search engine users, having gone through this process on more than one occasion, have learned to begin with targeted searches, and to quickly refine their queries if necessary, in order to find what they are looking for.

To target or refine your search, you would type in a more specific query. For example, “small green handmade widgets,” or “polymer x-shaped widgets.”

The number of matches for a targeted search is likely to be smaller, and to contain fewer brochureware sites, since marketing-oriented sites are less likely to have precise language.

### 2.3.2 Develop Unique and Detailed Content

In order to increase traffic, aim to make your website findable by users performing targeted searches.

To be found for a search on “polymer, x-shaped widgets”, it is not enough to have a few general pages saying that you sell all kinds of widgets. Yet, a single page deep in your site devoted to polymer, x-shaped widgets might be enough to get listed in the top 10 in Google for a search on this phrase.

**A website that only has a few pages of general information is not going to stand out from the competition. Aim to have many pages of precise information.**

Of course, your pages should be informative, and not designed solely for search engines. If you add a page of content about polymer x-shaped widgets, make it one that the visitor will want to read. List the advantages of x-shaped widgets. Describe the purpose they serve and how they are used. Explain the benefits of using polymer over steel. Then tell the visitor that you have the widgets in stock, along with lots of other types, and encourage them to follow the links to other products, or to your shopping cart page, or to your homepage.

### ***Landing Pages***

When a user finds a page of your site through Google or another search engine, this page becomes her *landing page*. The more landing pages you have on your site, the more chances you have of attracting visitors, particularly if each of the landing pages is about a different subject, and is optimized for a different set of keywords or phrases.

If you increase the number of landing pages on your website, you also increase the number of different keyphrases for which your site becomes findable.

### **2.3.3 Manage and Update Your Content**

Important as it is to add informative, text-rich pages to your site, you should not bite off more than you can chew. Search engine optimization should never take precedence over the quality of your content.

I have already said there is no point in having a great site if nobody can find it. The reverse is also true: there's no point in generating huge volumes of traffic to your site if visitors have a bad experience when they get there.

Your content should always be easy-to-read and informative. Your site should be usable and useful to the visitor. It should be accurate and up-to-date. Simple things like phone numbers, contact names, opening hours, links to other sites, and so on, are subject to quick change. Nurture your site and make adjustments where necessary.

Managing your site also involves making changes based on the types of visitors you are already getting and the keyphrases that they are entering. These can be gathered using a good statistics package. (See section 6.1, *Site Statistics*.)

Remember, a good website is a living, evolving document. If you're serious about creating a site that performs well in search engines *and* works as an effective marketing tool, you must be prepared to make a commitment to regularly maintain and update its content.

A good website will be updated at least once a month, if not more. I update and add pages to each of my websites several times a week.

## 2.4 Keyphrase Research

The goal of search engine optimization is to get as near as possible to the top of the search results for queries relevant to your site's subject matter.

Before you optimize your site – or any of the individual pages within your site – you should list the keyphrases for which you want to be found.

### 2.4.1 Keyphrases

Trying to optimize your page for individual keywords, rather than keyphrases, is a common mistake.

The competition on individual words is clearly higher, as there are a more limited number of individual words than there are two-, three- or four-word phrases. (It is unusual to optimize for phrases longer than three or four words.)

For example, if you are a flower shop and you concentrate on optimizing your homepage simply for the word “florist”, you are unlikely to get into the top 10 in Google's search results of these words. After all, you are competing with thousands of other florists on the web *and* many more pages about the flower industry, directories of florists, etc.

In any case, a user who is actually looking to buy flowers isn't likely to search simply for the word “florist” or “flowers”. She would be inundated with results (Google has *millions* of results for each of these examples).



This is why users tend to optimize for two- or three-word keyphrases, rather than simply keywords. Recent statistics on search queries confirm this:

**2 word phrases — 32.58%**  
**3 word phrases — 25.61%**  
**1 word phrases — 19.02%**  
**4 word phrases — 12.83%**  
**5 word phrases — 5.64%**  
**6 word phrases — 2.32%**  
**7 word phrases — 0.98%**

*Source: OneStat.com, February 2004*

If, for example, your flower shop is based in Jonesville, Michigan: you could optimize your homepage for phrases such as “florist Jonesville” or “flowers Jonesville”. (Note that Google ignores small words such as “in”, “at”, “the,” etc., so you don't need to optimize for the phrase “flowers in Jonesville”.)

Perhaps your flower shop specializes in certain types of flowers, say orchids and gloriosa. “Orchids Jonesville” or “gloriosa Michigan” might be appropriate phrases. Going through your services, you should be able to think of lots of other keyphrases – “wedding florist Jonesville,” “valentine florist Jonesville,” “exotic flowers Michigan,” “flower arrangements Michigan,” “flower delivery Jonesville,” and so on.

The trick, again, is to think like a searcher. Put yourself in the position of one of your potential customers – in this case, someone in the Jonesville area who is thinking of buying flowers.

Consider the different real-world scenarios that your potential site visitors – a man who wants to find a special type of flower that will impress his new girlfriend; a person looking for advice on what kind of bouquet is appropriate for a funeral; etc. Try to imagine the phrases that a searcher would enter.

### **Tips on drafting your initial list of keyphrases:**

- Repeat the same word in different keyphrases. You may re-use individual words in different keyphrases, as long as you only target one keyphrase per page (a strategy I encourage, for reasons I will explain later).
- Use regions or other common query-refining words. For a small, local business such as a florist, a regional keyword is important.
- Be specific about product or service niche – e.g. “martial arts footwear”, “large piece jigsaws”. In doing so, try to look at your industry from the outside. What words would the average Joe use to describe your products and services?
- Identify as many keyphrases as possible, particularly if you intend to add pages to the site over time (a strategy I strongly advise). As you add more pages, you will add more keyphrases, thus widening the net with which you “catch” visitors.
- Don’t worry about including many small or common words as part of a keyphrase. While it is difficult to create meaningful phrases without using such words, remember that Google ignores small, common words in user searches (except where the user encloses her search query with quotation marks).



Treat your initial list of keyphrases as a first draft. Be as liberal as you like and try to make the list as long as you can.

## 2.4.2 Keyphrase Suggestion Tools

The next step of the process is to identify keyphrases that users really search for. Doing so will make you delete some of the phrases from your initial list, where you discover that users aren't actually searching for these phrases. But this step will also provide keyphrase suggestions you hadn't thought of.

So how can you tell whether keyphrases correspond to real-world searches? In fact, there are various tools on the web that allow you access to real search queries – databases of the phrases that have been recently typed into major search engines, and their frequencies.

What follows is my overview of three of the best-known keyword suggestion tools.

### *Overture*

Overture's suggestion tool is quick to access and free to use. You can find it at:  
**<http://inventory.overture.com/d/searchinventory/suggestion/>**

Enter a search term to get similar and related phrases, taken from the database of queries entered into the Overture search engine. The tool also indicates the relative popularity of each phrase, by providing the number of times the query was entered into Overture in the previous month.

Overture has a very small share of the search engine market. However, the keyphrases that users enter do not vary much from one search engine to the next, so you can consider Overture's keyphrase suggestions as representative (in percentages) of queries conducted on other search engines.

### *Wordtracker*

Another popular, sophisticated keyword tracking service is provided by Wordtracker:  
**<http://www.wordtracker.com>**

The Wordtracker tool allows users to search a database of queries taken from the search engines MetaCrawler and Dogpile. Wordtracker provides access to query logs stretching back for two months, and the database is updated weekly.

Wordtracker also has a neat "lateral suggestion" feature. It looks at the text of web pages that contain the keyphrase you're looking for, and identifies other keyphrases that are frequent on those pages.

While you have to pay a subscription for the full version, a free trial of Wordtracker is offered on the site.

### *Google AdWords*

Another great way to check on the popularity of keywords is by using the Google AdWords keyword suggestion tool. You don't have to place a Google AdWords order to use the tool, although you have to go through a couple of steps *as if* you were placing an order.

Here's how to do it. Go to the AdWords homepage:

<https://adwords.google.com/select/>

1. Click on the gray "Click to Begin" button (you're not obliged to login or register).
2. Select your language and choose a country, or group of countries, or all countries. The tool will only return search queries that are popular in the regions you select here.
3. Click "Save and Continue".
4. Fill out all of the fields on the following page, as if you were creating an ad.
5. Click "Create Ad & Continue".

You will then be taken to the keyword suggestion tool, which mines Google's own database of queries.

Note that Google makes predictions of how many clicks a link would receive if it were in the Sponsored Links section of its results page. This prediction is most likely *lower* than the number of clicks a link would receive if it were in the natural results – which is what you're aiming for!

### **2.4.3 Common Words and the Importance of Word Proximity**

Often, when you conduct a search in Google, you will notice that it returns a result saying that a certain word "is a very common word and was not included in your search."

These common words are technically known as *stopwords*, and Search Engines typically do not include them in their analysis of page content, for obvious reasons.

**Here is my full (but unofficial) list of Google stopwords:**

**I, a, about, an, are, as, at, be, by, com, de, en, for, from, how, in, is, it, la, of, on, or, that, the, this, to, was, what, when, where, who, will, with, und, the, www.**

So, if Google ignores these words, you should leave them out of any keyphrase you create, right?

Unfortunately it's not that simple. To illustrate my point, take a look at the search results for the following two keyphrases:

Holidays in Ireland  
Holidays from Ireland  
Holidays and Ireland

Look at the top 10 Google results. Chances are, the results will be slightly different for each search. Now, if Google ignores the stopwords, "in", "from" and "and", why do the results differ?

One theory is that the answer has to do (at least partly) with *word proximity*. That is, while Google ignores the individual stopword, it notes the distance (in character spaces) between the keywords, and tries to match it.

If you look at the three phrases I gave in the example, the differences between the keywords that are actually searched is different in each case. I demonstrate this below, by counting the number of dots from the ‘H’ in Holiday to the ‘I’ in Ireland.

Holidays in Ireland  
0.....14

Holidays from Ireland  
0.....16

Holidays and Ireland  
0.....15

By taking the word proximity into account then, we can expect that the results for a Google search on *Holidays in Ireland* would be similar to a search on *Holidays to Ireland*. And they are – but they are not exactly the same.

I believe that the Google algorithm does take account of word proximity – but it is not the only factor that affect searches that use different stopwords.

Clearly, searches that use different common words produce different results. And searches that are contained in quotation marks – such as “holidays in Ireland” will force Google to search for the exact phrase, including the common word.

**Conclusion: While Google says it ignores common words, this doesn’t mean you should ignore them.** Treat variations of search phrases, which use different common words, as separate keyphrases that need to be targeted individually.

#### **2.4.4 Assessing the Competition**

You should now have a list of keyphrases for which you would like your site to show up highly in search engines.

For some of those phrases, getting into the top 10 may be achievable in the short- to mid-term. For other phrases, the competition may be a lot greater, and even getting into the top 100 would be an achievement.

##### ***Check the Number of Matches***

One simple way to analyze the overall competition is to look at how many matches Google displays for a particular keyphrase.

Generally speaking, the more matches there are for a given phrase, the harder it is to get into the top 10. As a very broad rule of thumb, a search that has less than 100,000 matches can be considered one with low competition, while over 1,000,000 matches indicates high competition.

Common words and phrases tend to generate more results, so there is less competition for rarer words, jargon, or words related to niche subjects. The majority of business websites still use the bland, common language of marketing-speak, and are typically *not* optimized for Google. There are vast areas of the dictionary – think of it as virgin territory – ready to be claimed by those with clear optimization strategies.

### ***Identify your Competitors***

Check who is currently coming up high in the results for the keyphrases you are targeting. Are these your real-world competitors? While you may not be able to get into the top five or 10 for your keyphrase, this may not be necessary if none of the sites listed for those results directly compete with you for business.

For example, if the top five results for search for “flowers in Dublin” don’t contain any flower shops, then a searcher who is looking for a florist is likely to click further down the list. Don’t get too hung up on positions; concentrate on outperforming your direct competitors.

### ***Check Competitor PageRank***

You should check the PageRank of the sites that are already in the top 10 for your target keyphrase, particularly those of your real-world competitors.

PageRank gives you a general, indicative sense of the competition – but it is always balanced with relevancy to the search query. If the site in tenth place has a PageRank of 3, this *does not mean* you have to achieve a higher PageRank to overtake it. However, if your page optimized such that its relevancy to the keyphrase was “equal” (as far as the search algorithm is concerned) to your competitor’s page, then a measure of PageRank would settle the contest.

### ***Check Competitor Backlinks***

Find out who is linking to your competitors, and whether you can get links from the same sites. How to check your competitors’ links, and why this information is important, is explained in Chapter 5, *Developing Links With Other Sites*.

### ***Assess Whether Competitor Sites Have Been Optimized***

Arguably the best way to assess competition is to analyze whether sites in the top 10 have been optimized and, if so, to what extent. It is not always easy to tell whether a site has been optimized, however, since the purpose of optimization is to make it look as though the keyphrases occur “naturally”.

However, as you become more practiced at findability, you should be able to tell whether other sites are optimized or not. (Those that are will follow similar guidelines to the ones you are discovering in this book.)

Aim to target keyphrases where at least some of the sites that are currently in the top 10 have not been deliberately optimized.

## **2.5 The Google Sandbox Effect**

### **2.5.1 Understanding the Sandbox Effect**

Until recently, those with a good knowledge of search engine optimization could launch a new site and, applying their expertise, expect it to show up in the top 10 Google results for certain keyphrases in a matter of weeks.

In early 2004, however, Google made a change to its algorithm to prevent such “quick wins”. Theories differ as to the specific nature of this change, but its outcome is clear: it is now extremely difficult to get a new site – *with a new domain name* – listed prominently in Google’s results within a short time frame.

This algorithmic change has been dubbed Google’s “Sandbox Effect”. There are a couple of good reasons as to why Google has implemented this change.

- a) Many spammers profited from techniques that they knew would cause their sites to get banned. They didn’t care; when the sites got banned, they immediately launched new domains with the exact same content, and got high listings once again. The sandbox effect prevents this tactic from succeeding.
- b) Age of a site is arguably an indication of quality. The longer established a particular domain name, the more likely it is to have been maintained, and to offer useful content. The sandbox effect prevents new sites from displacing more established competitors, even if the newer sites are otherwise better optimized.

The sandbox effect causes new sites to go through a “probationary period” of about 6-12 months before they can achieve good listings in Google.

Google representatives insist there is no official sandbox *policy*, and optimizers have reported successes in launching new sites that quickly got high positions – but only where the search terms were not very competitive.

#### **The Sandbox Effect – My Own Observations:**

- New domains still get visited immediately by Googlebot, if linked to by a high-quality site.
- Google still adds new domains to its index quite quickly (a search for “site:www.newdomainname.com” will show how many pages of a given site are indexed).
- New domains are unlikely to get high positions – for keyphrases that are even mildly competitive – for several months, usually 6-12 months.

### **2.5.2 Planning for the Sandbox Effect**

#### **If you want to get into Google’s search results quickly:**

Create a mini-site, which is a subset of an existing domain. Don’t register and use a new domain.

For example, when I launched this book, I decided to add a subsection to my existing site, **www.michaelheraghty.com**, as I wanted. I *didn't* register a new domain – a technique I probably would have implemented had I written the book a year earlier. Within days, the new pages began showing up in Google's results.

**If you want to launch a new site with its own domain:**

Be patient, but use all the techniques I describe in this book. You may not see results right away in Google, but within a year you will.

**If you want to launch a new site, with its own domain – and you want to get results in Google immediately:**

Consider buying (or otherwise obtaining) an existing website, with a domain name you can use. Aim for a site with good PageRank and lots of quality backlinks.

***Warning:*** *If you are buying an existing website, make sure it is a domain that has not expired. Once a domain name expires, Google resets the site's PageRank back to 0.*

## 3. Designing Your Website

Before getting stuck into individual pages, there are certain design decisions you will take at the outset, possibly when designing your page templates, or your site's look-and-feel.

In this short chapter, I want to address some of those decisions – particularly the choice of menu style, and the decisions whether or not to use cascading style sheets, or a site map. I will first explain the importance of usability in website design, and its relevance to findability.

### 3.1 Committing to Usability

Google rewards good usability. Sites that employ a simple, easy-to-use design consistently perform better than others in its search results.

Furthermore, the more usable your site, the more likely it is that visitors will:

- Interact with the site
- Buy products or order services
- Return to the site in the future.

Google's own site is highly usable, and the company has consistently stated it is committed to promoting the needs of the user in its search results.

There are many good books and sites on usability, and the topic is too broad to cover here. But there usability principles that have a particular bearing on search engine optimization, and I will highlight these throughout the following chapters.

For further reading on usability, I recommend Steve Krug's *Don't Make Me Think*, a very popular book among seasoned webmasters.

## 3.2 CSS

Cascading Style Sheets (CSS) are used to separate the stylistic elements of a page (layout, color, borders, font weight, and so on) from its content.

If you don't understand how to use CSS, and don't have the time to learn, you can simply leave it out. CSS is not essential to web page optimization, although it offers many advantages when used correctly.

### *Smaller File Sizes*

By taking styles out of the HTML page and putting it into a standalone style sheet (.CSS file), you can reduce the overall amount of code in your web pages. Pages with less code have smaller file sizes, and Google prefers pages with smaller file sizes.

Though Google doesn't offer specific advice on this matter, the search engine optimization community is generally agreed that 100kb is a good upper limit for page sizes.

### *Well Structured Pages*

CSS also allows you to structure your document according to HTML standards, without compromising the look-and-feel of the page.

Google rewards pages that are well structured, though many designers choose to ignore standards and guidelines, because they produce bland pages. Using CSS, designers can create attractive pages, with much flair, while adhering to the findability design principles that I will identify throughout this chapter (e.g. proper use of HTML heading tags).

### *Browser Compatibility*

If you are new to CSS, be aware that different browsers still interpret CSS standards in different ways, while some older browsers can't read CSS at all. Ensure that your CSS is as cross-browser compatible as possible, and that your HTML pages look acceptable even without CSS.

### *Resources*

There are many useful resources for those seeking to learn more about the effective use of cascading styles.

A useful place to start is at Glish.com. See: [www.glish.com/css/](http://www.glish.com/css/)

Good books on CSS include "Eric Meyer on CSS" by Eric Meyer, and "Designing with Web Standards" by Jeffrey Zeldman.

### 3.3. Menu Structure & Style

Your main navigation menu is of crucial importance to the usability and findability of your site.

A menu is the primary way of providing links from one page on your site to others. I have already stated that you must provide at least one text link to each of the pages in your site – a menu is the easiest and most efficient way of providing these links.

All too many website menus are composed of images, rather than text. Designers combine image menus with JavaScript to achieve a cursor “rollover” effect. Unfortunately however, search engines can’t read text when it is part of a graphic. I will explain this problem and its implications in more detail in the following chapter. For now, it is suffice to say that you strive to use menus that do not use graphics.

**I strongly advise that you make a text-based menu the cornerstone of your website navigation.**

The good news is that CSS will allow you to create colorful, text-based menus with neat rollover effects. Furthermore, a CSS menu is more usable than a graphical menu, since it downloads faster. It looks crisper too, as there is no *color dithering*.



*Example of rollover menu created using HTML and CSS, but no graphics.*

You can find a good tutorial on how to create a CSS-based menu at:  
[http://www.projectseven.com/tutorials/css\\_menus/list\\_01/](http://www.projectseven.com/tutorials/css_menus/list_01/)

### 3.4 Site Maps

A site map is a good way of enhancing your performance in search engines, and making your site easier to navigate. Google encourages the use of sitemaps (see: <http://www.google.com/webmasters/guidelines.html> ).

A good site map

- is a useful navigational aid for visitors
- ensures that search engine bots can reach every page on the site
- helps spread PageRank more evenly throughout the site.

Site maps can take different forms, but usually they are single-page representations of the overall site structure.

**Avoid visual maps that use Gif or Jpeg images. Create your sitemap using text links.**

Site maps become more complicated as the number of pages in a site grows. In this case, it may be useful to “chunk” the sitemap into discrete elements, with each element leading to a smaller map that provides more detail. Google recommends that none of the individual chunks should contain more than 100 links.

Take a look at the eBay sitemap or, indeed, the Google sitemap to see how a large volume of information can be indexed and linked in a straightforward manner:

**<http://pages.ebay.com/sitemap.html>**

**<http://www.google.com/sitemap.html>**

## 3.5 URLs and File Names

Each individual page on the web has its own unique address, better known as a URL (Uniform Resource Locator).

A URL is comprised of a domain name, e.g. **http://www.yoursite.com**, and (but not always) one or more directory names, e.g. **/about-us/**, and (but not always) a file name, e.g. **contact.html**.

The URL for this example would be  
**http://www.yoursite.com/about-us/contact.html**.

Google and Yahoo attach a modest value to the words contained in URLs. It is more important to have your keywords within the web page itself than within its URL; however, it is better yet to have the keywords in both.

Microsoft's new search engine (available in test mode at the time of writing – early July 2004) appears to assign a large value to keywords used in URLs, particularly in the domain name. Those seeking to target Microsoft searches, then, should pay particular attention to this section.

### 3.5.1 Choosing a Domain Name

Your URL is important for your general web marketing, as well as findability. Don't worry if you can't get a domain name that contains your desired keywords or keyphrase; you can still put those words into the file name of a particular page. In choosing a domain name, follow these general marketing principles:

#### *Go for Brevity*

One of the main ways in which first-time users find a site is by typing a name (e.g. a brand name) directly into the address bar of the browser, and adding a .com, or .co.uk, or other suffix to it.

Short domains are more likely to lead to type-ins, since users are more likely to remember them.

#### *Keep it Simple, Stupid*

Single-word domains are preferable to their multiple-word or hyphenated counterparts. A domain name should also be memorable and *easy to spell*, thereby increasing the likelihood of type-ins.

#### *Try a Different Suffix*

Unfortunately, single word domains (and short acronyms) are expensive and difficult to obtain, particularly if they are dot coms. If you can't get a short dot com domain, try another TLD (top level domain) – such as a country-specific TLD (e.g. .co.uk), or a .org or .net suffix.

### 3.5.2 File and Directory Naming Conventions

In optimizing a site for search engines, the idea is to try and make every single page findable, not just the home page. Thus, users will have many different ways of landing on your site (non-homepage entrances to the site are called *landing pages*).

#### *Use Keyphrases in the Filenames*

For reasons I will explain in the next chapter, I advise you to target **one keyphrase per web page**. The first place to include this keyphrase is in the file name (thus making it part of the URL).

#### *How to Separate Keywords*

**Don't** run the individual words of a keyphrase together, as follows:

**<http://www.yoursite.com/mydesiredkeyphrase.html>**

**Don't** use underscores (as shown below) to separate them either:

**[http://www.yoursite.com/my\\_desired\\_keyphrase.html](http://www.yoursite.com/my_desired_keyphrase.html)**

**Do** separate your keywords with hyphens, as in this example:

**<http://www.yoursite.com/my-desired-keyphrase.html>**

The reason? Google treats words separated by hyphens as individual words (as if they were separated by spaces).

#### *Avoid Auto-Generated File Names*

Another drawback of dynamically generated web pages is you often can't control the page names created by an automatic page-generating system.

If, however, you are committed to using a publishing system (such as a content management tool; or a database-driven site that uses PHP, JSP, ASP; etc.), do some research or get some advice to see if it can be configured so as to allow you to specify the file names.

Many of these systems can, with a little tweaking, be configured to allow you to control the names of the files they create.

#### *Use the Keyphrase in the Directory Name*

If you use the filename **index.html** (or `index.htm`, or `index.php`, etc.) as the root page of a particular directory (a standard convention), make sure that the directory name contains the target keyphrase.

For example, if you have want to target the keyphrase blue widgets, you could create a directory that contains the keyphrase, so that page would have the following name:

**<http://www.yoursite.com/blue-widgets/index.html>**

Of course, the same file could also be accessed with the following URL:

**<http://www.yoursite.com/blue-widgets/>**

Alternatively, avoid using the filename `index.html` (it's not compulsory!) and only use meaningful filenames, for example:

**`http://www.yoursite.com/products/blue-widgets.html`**

### **3.6 Don't Use Frames**

Let's be clear: I'm not encouraging you to "avoid" frames. I'm giving you a direct order *not to use them!*

Frames inflict major findability and usability problems on websites. Any website can be – and most websites are – created without relying on frames.

Here are some of the main reasons for not using frames:

**Search engines will index framed sites incorrectly**, because two or more frame pages will share a single URL.

**Frames cause the browser's back and forward buttons to behave in ways the user does not expect**, causing confusion.

**The URL of a frameset is not useful as a web address**, as its constituent pages cannot be bookmarked.

**Frames break the notion of "place" within a website.** One of the main principles of website navigation is that the user should always know where he/she is within the site. This sense of context is typically sacrificed when frames are used.

## 4. Optimizing Individual Pages

Once you have laid the foundations of good overall site design and structure, as explained in the previous sections, you can begin to work on the nuts and bolts of optimizing individual web pages.

At the individual page level, you target the keyphrases in your repertoire. Those who like a challenge may thoroughly enjoy making individual web pages findable, as the process requires both discipline and creativity.

For those of you who simply find web page optimization hard work, don't worry: if you follow my tips, you *will* be rewarded with better search engine listings and increased traffic to your site.

### 4.1 Page Titles

Coming up with a page title sounds like a simple task, yet a good title is one of the most important factors in marketing a page through search engines, so it's important to take your time and get it right.

**The page title is the heading that appears in the blue bar across the top of the browser – not any other heading that appears within the page.**

You create the page title in HTML, by putting text between the <title> and </title> tags, as follows: <title>Example of a Page Title</title>.

Note that the Page Title must itself be contained somewhere within the header of the HTML page – the initial portion of the source code, which is enclosed by <head> and </head> tags.

In writing a page title, observe all of the principles explained in the following sections:

#### 4.1.1 Include the Relevant Keyphrase in the Page Title

Of all the places you should put your keyphrase on the web page, none is more important than the page title. Simply put, you *must* include the target keyphrase in the page title, otherwise you cannot consider the page optimized for that phrase.

For an extra boost, try to put your keyphrase as close as possible to the left of the title. Closeness to the left can help your performance slightly, but ignore this guideline if it impacts on your ability to make the title sound natural or attention grabbing to a searcher.

Some optimizers recommend using no words other than the keyphrase in the page title. This isn't always a feasible solution, as the keyphrase might be too short to serve as an attention-grabbing page title. But if this approach works in your case, go for it.

**Ensure that the keyphrase you use in the page title is consistent with the keyphrase you are targeting within the body of the page.**

#### **4.1.2 Keep it Short**

Google will display up to 67 characters of a Page Title, including spaces and any other symbols. If your Page Title exceeds 67 characters, Google will display as many whole words as it can under the 67-character limit, followed by an ellipsis (the “...” symbol).

Yahoo displays more characters, but its best to stick with Google’s guidelines, so that you will be optimized for both.

Consider the following examples:

```
<title>Online Guide to English Grammar, Spelling and Punctuation  
Markings</title>
```

– 66 characters. This Page Title will display in full in Google’s results.

```
<title>The Online Guide to English Grammar, Spelling and Punctuation  
Markings</title>
```

– 70 characters. This Page Title will display in truncated form in Google’s results, as shown here:

[The Online Guide to English Grammar, Spelling and Punctuation ...](#)

Note that by stretching the Page Title beyond the 67-character limit, we end up with only 60 characters in Google’s results.

One last tip: avoid excessive punctuation, as this uses up characters. A page title must be clear and understandable, but you can use poetic license. For example, always leave out the period at the end of the sentence.

#### **4.1.3 Grab the Searcher’s Attention**

Ultimately, you are writing your title for a reader, and trying to get him to click on it when he sees it in a list of search results. Readers don’t typically look at the blue title bar of a web page, so it’s only purpose, as far as you are concerned, is to grab attention in the search results.

A well-written page title will stand out from the poorly written titles of competitors (and there are *lots* of those) on the results page.

#### **4.1.4 Avoid Keyword Lists**

Many people try to stuff their page titles with lots of keywords, creating a list of words that won’t necessarily mean anything to the user.

Avoid writing page titles like this:

[Flowers bulbs seeds roses tulips carnations.](#)

Not only will this be meaningless to the reader (does it lead to a flower enthusiast's hobby page, a florist's site, or something else entirely?), it may also trigger a filter in Google's search algorithm that marks it as looking suspiciously "spammy".

Instead, try a title like this:

[How to plant roses, tulips, carnations and other flowers - online tips](#)

Or:

[Flowers Online: Seeds and bulbs available for roses, tulips and carnations](#)

The latter example is slightly better optimized for the keyphrase "flowers online," since it occurs closer to the left of the title.

Be creative with your titles; change them if they aren't working. But don't fall foul of Google – avoid lists.

#### **4.1.5 One Keyphrase per Page Title**

Since you have to keep page titles short, and meaningful to users, it's a good idea to concentrate on optimizing for a single keyphrase per page title.

A keyphrase can be two, three or even four words long. The longer it is, the more it will eat into your 67-character limit.

Instead of trying to cram several keyphrases into one page title, why not create a single page for each of the keyphrases?

This will take a little more work, true, but you are far more likely to be successful using this approach.

## 4.2 Meta Tags

Meta tags are used to denote metadata – information *about* a web page. Despite some popular myths, meta tags are not very important in the search engine optimization process.

Google largely ignores metadata, since many pages don't use them, and they don't necessarily reflect the content of the web page. Google's principle is that its results should reflect the content that web page visitors can see, not any meta or hidden information.

Yahoo has a similar policy, though it pays a little more attention to meta tags than Google. Some of the smaller search engines too take particular tags into account, such as the meta description tags.

Just to illustrate what meta tag code looks like, the following is an example of meta tags used in the header of a web page (that I chose at random) from the projectseven.com website.

### *Example of use of meta tags:*

```
<head>
<title>Dreamweaver Resources</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="Classification"
content="dreamweaver, templates, resource, experts, themes, extensions, tutorials, tips">
<meta name="KeyWords" content="dreamweaver templates, tutorials, web design themes, fireworks dreamweaver studio, design packs, dropdown menus, dhtml, javascript, macromedia resources, dreamweaver resources, dreamweaver behaviors, commands, objects">
<meta name="Description" content="Harness the full power of Macromedia Dreamweaver and Fireworks">
</head>
```

### 4.2.1 Meta Description

Google sometimes uses the information in the meta description tags as its own description of a page – usually if cannot find a “better” description on the page itself (this analysis is all performed by its algorithms, of course, not by human editors).

It is thus good practice to provide an accurate, attention-grabbing page description.

Write a description that will provide a good hook for your page, while including your keywords or keyphrase(s), but make it at least slightly different to your page title, which it will follow directly in the search results page.

The following are examples of useful page descriptions:

#### *For Page Title: “Lillie’s Bouquets – Florist, Jonesville”*

```
<meta name="Description" content="Florist in Jonesville, Michigan, with flowers for all occasions.">
```

***For Page Title: “Blue Widgets from WidgetWorld”***

```
<meta name="Description" content="Information and price list for blue widgets and other widget products available at WidgetWorld.">
```

#### **4.2.2 Meta Keywords**

One of the most common myths about search engine optimization is that it involves clever use of meta keywords. Nothing could be further from the truth. Meta keywords hold little or no value when optimizing for the main search engines.

**For the vast majority of web pages, Google ignores information contained in the meta keywords tags.**

Google only looks at meta keywords when all else fails – that is, if there is little or no text content on the page. A page with such low text content is unlikely to achieve high positions in Google’s search results.

Yahoo attributes a small degree of importance to meta keywords, but their inclusion is not necessary unless if you are optimizing for Yahoo on particularly competitive keyphrases.

If you are using the meta keywords tag to target Yahoo, make sure you only include words that also appear within the content of the page. Yahoo regards words in the meta tag that are not used elsewhere on the page as an indicative of spam.

Remember too to review your meta keyword tags whenever you edit the text content of a page, ensuring they are consistent with one another.

**A simple but effective policy on meta keywords is to include only your target keyphrase in the tags.**

## 4.3 Page Structure & Layout

When Google recommends that web sites should have “a clear hierarchy,” this applies to individual pages as well as overall site structures.

Layouts should be simple and uncluttered. Try to make the text items on the page flow sequentially, and avoid complicated designs.

### 4.3.1 Heading Tags & Keywords

A good approach to designing a structured page is to use HTML heading tags. Start with a single level one heading – a heading enclosed in <H1> tags – near the top of the page. Follow this with a series of level two subheadings, and include nested headings (level three, and so on) if necessary.

Using headings to structure your page will also give you an opportunity to insert your main keyphrase into a heading tag. It's best to insert the keyphrase into a size one <H1> heading.

In subsequent headings, avoid repeating any of the words from your target keyphrase. It is enough to use the keyphrase *once* in a heading – any more than that and Google might consider the page to be a little spammy.

You can combine <H> tags with CSS, to produce attractive headlines.

#### Example of the opening section of a structured page that uses keyphrases in heading tags, combined with CSS.

```
<HTML>
<HEAD>
<TITLE>Lillie's Flowers, Jonesville, Michigan - About Us</TITLE>
</HEAD>
<BODY>
<H1 class="brash">History of Lillie's Flowers in Jonesville</H1>
Lillie's Flowers has a long, established tradition in Jonesville.
We've been serving quality bouquets to Dubliners for over 60 years
now. Here's how it happened:

<H2 class="fancy">Moore Street Stall</H2>
Lillie's grandfather Joseph was well known for his popular stall in
Tulip Street, where he worked throughout the 1920s. Joe was famous
for his decorative bouquets, and was renowned for his stunning floral
displays, which were often used in the town's civic ceremonies.

<H2="fancy">A Family of Florists</H2>
Joe got so busy that he began to enlist the other members of the
family...
```

#### The accompanying CSS code:

```
.brash{
  color: #D2691E;
  text-transform: uppercase;
  line-height: 140%;
  letter-spacing: 0.3em;
```

```
font-family: Verdana, Arial;
font-size: large
}

.fancy{
color: #D2691E;
text-transform: capitalize;
line-height: 130%;
letter-spacing: 0.2em;
font-family: Book Antiqua, Verdana, Arial;
font-size: medium
}
```

### 4.3.2 Page Layout

Put the first instance of your target keyphrase as close as possible to the beginning of your HTML source code (after the opening <body> tag). Starting a page with a heading, as shown in the example code above, is a good way to do this.

For subsequent occurrences of the keyphrase, put a reasonable amount of text (a sentence or two, at the very least) between each occurrence. In the following section, I will explain what format subsequent occurrences should take.

Try to keep the layout of your page simple and ordered, with clear headings and paragraphs.

### 4.3.3 Table-based Layouts vs. CSS Layouts

The HTML <table> tag was originally intended to allow designers to include tabular data on a web page. However, designers found it allowed them to create grid-like frameworks on the page, and thus tables became the standard way of creating columns, sidebars, and other box-like features on a web page.

The use of tables is fine as far as Google is concerned – its bots won't have any problem crawling the page. However, tables will skew the apparent order of text elements of the page, making text items seem (to the bot) to appear in a different order than they actually do (to the human visitor).

For example, a heading might appear at the top of the page to the human visitor but, if your page uses a table-based structure, to the bot, it might appear further down the page. To get a better idea of how your web pages look to a bot, use a text-only browser such as Lynx, or use a Lynx emulator.

There are many Lynx emulators on the web that allow you to type the URL of any page, and see it in text-only format. A good one can be found at:

**<http://www.delorie.com/web/lynxview.html>**

Many web designers are now migrating away from table-based methods. The newer approach is to use CSS to create columns and other framing elements on the page. Implemented properly, a CSS solution allows you to give both human visitors and bots the same text, in the same order.

## 4.4 Keyword Formats & Density

The format and density (the amount of times they occur) of keyphrases within the page are important to how search engines, particularly Google, score that phrase.

Thankfully, formatting and controlling the density of keyphrases are two of the easiest aspects of optimization.

### 4.4.1 Formatting of Keyphrase Within the Page Content

By the time you get around to creating the content of the page, you should have included your keyphrase in the following ways:

- once in the page title
- once in the meta description
- once in a heading (preferably a size one heading, near the beginning of the page).

To complete your on-page optimization, add the phrase in the following ways:

- once in regular text
- once in **bold** text
- once in *italicized* text
- once in an [outgoing link](#) (see the next section)

In each of the above occurrences, try to ensure that the font size is not smaller than HTML size 2, or 12 pixels if using CSS.

As mentioned previously, try to leave at least one or two sentences between each occurrence of the keyphrase.

**It is better, but not essential, to keep the individual words of the keyphrase together in each occurrence.**

If you feel you can't use the keyphrase this often, without making the text seem contrived to the reader, try including the individual words of the keyphrase separately, but in close proximity.

### 4.4.2 Outgoing Links

Include the keyphrase once within an outgoing link – i.e. a link that leads to another page.

**Text that appears in an outgoing link is called anchor text.** Try not to include words other than the keyphrase in the anchor text.

It is important, too, that you have at least one or two more outgoing links on your page that point to related pages elsewhere on the web. By related pages, I mean pages that contain content similar to the content on your page. But only use the keyphrase as the anchor text in one of these links.

#### **4.4.3 Keyphrase Density**

If the overall text content of your page is low (less than 150 words, say), do not add the keyphrase in all of these ways described above. Leave out the italicized occurrence, or the bold occurrence, or both if necessary.

A very general rule of thumb is that your keyphrase should make up no more than 15% of the content of the page. (Don't worry about this figure too much – like I said, it's only a loose guideline.)

Your keyphrase density will be affected by any occurrences of the individual words that make up the keyphrase, even when they appear separately.

If your keyphrase is “gold rings,” then, and you have used this phrase in each of the ways described in the previous section, you must not inadvertently use the individual words “gold” or “ring” in other contexts within the page.

To check for such occurrences, there are some good free online tools. These keyword/keyphrase density checkers will also give you the percentage occurrence of phrases on any web page.

#### **Popular keyword density analyzers:**

RANKS keyword density analyzer, with placement and prominence analysis:  
**<http://www.ranks.nl/tools/spider.html>**

WebmasterWorld keyword density analyzer:  
**<http://www.searchengineworld.com/cgi-bin/kwda.cgi>**

## 4.5 The Art of Writing Optimized Text

Integrating keyphrases into web page content, with the appropriate frequency and formatting, would be a straightforward task were it not for one small but crucial snag: *human* readers will not enjoy reading text that has obviously been written for a search engine.

The text on a web page must serve, primarily, to meet that page's objectives. The page must convince the visitor to buy product x or try service y, or communicate other information as intended by you, the author.

In optimizing a web page, you must not compromise these goals. Thus, keywords should be integrated as seamlessly and invisibly as possible. Ideally, the content should read as though it could not have been better written *without* the keywords, or with any other formatting. This is a tall order, especially for inexperienced writers, and to do it well takes lots of practice – and lots of drafting, and redrafting!

Up to now, I have explained how to create a site that performs well in search results; in this section, I will introduce you to the art of *concealing* those practices.

### 4.5.1 Writing for the Web

Writing for the web is a specialized skill. Reading from the screen is not the same as reading from the printed page. The author also has to contend with links – ways for readers to leave the page before finishing.

When writing for the web, here are some broad principles to keep in mind:

#### ***Keep Pages Short ...***

Aim for an overall page size of no more than 150k. Users don't like to spend too long on any single web page. The sight of a dense, lengthy text page is enough to make a visitor find the nearest exit click immediately.

#### ***... But Not Too Short***

Don't over-compromise. A page must have at least some meaty content to entice the visitor to read. Bear in mind that users are happy to scroll down a page; it's scrolling left to right they hate.

#### ***Employ a Short, Clear Sentence Style***

Unlike the print reader, the web reader is in an active state of mind; is likely to be multitasking; and is unlikely to be committed to reading the page. As a writer, you must strive to retain your reader's attention.

Brevity, clarity and succinctness are crucial to keeping your reader locked in.

#### ***Use Bullet Lists and Other Scannable Features***

The visual cues on web pages also carry much more weight than they do in print pages, since web readers scan pages before they decide to read them, or sections of them.

Bullet-pointed lists are a great way of making a text item stand out, as are other simple features, such as:

- Clear headings
- Text boxes
- Neat paragraphs

Keep your paragraphs short; a single sentence may constitute an entire paragraph.

Be careful not to clutter your page; web pages should *not* look like PowerPoint slides!

### ***Check Spelling and Grammar***

An informal style doesn't mean that you should break the rules of spelling and grammar. Even the smallest mistake in either of these domains will make your page look careless and unprofessional.

Employ a spell checker, and ask a friend or colleague to proofread any text you intend posting to your site.

### ***Obey Universal Writing Principles***

There are many other general tips on writing that apply to *all* media, such as “omit needless words.”

Get your hands on a good book on writing style. I recommend *The Elements of Style* by William Strunk. First published in 1918, it is just as relevant today. This short book is available free online at:

<http://www.bartleby.com/141/>

## **4.5.2 Integrating Keyphrases into Your Writing**

Here is my four-step approach:

**Step 1:** Write the text of the page without concern for the keyphrase for which you are optimizing.

**Step 2:** When you have written your page, go back and try to insert each occurrence of the keyphrases, re-editing the text where necessary. Try each formatting method (heading, bold, italics) one at a time.

**Step 3:** Leave the page for a few hours or longer, then re-read it.

**Step 4:** If the text seems artificial, with too many repetitions of the keyphrase, remove one, or even two, of the less important occurrences (i.e. those in bold, italics and regular text).

The last step may impact slightly on your performance, but ask yourself which is better: to be second in Google's results for a particular query with a page that reads convincingly, or to be first with a page that reads poorly?



## 4.6 ALT and TITLE Attributes

### 4.6.1 ALT Attribute

The ALT attribute allows web designers to provide text descriptions of images. It is an attribute of the <IMG> (image) tag.

For example, the HTML code for a logo on a homepage might look like this:

```
<IMG SRC="logo.gif" ALT="WidgetWorld Logo">
```

ALT text is an important usability element of a website, as it benefits the following groups of users:

- Visually impaired users, who rely on screen-reading software
- Users of text-only or image-disabled browsers
- Users with slow internet connections (the ALT text appears while users are waiting for the image to download).

While Google recommends that ALT attributes are descriptive and accurate, **ALT text has little bearing on Google's web search results**. This is understandable; otherwise, spammers would stuff hidden keywords into ALT attributes.

### 4.6.2 ALT Text Combined with Anchor Tag

The text used in an ALT attribute becomes more important to Google where the image is used as a link. In such a case, Google associates the ALT text *with the linked-to page*, rather than the page on which the image resides.

For example, if your company logo doubles as a link to your homepage (a common convention on websites), you should put the keyphrase that you are targeting on your homepage within the image's ALT attribute.

If the target keyphrase was "blue widgets", the HTML code for each occurrence of the logo might look as follows:

```
<A HREF="http://www.widgetworld.com"><IMG SRC="logo.gif"  
ALT="WidgetWorld Logo, with Blue Widgets"></A>
```

### 4.6.3 ALT Attribute and Image Search

Google's Image Search function does take the words used in ALT attributes into consideration, as well as the text in the vicinity of the image.

However, image-searchers have different motivations to general information-seekers; and are less likely to engage with your site in the ways you intended.

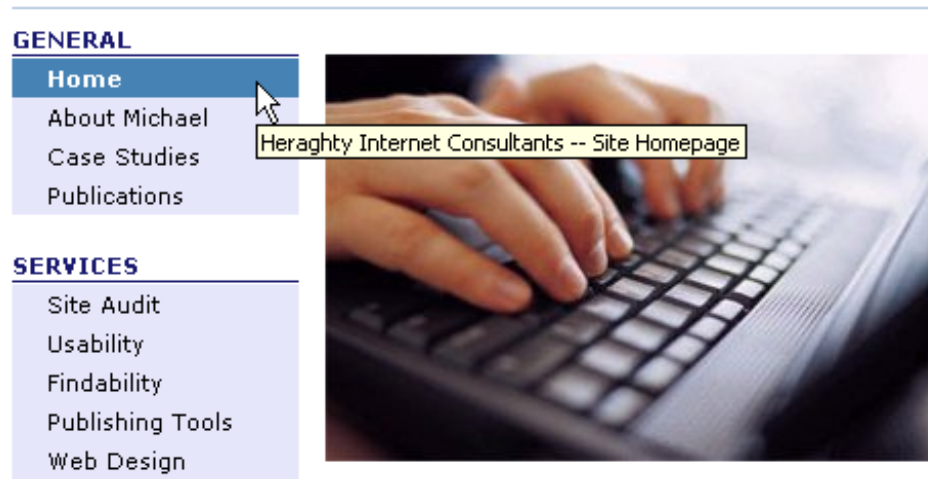
### 4.6.4 The Title Attribute

The TITLE attribute is used as part of the <A> or anchor tag to describe the destination of a link.

For example:

```
<A HREF="index.html" TITLE="Widget World Homepage">Home</A>
```

If the user holds the cursor over the link, the description used in the TITLE attribute appears to hover over the web page, as illustrated below.



*Example of Text in Title Attribute as it Appears on the Web Page*

The TITLE attribute carries a little weight with Google, but not a lot (probably less than the ALT attribute). However, every little helps, and it is a good practice to use Title tags.

## 4.7 Graphics vs. Text

Graphics are great for enhancing websites, providing you observe the principle that *less is more*.

Image files add to a page's overall size, which makes it slower to download (especially on a dial-up modem connection), and less likely to be fully crawled by search engine bots.

### **Avoid using graphics to display text!**

Search engine bots cannot read text contained within images. Yet a huge number of websites use graphics to display *the most important keywords* on the page (such as the company or product name). Typically, this mistake is made at the top of the page, where the company or product logo appears.

Try to separate the visual content of your company logo from its textual content. Even if you feel you can't render the company name as text, try to render the strapline – the one-liner that describes what your company does – as HTML.

Text straplines are often the best places to include the keyphrase for which, above all others, you would like to be found.



*Example of Keyphrase used in Strapline. As a rule of thumb, any text you can drag over with your cursor is HTML text, and is readable by search engine bots.*

Site menus are the second most common place where text gets rendered using images. When you create menus using graphics, you lose all of the optimization potential of the menus. (See Section 3.3, *Menu Structure & Style*, on how to create menus using CSS.)

## 4.8 Flash And Other Multimedia Elements

For multimedia elements such as video clips, audio clips – or anything else that whoops, whirrs or whizzes – here’s my advice: **just say no!**

The only reason you should ever include any of these multimedia elements is if your site depends on their inclusion. Only a handful of sites, usually those involved in music or film promotion, genuinely depend on these otherwise gimmicky features.

You won’t be able to optimize Macromedia Flash files (a common and popular format) or other multimedia elements on your site. Google has only a very limited ability to read text from flash files, and can’t index any of the words that are used in video or audio clips.

## 4.9 JavaScript and Other Client-Side Languages

While the use of JavaScript isn't as problematic as the use of multimedia, I would urge that you avoid using this language, unless you feel it is necessary, from a user experience perspective.

Avoid using JavaScript to display text – especially if that text contains your target keyphrase – on your web page; search engine bots may not be able to read this text.

JavaScript may also bloat your page size unnecessarily. Where you are using a lot of JavaScript for essential reasons – to validate form fields, for example – put it in an external “.js” file, thus reducing page size.

JavaScript is the most popular *client-side scripting language* – which means code that performs a function on your local machine, not on an internet server. But it is not the only one; others include DOM and VBScript. You should generally try to avoid using any client-side scripting languages. Stick to HTML and CSS.

## 5. Developing Links with Other Sites

While optimizing the text on your web page is essential to getting found for search queries, the position the page achieves in search results is heavily dependent on the number and relevance of its incoming links. The number and quality of links required rises heavily for search phrases that are more competitive. For example, travel related phrases require a lot more links from quality travel-related sites, than would be required for other, less competitive subjects.

If search engines didn't exist, it would still make sense to develop links with other sites. One of the main ways visitors will find your site is by clicking a link on another site, particularly a site that has similar or related content. A link is like a recommendation.

Building links is one of the most challenging aspects of search engine optimisation – and one that will take much of your time and energy.

It is not advisable to simply exchange links with as many other sites as you can. Wanton link exchanging may lead to your site performing poorly in search engines, or even being penalized.

In this chapter, then, I will explain why and how to foster links with other sites, and when to say no to a link exchange offer.

### 5.1 Why Your Site Needs Links

As I explained in Chapter 1, PageRank is Google's way of measuring the popularity, and thus the importance, of a web page. In order for a page to increase its PR, it must increase its number of incoming links, or *backlinks*.

Backlink quantity is important for gaining PR, then, but there are many other factors – mostly pertaining to pages from which the backlinks originate – that influence how your web page will perform for a given search query.

#### 5.1.1 Links from Pages with Higher PageRank

Not all links are equal!

A link from a page that has PR1 is worth less than a link from a page with PR2, which in turn is worth less than a link from a PR3 page, and so on up the scale.

Pages are said to “transfer” some of their PageRank to the pages they link to.

The general rule of thumb is as follows:

*A link from a page with a PageRank of X, if it has no other outgoing links, will confer a PageRank of X-1 on the linked-to page.*

This math of PR transfer is not always this straightforward; there are many other factors that Google's algorithm takes into account when assigning PR, as we shall see.

Nonetheless, the basic idea is that a single link from a page with PR5 should be enough to give your page a PageRank of 4. On the other hand, even 20 links from PR1 pages would not be enough to give your page a PageRank of 4.

**Try to get links from pages that have higher PageRank than yours.**

But do not refuse links simply because they are from low PR sites! All backlinks will increase your PageRank, even if only modestly. And since the PageRank of all pages is subject to change over time, a link from a PR2 page today may turn out to be a link from a PR3 page next week.

**Try to get links from as many different sites as possible.** A set of ten links from different sites is worth far more than a set of ten links from pages within the same site, assuming all the links are from pages that have the same PR.

If you want to find out what pages are linking to your page, do a search on the URL in Google, preceded by “link: ” (without the quotes, and with no space). For example, **link:www.yoursite.com**



**Web** Results 1 - 10 of about 254,000 linking to **www.amazon.com**. (0.10 seconds)

[iLoveLanguages - Your Guide to Languages on the Web](#)

iLoveLanguages is part of the WWW Virtual Library: Education > Languages. ...

[www.ilovelanguages.com/index.php?category=Commercial](http://www.ilovelanguages.com/index.php?category=Commercial) - 75k - [Cached](#) - [Similar pages](#)

[PHP-Nuke](#)

Main Menu. ...

[phpnuke.org/modules.php?name=Commercial\\_License](http://phpnuke.org/modules.php?name=Commercial_License) - 19k - [Cached](#) - [Similar pages](#)

[Davis Cup Tennis Schedule, Statistics & Information](#)

DAVIS CUP Tennis Schedule, Statistics & Information. 2004 WORLD GROUP

DRAW. First round Feb. 6-8, 2004: Sweden at Australia Austria ...

[www.tennis-x.com/tournaments/daviscup.shtml](http://www.tennis-x.com/tournaments/daviscup.shtml) - 73k - [Cached](#) - [Similar pages](#)

*Google's results for links to Amazon.com*

**Notes on Checking Your Backlinks with Google:**

- Google does not display *all* of your backlinks – it generally leaves out links from pages that have PR of 3 or less.
- The majority of pages on the web have a PageRank of 3 or less.
- Google only updates its lists of backlinks once every two months or so.

### 5.1.2 In-Site Links: The Pyramid Structure

It is extremely important that you link together the pages within your site. In the section on Section 2.2, *Site Structure and Integration*, I discussed ways to ensure your site pages are well interlinked – such as using a good menu system and site map.

You can enhance the performance of keyphrases within your site by directly linking pages that share similar themes or topics.

Some SEOs recommend building *theme pyramids* – hierarchies of pages, at least three levels deep, that share the same theme. Each page in a pyramid should point to its parent and children pages in the pyramid.

In this strategy, pages higher up in the pyramid (for example, the homepage – which is at the very top of the pyramid) are used to target general keyphrases (e.g. widgets), while pages lower down are used to target more specific keyphrases (e.g. “green widgets” targeted on a page in level two, while “green triangular widgets” targeted on a page in level three; etc.).

The pyramid structure is useful and logical, for website maintenance and organization as much as SEO purposes, but it is not essential. Whatever way you structure your site, make sure that its pages are well interlinked.

### 5.1.3 Note on Relative Links vs. Absolute Links

An absolute link is a link that contains a complete URL. For example:

```
<A HREF="http://www.widgetworld.com/about-us/contact-details.html">
```

A relative link contains only the path to the linked page, relative to the referencing page, for example:

```
<A HREF=" ../about-us/contact-details.html ">
```

SEO's argue about which form of link is “better” for optimization, though no consensus has been reached.

My opinion is that any advantages one method may have over the other are trivial. Google's bots will be able to spider the pages on site *regardless* of whether you use relative or absolute links. Go with the method you personally prefer.

### 5.1.4 Avoid Superstar Pages

Sites will typically contain one or more “superstar” pages – pages that get far more backlinks than other pages in the site. The homepage is often the superstar pages on any site.

The problem with superstar pages is that they can soak up PageRank, which could otherwise be spread across your site. So, if one of your pages has a PR7, this doesn't guarantee that any of the other pages on your site will have a high PR. It is not

uncommon for a site homepage to have PR7, with all the other pages in the site having PR3 or lower.

You can improve matters by linking the high-performing page with as many of the other site pages as is feasible.

Another good strategy is to get backlinks to the other pages in the site – not just to the superstar page. Instead of asking another site to link to your homepage, try to get them to link to a specific page within your site (for example, an article you have written).

The more pages that achieve high PR via external backlinks, the more the PR will spread throughout your site. **Make your pages perform as a team, not as isolated individuals.**

## 5.2 Good Links, Bad Links

Google's explanation of PageRank – that links to your site count as votes – turns out to be a little oversimplified.

Not *all* votes are counted. There are certain types of links of which Google doesn't approve. In some cases, Google will simply ignore these “bad” links, and they will not influence your PR. Worse, certain links can act as a vote *against* your site – giving you a “thumbs down”, as it were.

Worse again, links from sites that use spam techniques (see Section 7.1, *Techniques to Avoid*), or that come from non-relevant pages, can get your site de-listed for searches on the very keyphrases you are trying to target.

In the following sections, I will explain how to distinguish between useful, neutral and harmful links.

### 5.2.1 Relevant Links

While any link to your page is likely to boost its PR, links that are relevant to your target keyphrase will boost your page's performance in searches on that keyphrase.

Consider a page that is optimized for the keyphrase “Jonesville florist”. A link from any page with a high PageRank (PR4 or more) is likely to boost the performance of the Jonesville florist page in a Google search.

But a link from a page that has a similar theme to the keyphrase – a page about flowers, about florists, about Jonesville, or – best of all – about florists in Jonesville, is even more valuable.

Such a link is called a “relevant” link, and it will specifically enhance the performance of the keyphrase for which your page is optimized.

### Avoid Non-Relevant Links!

Google frowns upon non-relevant links. If your backlinks come mostly from pages whose themes have little in common with your keyphrase, your page may drop in, or disappear from, Google's search results for that phrase.

Note that Google uses software to determine whether the content of one page is relevant to another. Google employees do not normally intervene in this process, except when investigating suspected spam or cheating.

### 5.2.2 Links from Expert and Hub Sites

Even better than a relevant link is a link from a site that is considered by Google to be an authority on the subject referred to in your keyphrase. SEOs call these *expert* sites. Where reputable sites contain many links to other sites on a shared topic, they are called *hub* sites.

Google doesn't reveal how it decides which websites are expert or hub sites, although we can be sure (based on other statements) that the company does not use human editors to decide whether certain sites qualify for expert status.

You can identify many expert and hub sites by taking the following steps:

1. Search Google for the keyphrase for which you want to optimize (e.g. blue widgets).
2. For the top ten search results, check the *backlinks* to each of the sites in the list. (To do so, type the URL of each site into Google, preceded by "link:" For example, "link: http://www.widgetworld.com ).
3. Make notes of the top 5 backlinks for each site.

At the end of this process, you should have a list of about 50 sites. Is there a site that is figuring more prominently than others in this list?

If so, chances are that this is an expert or hub site. If you manage to get a link from this site, it will enhance your chances of performing well for the keyphrase.

### **5.2.3 Links Beyond 100**

I have already mentioned that PR gets transferred from one page to another, such that a link from a PR(X) page is worth PR(X-1).

The bad news is that this is true only if there is a small number of outgoing links on the PR(X) page. The more links that appear on any page, the more the value of its outgoing links will be diluted.

When looking for pages to provide links to your site, try to choose those that have as few outgoing links as possible.

If a page has more than 100 outgoing links, Google will ignore all links after the hundredth.

### **5.2.4 Sponsored Links**

As a general rule, you shouldn't have to pay for links to your site. After all, hyperlinks are the essence of the web – they are one of the main things that make websites different to print and other media. Create a good, quality site, and lots of webmasters will be happy to link to it, at no cost.

Some SEOs nevertheless pay for links, particularly from sites with high PageRank. Google frowns upon this process, and has taken measures against it in the past.

Many of those who offer paid links are in fact engaging in spam techniques – so the link may turn out to be worthless.

Even if the person you are paying for the link runs a reputable site, your link may be labeled as "sponsored," and/or be placed on an obscure part of the page. In these

cases, visitors are less likely to click through to your site – which, after all, is the primary reason for getting links; traffic doesn't come from Google alone.

Your link could be placed on a page that contains lots of outgoing links, thus diluting the PR value of the link.

Concentrate on trying to encourage links from other sites without paying, by making your content as interesting, entertaining and useful as possible. If you do have a budget to pay for links, and are eager to spend it, consider using it for Google AdWords instead.

### **5.2.3 Reciprocal Links vs. One-way Links**

Reciprocal links are less valuable than one-way links. From Google's point of view, a one-way link is an unsolicited vote of approval for your site, while reciprocal links tend to indicate agreed exchanges.

Reciprocal links do have value, just not as much as one-way links. But you will find it easier to get reciprocal links than one-way links, so don't turn either form down.

### **5.2.4 Cross-Linking Among Sites**

One way to get more links is by creating more than one site. With two sites, you can obviously link one to the other; with three sites you can link all three together, etc.

However, there are a few drawbacks to this approach.

If you link among sites that you own, try to ensure that the links are not reciprocal, or otherwise cross-linked.

For example, imagine you own three sites, called A, B and C. Site A links to Site B, and Site B links to Site C, and Site C back to Site A. You may regard this as a clever way to avoid using reciprocal links. However, the link pattern described above is a form of "cross-linking" known as triangular linking.

Search engines can easily detect triangular linking, or other crude cross linking strategies, and they will recognize these links as being similar to reciprocal links, and will reward them less than independent, one-way incoming links.

Google penalizes blatant cross-linking among a network of sites, particularly if such sites share IP addresses or have other associative features. Google does not approve of affiliations that are created solely for the purpose of increasing PageRank or otherwise "gaming" its system, and is continually updating its map of the web in search of unusual or incestuous linking patterns. Yahoo and Microsoft are increasingly using similar techniques, and they too frown upon this type of incestuous linking.

As the number of sites you control grows, it becomes more difficult to cross-link in such a way that they do not appear enmeshed. Try not to make your sites seem part of one large "neighborhood."

### **5.2.5 Links from Links Pages**

Often, when someone agrees to create a link to your site, they will add it on their “links” page.

Getting a backlink from a links page is sometimes desirable, sometimes not. If the links page has a high PR, and links to sites that appear highly in searches relating to the topic of interest, then the links page might be considered a “hub” or “expert” resource by Google. A link from such a page is valuable.

Hub pages tend to have good descriptions after each of the links, for example:

#### **Widget World**

Useful site about widgets, with succinct descriptions and colorful pictures.

<http://www.widget-world.com>

#### **Joe's Widgets**

Joe Soap's personally updated widget bible; everything you ever wanted to know about widgets but were afraid to ask.

<http://www.joeswidgets.com>

Google does not assign value to links from pages that are little more than lists of clickable URLs:

<http://www.widget-world.com>

<http://www.joeswidgets.com>

<http://www.widgets2go.com>

Or clickable descriptions:

[Widget World](#)

[Joe's Widgets Page](#)

[Widgets 2 Go](#)

Of course, you can't control the way others create their links pages – but this information may be useful to you when identifying sites from which you would like to get links.

If the links page does not look like a hub or well-structured links page, it will be more valuable to get a link from a regular content page. Regular content pages tend to have higher PR and fewer outgoing links than links pages.

Beware of links pages that look spammy – especially those that do not seem to be part of sites with good content. Remember that Google does not even follow any links beyond the first 100 on any given page.

## 5.3 Anchor Text

In the section on outgoing links, I explained that anchor text describes the words that form a link – words that, on the web page, are clickable. In HTML code, anchor text appears within <A> or anchor tags.

Here is how the anchor text “advanced widgets” might look on a web page:

Find out more about the history and design of [advanced widgets](#).

The HTML code for this sentence would look as follows:

```
Find out more about the history and design of <A  
HREF="http://www.widget-world.com/advanced-widgets.html">advanced  
widgets</A>.
```

### 5.3.1 Anchor Text and Keyphrases

Anchor text is one of the most important off-page factors in influencing the performance of your keyphrase in Google’s results.

Try to include your exact keyphrase in the link; avoid including additional words.

In the example I gave in the previous section, notice that I made “advanced widgets” the anchor text and not “Find out more” – which many webmasters would have opted for.

#### **Appropriate Anchor Text:**

Find out more about the history and design of [advanced widgets](#).

#### **Inappropriate Anchor Text:**

[Find out more](#) about the history and design of advanced widgets.

Using your keyphrase as anchor text in at least one link to your site is a powerful way of influencing the performance of your keyphrase in Google’s search results.

There are also good usability reasons for making your keyphrase the anchor text. “Find out more” is not very informative on its own, and the user has to read further on in the sentence to see what she is going to find out more *about*. By presenting the link in this way, you are asking to reader to hold a “marker” in her head, while reading on. Chances are your reader will become annoyed with this needless challenge, especially when sentences are long.

(Web designers make this type of mistake all too often. The classic example is the [click here](#) link, which reveals nothing about the destination page.)

The more relevant and higher the PR of the page containing the link, the more valuable it is to get your keyphrase into the anchor text of that link – especially if it is a hub or expert site.

### 5.3.2 Anchor Text Within Your Site

You may find it difficult to get your keyphrase into the anchor text of pages that link to yours. After all, to request a link is one thing, but to specify the exact nature of the text you want to include in the link is another (but don't let that stop you from trying!).

Nevertheless, there is one set of pages from which you can exercise complete control over your anchor text – pages within your own site!

Anchor text in links from pages within your own site is highly valuable in optimizing for your target keyphrase.

This is yet another reason why you should use text-based menus, which allow for many links from within your site. See Section 3.3, *Menu Structure and Style*, for more on text-based menus.

### 5.3.3 Anchor Text Variation

Not all of the backlinks to your site have to contain your keyphrase as anchor text. Indeed, there is evidence to suggest that Google now regards too much consistency in anchor text as deliberate manipulation.

**Warning: If all or most of your backlinks contain exactly the same keyphrase, you may drop or disappear in Google's results for that keyphrase.**

For every three backlinks you get, try to ensure that at least one of them does not contain your keyphrase.

## 5.4 How to Get Links

Getting links to your site involves interacting with other people who own or maintain websites with content that could be considered relevant to yours.

Obtaining good, quality links may take hard work and persistence – but such links are important to your success in Google, especially if you are optimizing for competitive keyphrases.

Before you embark on a linking strategy, make sure you have created an attractive site with well-written, unique content. It will be easier to get links if your site appears worth linking to – and such a site is more likely to generate unsolicited links! This is not an unusual occurrence; unsolicited linking is the essence of the web – anyone can link to anyone else.

To find out what sites are linking to yours, you should use a good statistics software package on your server (as well as checking your backlinks in Google, using the *link:* prefix, as explained earlier). Many hosting providers offer such statistics as standard. A good statistics package will list all referring URLs to your site, on a daily or hourly basis.

The stats package will only show you links that have actually been clicked (whether by a human visitor or a search engine bot) to get to your site. Links that exist but are not clicked will not show up in your site statistics.

*See also Section 6.1.3, Site Statistics Software.*

### 5.4.1 Link Exchange Request

One of the most common ways that site creators get backlinks to their sites is by exchanging links with others. This practice is called reciprocal linking.

The great thing about reciprocal links is that you can get them from complete strangers, after a brief email exchange.

A good approach is to identify sites similar or related to yours, then email (or better still, call) the person responsible for the site, requesting an exchange of links. Courteousness and sincerity are all important in such emails, as requests for email exchanges with dubious partners are unfortunately frequent.

Indeed, you should be cautious of any such requests that you receive. Check out the sender's site and satisfy yourself that it is genuine. Beware of Google-cheats! A link from such a site could potentially harm your site. Make sure, too, that the sender's site is relevant to yours.

### 5.4.2 Directories

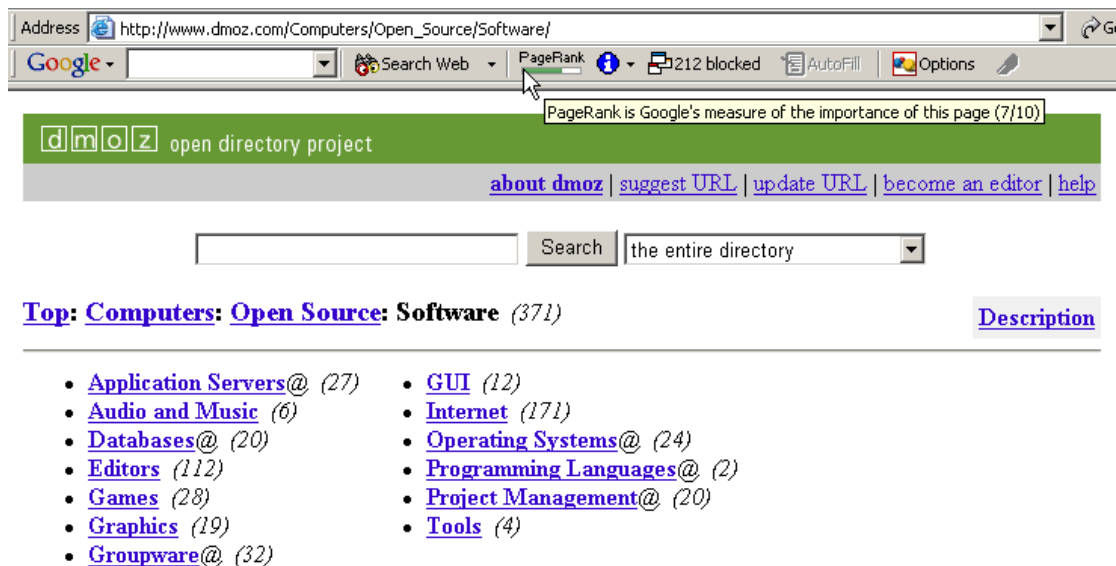
Directories are sites devoted to providing an organized set of links to other sites. Directories have a hierarchical structure of categories and sub-categories, and are often themed, or restricted to sites from a geographical region.

Links from popular directories are valuable, particularly if Google regards them as hub or expert sites.

Before submitting your site to a directory, you will probably find multiple categories under which your site is eligible to be listed. Have a look at the relative category pages and, when submitting your site, try to ensure that your site's link will appear on the page that has the highest PR.

Higher-level category pages typically have higher PR. For example, it may be more useful to get listed under the [Top > Business](#) section of a directory than under the [Top > Business > Widgets > Widget Makers](#) section, which is likely to have pages with lower PR.

If listings are alphabetical, you may find yourself on the last page – with the lowest PR – if your company begins with a letter that occurs lower in the alphabet. (Widget World, beware!) If your company name begins with an alphabetically-challenged letter, try to get into a category that isn't crowded.



Address [http://www.dmoz.com/Computers/Open\\_Source/Software/](http://www.dmoz.com/Computers/Open_Source/Software/)

PageRank is Google's measure of the importance of this page (7/10)

**Top: Computers: Open Source: Software** (371) [Description](#)

- [Application Servers@](#) (27)
- [Audio and Music](#) (6)
- [Databases@](#) (20)
- [Editors](#) (112)
- [Games](#) (28)
- [Graphics](#) (19)
- [Groupware@](#) (32)
- [GUI](#) (12)
- [Internet](#) (171)
- [Operating Systems@](#) (24)
- [Programming Languages@](#) (2)
- [Project Management@](#) (20)
- [Tools](#) (4)

*In this example, a link from the Top: Computers: Open Source: Software page (PR7, above) is worth more than a link from the Top: Computers: Open Source: Software: Audio and Music page (PR5, below).*



Address [http://www.dmoz.com/Computers/Open\\_Source/Software/Audio\\_and\\_Music/](http://www.dmoz.com/Computers/Open_Source/Software/Audio_and_Music/)

PageRank is Google's measure of the importance of this page (5/10)

**Top: Computers: Open Source: Software: Audio and Music** (6) [Description](#)

When submitting your site description, include the keyphrase.

Be aware that spam sites often pose as directories. Exercise caution in choosing directories to which you submit your URL. You may be contacted and asked for a payment after submitting for what you thought was a free listing. If asked for a payment, it's best to decline.

Concentrate on getting listed in free directories – there are many free directories that Google regards highly, particularly regional ones. Ignore any unsolicited emails offering directory listings.

### **5.4.3 The Open Directory Project**

One site with which Google has a special relationship is the Open Directory Project (ODP) – <http://www.dmoz.org>

Google relies heavily on the ODP for its own directory – <http://directory.google.com>

Sites listed in Google's directory are different to sites that appear in its regular search results, because human editors review and approve each of the directory entries.

It is worthwhile submitting your site to the ODP, which will review and list it for free. The only drawback is that the review process can take months – even a year or more. From there, it will take several months more for it to enter the Google directory, but once again inclusion is free.

Before you submit your site to be reviewed by the ODP, have a look the pages of each category relevant to your site, then check the PR for each of these pages. Submit your site to the category page with the highest PageRank.

### **5.4.4 Contributing to Other Sites**

A great way to get quality, relevant backlinks is to write an article, or otherwise contribute to informative websites. No matter what the subject of your site, chances are there are one or more sites devoted to the same subject that would welcome contributions.

Usually these are small, independently run, industry-related news sites, weblogs, or community sites.

Try to identify such sites. Many sites openly request contributions; they are typically run by individuals or small groups, who do so in their spare time, and are hungry for content. Even a request for content doesn't appear explicitly on the site, does it contain articles or other contributions from third parties? If not, is it regularly updated? Chances are the author(s) will be glad of a little help!

Start by coming up with an idea for an article. Perhaps even write a brief, two- or three- line synopsis. Stick to the subjects you know best – just as you did when coming up with content for your own website. Perhaps you can re-purpose some

material that you've already used on your own website, or for your own offline marketing purposes?

Email the owner of the site on which you'd like to get published with your proposal, and see if they are interested.

Of course, request that your article is accompanied by a short biography and a link back to your site. Most site owners will be happy to do so – but get this confirmed in advance.

A link from this type of site tends to be highly valuable, especially as sites that have lots of content and are often updated tend to have good PageRank, or even to be regarded as expert sites. Again, if a site agrees to publish your contribution, make sure you get a bio and link back to your site – and make sure any such link is not confined to a members-only area of the site (unreachable by Google's bots).

If you write one good article, you may be able to re-purpose it and get it published on another site.

There are additional ways you can contribute to other sites, such as providing news tips, becoming an editor, or becoming a forum moderator. Be warned that the latter two methods require a serious commitment, which you should make only if you have a genuine interest in the role.

#### **5.4.5 Responding to Link Suggestion Requests**

Many sites openly request suggestions for links. Usually, this means they want to provide links to sites with similar content, or which may be useful to their visitors. The great thing about getting a link from this type of site is that it is likely to be a one-way link from a relevant site, possibly even a hub or authority site (if it already has outgoing links to similar sites). As I explained previously, this is the best type of link you can get.

To find such opportunities, try searching for “*your-subject suggest a link*” or “*your-subject suggest a site*”.

If your site has quality content and offers information that would be valuable to readers interested in the site requesting suggestions, you are much more likely to see your link included.

Also, try to personalize the link suggestion. Tell them why you think your site would be a good match. If possible, email the site owners directly. If you have to use a form, use the field that asks for “other comments” or similar (if there is one) to include a personal message. Try to show that you have actually visited and read through the recipient's site. The personal touch leads to a higher success rate than suggestions that only contain linking information.

**Warning Note:** Beware sites that request URL suggestions, but which look for a link in return – particularly sites that ask for you to link *first*. As with many of the email link exchange requests, the site in question may never link back to you. Even if they do link to you, they will probably also link to many other sites – including sites of

dubious quality. The quality of any link you receive may therefore be diluted, and may even hinder your ranking efforts.

#### 5.4.6 Linkware

*Note: Readers with some programming and web development skills may be interested in the concept of linkware. For regular web users or designers, it's probably not an option, so you may want to skip over this section.*

Most readers will be familiar with the concept of “freeware” or “shareware” software, which is available to use for free, or for a small donation (if you find it useful). Usually these are small software programs such as desktop utilities; screensavers; plug-ins for larger programs; etc. – which are unlikely to make much money if they were available for sale. So, the makers distribute them freely (sometimes with a donation request) in order to “get their names out there”.

In recent years, linkware products have begun to proliferate. Instead of donations, these products are distributed in return for links from the users' websites.

Linkware products tend to be web-based applications – this ensures that the target market actually have websites, so they can contribute the links. In fact, most linkware products actually build the link in to the applications they distribute, and request that – as part of the “terms of use” agreement – the user does not remove the link from the application when using it. Examples of linkware products include website or weblog “skins”, statistics counters; JavaScript utilities; etc.

The great thing about a linkware product is that, if it's good, it will also become a “viral marketing” tool, with people noticing it on other sites and then downloading it for themselves.

#### 5.4.7 Blog Comments

Another popular type of site that allows users to post URLs is a weblog, or blog – a diary-style site. The blog is usually maintained by a single individual, and updated regularly.

Many weblogs allow readers to post comments on each of the individual entries. Comments are automatically published to the site. As well as submitting your name and comment, most blogs allow you to post your URL.

Google does not take the same stance with weblog comments as it does with guestbooks – the link you post *may* count as a valid backlink.

**Do not post a comment that appears solely designed to produce a link to your URL. This is known as *comment spam*.**

Bloggers are typically web-savvy individuals. If you engage in comment spam, the blog owner will probably delete it, or add you to the weblog community's blacklist of spammers.

He may even report you to Google or Yahoo, who encourage such reports, and penalize persistent spammers.

On the other hand, most bloggers welcome insightful, engaging comments, so weblogs can be a great way of generating links to your site.

When commenting on weblogs, observe the following guidelines:

- **Read the individual entry, then make a relevant comment.** Spammers rarely read weblog entries, so their comments are usually non-relevant. Bloggers loathe false, generic statements, such as “Great site!”
- **Comment on a relevant topic.** Is the theme of the entry similar to that of your own page? Non-relevant links may decrease your page’s performance. Anyway, bloggers prefer comments from those who share their interests.
- **Comment on a recent article, not one deep in the archives.** For each new comment, the blogger receives an email notification. If you reply to an entry that is months or years old, you invite suspicion.
- **Enter your name in the name field.** Stating the obvious? Perhaps, but entering “keyword1 keyword2” instead of your name is a sure sign of comment spam. Play it safe and let your name become the link’s anchor text.
- **You don’t necessarily have to link to your homepage.** Enter the URL of a specific page on your site – but *only* if that page is relevant to the entry on which you’re commenting.
- **Return to the weblog in the days following your comment.** Treat the weblog owner with respect. Visit again to see if she has replied to your comment. A commenter who never returns may be seen as a spammer.

#### 5.4.8 A Note About Guestbooks

There are certain types of links that you can generate automatically on a web page, by typing your URLs into a certain types of online form. Google strives to identify, and thus attribute zero importance, to such automatic links. Other search engines are following suit.

The most popular and prevalent type of auto-generated link is the link that is created on a website guestbook page. Along with your name and location, many guestbook forms prompt you to enter your URL. This URL will subsequently appear as a link on the guestbook entries page.

Signing a guestbook will neither harm nor improve your PageRank – Google ignores guestbook links.

## 6. Maintaining and Increasing Findability

### *The Waiting Game*

So, how long before you start to see your site's performance improving in the search engines?

Whether you are optimizing an existing site or launching a new site or, you should begin to notice results (an improvement in your position, or a first appearance in the listings) within weeks, in Google at least – providing you have followed my guidelines closely.

Be advised that it may take time to get *great* results – such as a top ten position in Google or Yahoo. The time it takes depends on

- 1) Your proficiency in creating and adding optimized content to your site
- 2) Your ability to get the right kinds of links
- 3) The strength of the competition

For those launching new sites, obtaining backlinks is the most difficult aspect of search engine optimization, particularly when competing with sites that are more established.

Getting listed in directories – particularly the Open Directory – can take a long time too.

For more competitive keywords, it is going to take longer to get good results in Google's listings. Niche phrases will get quicker results than common phrases.

If your pages aren't showing up in good positions in Google within a few weeks, don't despair! Just keep following the guidelines. Success will come eventually.

### *Don't Stop*

Once you achieve high positions in the search engine results for your target keyphrases, your work is done, right?

Wrong! The scramble for places in search results grows more competitive by the day. You may be in the top ten today but, if you are complacent, your competitors may soon displace you, by increasing their own optimization efforts.

Never regard your site as finished. Think of it as a living, organic entity that needs ongoing care and attention.

By acting now, you can achieve a great platform from which to build in the coming months and years. With dedication and persistence, you will not only improve your existing search engine results, you will continue to grow the repertoire of keyphrases for which you perform well in searches.

## 6.1 Site Statistics

To monitor your website's performance, you'll need a robust site statistics package. A decent stats package will give you the detailed, up-to-date information you need about how visitors found your site, and how they interacted with it.

### 6.1.1 Hit Counters Don't Count!

Site statistics packages are not the same as hit counters. The latter are feeble pieces of software that can only give you limited information about your site's visitors.



*Examples of Hit Counters.*

Hit counters don't analyze the log files on your website's server. Instead, they connect to a third-party site that remotely calculates simple (and not always reliable) traffic figures.

Counters usually take the form of gimmicky-looking images. They allow your visitors – and competitors – to see your traffic figures. Hit counters are okay for amateur sites, but not for business or professional sites.

### 6.1.2 Site Statistics Software

A statistics package will analyze your site's log files – the records the web server keeps of each visit to your site. It is difficult to interpret the data in raw log files. A good statistics package will extract the information and present it to you in neat, visual formats.

There are many good software packages available, and many hosting providers offer a stats package as part of the hosting package. When choosing a hosting provider, find what stats software they offer. Software packages should be included in the normal hosting price.

If you run your own server, you should purchase and install statistics software, or install a robust freeware solution (such as AwStats).

Popular stats package choices among webmasters include:

- Urchin
- Analog
- Net Tracker
- Webtrends
- Hitslink
- Webalizer
- Awstats
- FastStats
- ClickTracks

Each of these packages is different, and many offer specialist functions. For example, FastStats is designed for sites that get large volumes of traffic.

Note that any statistics based on log file analysis are rarely accurate, for technical reasons (such as the use of proxy servers; users sharing computers or IP addresses; and so on).

Assume a margin of error of around 10%, or up to 40% if your site gets high-volume traffic (i.e. over 100,000 visitors per day).

Below are the features you should look out for in a stats product, and what they mean. Note that not all of the packages listed above offer all of the features below, and that many packages have “pro” versions, which offer more features.

I have listed these features in what I consider to be their order of importance.

### ***Number of Unique Visitors per Time Range***

The most important of all your stats is how many *people* visit your site (not how many “hits” or “bytes” it gets).

You will want to know how many visitors your site received per day, per week, per month, per year – or during a date range that you specify.

An individual or *unique* visitor may return to your site more than once during a specified time range. Most packages will give separate stats for the number of visitors and the number of visits.

Good statistics packages will also identify search engine bots such as Google’s bots, or Yahoo’s Slurp. A stats package that doesn’t recognize bots will give you bloated visitor figures.

### ***Number and Breakdown of Search Engine Referrals per Time Range***

You will definitely want to know what keyphrases were entered into search engines to find your site, within any given time frame. A good package will be able to identify which of the keyphrases were entered into which of the search engines.

For example, you should be able to see that on July 2<sup>nd</sup> of this year, 6 people found you in Google.com for the keyphrase “white widgets”, 4 people found you by entering the same phrase in Google.co.uk, while 2 people found you by typing “widget world” in Yahoo.com.

### ***Number of Page Views/Duration of Visit per Keyphrase***

For each of the users that found you through a search engine, it is useful to know how many of your site pages they looked at, and/or how long they stayed on your website.

If users who search for certain keyphrases don’t stay on your site for long, or don’t visit many pages, you may have to consider changing your site, or targeting more relevant keyphrases.

### ***Details of All Referrals per Time Range***

For any time range you specify, you will want to see where your general traffic (not just your search engine traffic) comes from. If they came from another site, you will see the URL. This information will allow you to check for new backlinks, and assess how your existing backlinks are performing.

Many of your site visitors may show up in your statistics as having “no referrals.” These visitors probably found your site by typing its URL directly into the browser or by added it to their bookmarks or favorites.

### ***Bot Identification***

A good statistics package can easily identify all bots that visit the site. When you launch a new site, or make changes to your site, you will want to know that Google’s bots have visited the relevant pages.

As mentioned earlier, a good package will be able to tell the bots from the human visitors, and will separate the statistics accordingly.

### ***Good Presentation with User-Friendly Interface and Printable Pages***

Your package should:

- Present its statistics in visual, easy-to-garner ways
- Allow you to navigate quickly and easily between different reports
- Offer printable versions of all reports
- Allow you to export the data to other software (e.g. a spreadsheet)

### ***Page Performance***

Most packages will give you a report of the comparative performances of individual pages. On any given site, certain pages will attract more and longer visits than others. Use this feature to identify under-performing pages and consider how you can improve them.

### ***Visitor Paths Through Site***

Most packages will give you some idea of the paths that visitors take through your site, allowing you to answer questions such as:

- What is the most popular entry page on the site?
- What is the most popular exit page?
- Where do visitors of page X go next?

(The latter question will be answered in the following way: 60% go to page Y; 30% go to page Z; 10% exit the site.)

This information is useful in determining whether your visitors are engaging with the site in the way that you intended, and whether you need to change the design, or navigation.

Some of the more expensive packages will actually tag visitors (using a cookie), allowing detailed per-visitor reports of their “click-streams” through the site. However, you should avoid any package that uses Session IDs for this purpose, for reasons explained in Section 2.27, *Session IDs*.

Click-stream analysis may be useful for large e-commerce sites, but smaller sites should have enough information in that provided by standard packages.

## 6.2 Monitoring Keyphrase Performance

As I mentioned, getting into the top ten of the search results takes hard work, but so does staying there. By monitoring the performance of your keywords, you can create such a strategy that can lead to sustained performance in Google and other search engines.

### 6.2.1 Listings in Flux

There are various reasons why your keyphrase may change in position in a search engine's results from one week to the next.

#### 1. The Competition is Always Increasing.

Search engine optimization is a relatively new “science” and more and more people (including you) are learning how to get good at it.

Since search engine bots crawl the web regularly – Google's bots crawl on a daily basis – listings will be refreshed often, and competitors can quickly steal a march on one another.

#### 2. Search Engines Regularly Update their Search Technology

Google makes a *major* update to its search algorithms once every two or three months, and continually makes minor tweaks. Other search engines regularly tweak their algorithms too, if only to keep up with Google.

You've got to stay in touch with new developments in the search engine industry, and make the necessary adjustments to your own site, to ensure the competition doesn't overtake you.

#### 3. Your Content and Backlinks have Changed

If you're serious about your website, you'll regularly edit and add to its content. Changes you make will (sometimes inadvertently) affect the optimization of your site. Mostly, they will improve your performance but, from time to time, you may make a change that sees a page dropping in the listings.

Even if you don't change your site content, those that link to you may change theirs. Backlinks can disappear, or lose their value. Don't rely on the backlinks you already have; always aim for more.

### 6.2.2 Spreadsheet Solutions

A popular way to track the performance of your targeted keyphrases is to create a spreadsheet, listing each of the keyphrases, and their positions in Google's results on certain key dates (for example, the first Monday of every week).

The spreadsheet approach can also be used to monitor your positions with other search engines; or to monitor the positions of your competitors for the same keyphrases.

	A	B	C	D	E	F
1						
2						
3		<b>Comprend.ie</b>			<b>Aspire.ie</b>	
4	<b>Query</b>	<b>Google</b>	<b>Yahoo</b>		<b>Google</b>	<b>Yahoo</b>
5	Marketing Consultants Ireland	>100th	>100th		1st	1st
6	Strategy Consultants Ireland	3rd	>100th		20th	12th
7	Sales Management Consultants Ireland	>100th	>100th		>100th	>100th
8	Business Consultants Ireland	>100th	>100th		29th	>100th
9	Brand Consultants Ireland	7th	>100th		3rd	3rd
10	Strategic Marketing Consultants Ireland	1st	>100th		>100th	2nd

*Example of Spreadsheet used to Track Competitors*

### 6.2.3 Monitoring Keyphrases in Your Site Statistics

As I suggested, you should choose a statistics package that lets you see the keyphrases visitors searched before coming to your site, and what engines they searched on.

When analyzing your statistics, consider the following questions:

#### **Are your high-performing keyphrases leading directly to traffic?**

You're close to having a successful findability strategy when your site shows up in the first page of search results for your targeted keyphrase. But true findability success is when your statistics indicate a stream of search engine visitor referrals for that keyphrase.

If your targeted keyphrase gets good positions and brings you traffic, great! Just keep monitoring this keyphrase and consider making adjustments if the volume of traffic slips.

If you are getting unsatisfactory levels of traffic, despite high positions in the search results, there are two possible explanations.

1. **The search query isn't popular.** Did you research the popularity of the query, as explained in Section 2.4, *Keyword Research*? Read this section again and try to identify and target more popular queries.
2. **Is the description of your site not enticing the users to click?** Identify the sentence(s) on your site that the search engine selects to create the description. Change this to create a more enticing sentence. (Note: the description may come from the meta description tag.)

#### **Are "poor" performing keyphrases nevertheless bringing traffic?**

The majority of searchers don't look beyond the first page of results. But there are still a significant minority of occasions when users are unsatisfied with the top 10 or 20 results, and will dig a little deeper.

Work patiently at get even higher listings for all keyphrases that are bringing you traffic; they *must* be popular search terms. Work hard, too, at providing content that is a better match to the query than all of the other sites that show up.

#### **Are you getting traffic from keyphrases you didn't intentionally target?**

As your site builds its number of backlinks and grows in content, you are sure to notice traffic coming from search queries that you hadn't anticipated.

All traffic should be welcomed, of course, and this is one of the great side-benefits of putting time and effort into optimization.

Pay attention to such keyphrases if they appear more than once. They may represent SEO niches that haven't yet been filled – virgin territory. If your site is coming up highly without any deliberate effort on your part, you can perform even better if you *do* optimize for these phrases.

Work at creating new content and getting relevant backlinks.

#### **6.2.4 Monitoring Competitor Backlinks**

If one of your competitors suddenly jumps to a higher position for a certain keyphrase, and you have stayed in the same position, chances are they have made some changes to their site, or added to their backlinks.

For each of your main competitors, keep a list of their most important backlinks (the top 10 or so to appear in a Google search for link: [www.yourcompetitor.com](http://www.yourcompetitor.com) ). Do this using a spreadsheet, along the lines I've shown in Section 6.2.2.

If a competitor suddenly jumps in the rankings, look to see whether they have got any new backlinks. Can you get backlinks from the same sites?

## 6.3 Growing Your Site

### 6.3.1 Adding Content

One of the best ways to improve your findability is to add new pages of content to your site. It's a no-brainer: the more words you have in a site, the more likely it is that combinations of those words will show up in search engines.

Sites that have higher numbers of informative, cross-linked pages perform far better in search results than sites with fewer pages. With every new page comes the opportunity of optimizing for an additional keyphrase.

Use the various keyword suggestion tools (see Section 2.4.2, *Keyword Suggestion Tools*) to come up with new keyphrases to target on your new site pages. These tools reflect search trends, so watch out for queries that have become popular since you conducted your initial research.

Add content in an organic fashion – re-purposing a report you have written or an article you have had printed in the local newspaper, for example, by finding a place for it on your own site.

Remember to try and make links within your site among pages that share themes or topics.

If your site has a section that promises updates – such as a “company news” section – be sure to add fresh content regularly. Nothing looks worse than a news/updates/press releases section on a site that hasn't been updated for months, or even years. (This is all too common on corporate websites.)

An attractive, findable, well-maintained site will enhance your brand. An amateur-looking, outdated site that doesn't get listed in search engine results will *harm* your brand.

### 6.3.2 Experimenting with Content and Keywords

As well as adding new pages, you should also make changes to existing pages – particularly pages that aren't getting high positions in search engines for their target keyphrases.

Before making changes, make sure the page has good PageRank and decent backlinks (check against the competitors' pages for the target keyphrase to get an idea of what you need to achieve here). If you are satisfied that the page should be performing better, start tinkering.

Play around with new combinations of keywords. Move the keyphrase into a new position in the text. Add or remove an occurrence of the keyphrase, depending on the occurrence density. Remove words that might be considered variants of the keywords. Change the Page Title.

Don't be afraid of making mistakes. Search engine optimization is like any other skill – you *learn* from your mistakes. Even if you make a change, you will quickly see the effects in the listings, in Google at least.

Keep backups before making any changes, and revert to older page versions if the changes aren't improving your position.

### **6.3.3. Content Management Systems**

A Content Management System (CMS) is software that allows users to publish content to a website. With a good CMS, users do not need to know HTML or other code; they simply enter the text and click a button to publish it to the site. A good CMS will itself be accessible via the web, so that users (who have access) can update their websites from anywhere, at any time.

If you are serious about regularly updating your site content, a content management system is a great solution, particularly if you are not a web designer.

Avoid proprietary CMS systems – those that are created or tailored (usually by IT consultants) for your site. These usually cost a lot of money, and they often produce pages that are dynamic pages, or otherwise Google-unfriendly.

It is better to use an off-the-shelf type of CMS.

#### ***Tech Tip:***

If you are confident installing Perl/CGI scripts on your webserver, I highly recommend the popular Movable Type (MT) system. MT is free for use on personal sites, while a commercial license costs \$150. MT is best known as a weblog system, or “personal publishing tool,” but it is essentially a content management system.

I like Movable Type because:

- It is user-friendly
- It comes with very Google-friendly templates
- It has a well-linked archive structure
- Pages can be added quickly and easily.

You may need to do a little research (use Google!) to learn how to configure it to power a non-weblog site.

### **6.3.4 Temporarily Popular Search Queries**

If you update your site regularly, you may be able to gain high positions for fad searches – such as the name of a person who has unexpectedly made news headlines, or a new buzzword that are spreading throughout a cultural community.

These searches can bring short-lasting but high-volume spikes to your site.

The Google Zeitgeist page – <http://www.google.com/press/zeitgeist.html> – lists recent search trends, which often contain details of such trends.

However, before targeting such keyphrases, you should ask yourself these questions:

Will the content you create to target these keyphrases genuinely be relevant to your site? Will these keyphrases help you get the kind of visitors you really want? Can you convert these visitors into repeat visitors?

Temporarily popular queries certainly offer you the opportunity to get high positions and high traffic quickly – particularly if you have a knack for predicting search trends. On the other hand, you may put a lot of effort into a pursuit that doesn't meet your site's aims and objectives.

## 6.4 Search Engine Marketing and Paid-For Listings

Throughout this book I have described ways to make your site findable and increase your traffic – without spending money.

Yet, some commercial sites have aggressive traffic targets, and have budgets to spend on achieving this traffic. Such sites may engage in a broader search engine marketing (SEM) strategy that includes, but is not limited to, optimization.

SEM consultants aim to get sites found through regular *and* paid-for methods. Google accepts payments for its AdWords listings, which appear separate to, and distinct from, its regular search results. (Note, however, that searchers are more likely to click on natural results than on sponsored links.)

Other search engines, notably Yahoo, offer sponsored links and other ways to pay to get included in their index of results.

Two common pricing models used in SEM are:

- i) keyphrase auctions
- ii) pay-per-click (PPC).

In keyphrase auctions, the search engine gives the highest position in its sponsored links section, for a given keyphrase, to the highest bidder. (Google uses this system to determine the price and position of its AdWords results.)

PPC means that you pay for each click that is received via the sponsored link.

SEM consultants advise companies on the best way to spend their budgets; what keyphrases to bid for; etc. They may help you generate some extra traffic if you are willing to spend enough money.

Of course, you don't need to engage an SEM consultant in order to create an SEM strategy – you can deal directly with the search engines. If you are determined to spend money on SEM, I recommend the Google AdWords program.

## 7. Search Engine Spam

Most internet users think of spam as unwanted, unsolicited emails. In the search engine world however, another type of pest – called *index spam* or *search engine spam* – is a fast-growing problem.

### 7.1 Defining Search Engine Spam

There is no agreement on what constitutes search engine spam.

Some academics define spam as *any* action taken by a website publisher that influences the ranking of search engine results. I find this definition unfairly pejorative, as it fails to distinguish between honest-spirited and dishonest-spirited attempts to win favorable search engine rankings. Moreover, it is simply naive to expect publishers to resist attempts to influence search engine results.

The following, in my view, is a better definition:

Search engine spam is any document that, in an effort to gain favorable ranking, either

- a) compromises the quality of its information, OR
- b) attempts to deceive a search engine bot, OR
- c) attempts to manipulate a search engine algorithm, OR
- d) attempts to deceive the end user.

I believe that the definition offered above articulates the principles adopted by Google, Yahoo and Microsoft vis-à-vis spam.

I want to differentiate between *manipulating a search engine algorithm* (which, in my view, involves an element of deception) and *capitalizing on knowledge of how a search engine works*. The latter describes what a good SEO consultant does, and should not involve compromising the design or quality of the website.

The area between manipulating an algorithm and capitalizing on knowledge is, I admit, sometimes a gray one. However, spammers and SEO consultants each tend to have distinguishing characteristics, which I outline below:

#### **Spammers – a.k.a. “Blackhats”:**

- Look for quick jackpots
- Try to find and exploit new weaknesses/loopholes in the SE algorithms
- Disregard website visitors’ needs
- Typically produce shoddy websites with poor content
- Run continual risk of penalization by SEs

#### **SEO Consultants – a.k.a. “Whitehats”:**

- Work patiently to achieve steady improvement in rankings
- Continually improve their optimization techniques to reflect evolving SE policies
- Put website visitors first
- Create quality, information-rich websites
- Will not get penalized by SEs

*SEO consultants try to beat the system, while spammers try to cheat the system.*

## **7.2 To Spam or Not to Spam**

While email spam is deemed in many regions to be an illegal activity (leading to prosecutions in some US states), search engine spam is not illegal – and is unlikely ever to be.

Search engine spammers don't force unsolicited messages into search engine results pages. All they do is publish websites. Spammers don't *ask* for search engine bots to crawl their sites. The search engines crawl sites without any explicit permission to do so.

The onus is on the search engine company to filter out attempts to deceive or exploit its algorithms. Otherwise, the quality of results will deteriorate, and users will abandon that search engine in droves.

Nevertheless, it is in the interest of web publishers to avoid using techniques that might be considered spam – because the risks of being caught and penalized by search engines are too severe. **Sites found (whether through manual or automatic detection) to be using spam techniques can lose their high positions, or even be banned from search engine results altogether.**

This can not only harm the performance of your website, it can harm your *brand*.

## **7.3 Been Caught Spamming?**

It is difficult – but not impossible – to *inadvertently* use spam techniques.

If your site has been found guilty of spam, you should enquire with the relevant search engine company to find out why your site has been de-listed, then take the necessary steps to undo the offending techniques.

Individual sites are rarely banned indefinitely; those that clean up their acts are usually re-included in the index.

However, large scale spamming operations that are based on networks of interlinked sites can lead to entire “communities” of websites or servers being banned long-term.

## **7.4 Common Types of Search Engine Spam**

The following is a list of techniques that search engines considered deceptive:

### ***Cloaking***

Cloaking is the practice of serving two different sets of content – one version to the search engine bot that crawls the site, and another version to the site visitor.

Cloaking is a relatively sophisticated technique that usually involves checking the IP addresses of all visitors. Search engines consider cloaking to be a major offence, since it is a deliberate and technically advanced attempt at deception.

### ***Refresh Pages or Doorway Pages***

Doorways are web pages that have been optimized for a keyphrase, but which immediately re-direct to another page (usually not relevant to that keyphrase). Such pages often contain large volumes of meaningless, auto-generated text.

In fact, you are better to completely avoid using the <meta refresh...> tag. While this is a legitimate HTML tag, search engines are not likely to rank highly any pages that use it, since the threat of deception is too great.

### ***Meta Keyword Spam***

Yahoo checks to make sure that the words used in the <meta keywords...> tags also appear in the body of the document. Where meta keywords do not also occur within the web page, Yahoo considers this spam. Google ignores meta keywords altogether, except in cases of rare or unusual content.

### ***Excessive Keyphrase Repetition***

Search engine optimization is so called because it requires an *optimal* blend of keyphrases within the content of a web page. Web pages that overuse a certain word or phrase may disappear in the search results for that query.

### ***Bait and Switch***

In this case, the spammer creates a search-engine friendly page and, when it gets a good ranking, changes it for another page, with different content. Since the major search engines update their indexes regularly (e.g. Google updates several times a week), this strategy is doomed to failure.

### ***Page-Jacking, Content Duplication, Plagiarism***

Page-jacking is a blatant form of plagiarism. Basically, the spammer copies word for word a page that is already performing well in the search engine results for a given topic. This technique is unlikely to succeed, as search engines penalize “duplicate content” – where the content of one page has clearly been copied from another page that was online before it.

### ***Hidden Text***

One of the oldest search engine spam tricks is to use “invisible” text that is the same color as the web page background. Search engine algorithms can easily spot this, as well as more sophisticated efforts at concealment (such as those involving CSS).

### ***Keywords in Comments***

A variation on the above, stuffing keywords in HTML comment tags is a lame trick that search engines caught on to long ago.

### ***Hidden Links***

As with text, any attempt to conceal links on a web page is seen as blatant deception.

### ***Cross Linking Networks and Link Farms***

Search engines are particularly wary of “mutual admiration societies” – networks of sites that link excessively among each other. These are also called “bad

neighborhoods.” A reciprocal link to a page in such a neighborhood could have adverse effects on your optimization efforts.

### ***Misleading Anchor Text***

The text used in a link should correspond to the content of the page in which it appears *and* the page it links to. Google is growing tougher on deceptive anchor text.

### ***Keyword Stuffing in ALT Text***

Stuffing the ALT text of images – particularly 1x1 transparent pixels – with lists of keywords is a technique that was once highly prevalent, until Google cottoned on and adapted its algorithm to detect this practice.

### ***Domain Flooding***

While some sites understandably use more than one domain name, a common spam technique is to register a large number of domains, which all point or redirect to the same website. Search engines easily detect this technique, since they continually update their maps of the web.

### ***Over-Submitting***

Many search engines and directories don't like it when sites make multiple submissions via the “Add URL” button. On some, such as DMOZ, this action is counter-productive as it may lead to a delay in, or refusal to, include your URL at all.

### ***Comment/Guestbook Spam***

Sites such as weblogs and guestbooks offer various ways for visitors to post their URLs, using forms. Spammers try to exploit these features to create links to their sites. Often, they use automated programs to find as many such pages as possible across the web, and post URLs to them. Google ignores guestbook links, and is beginning to take a tougher stance on comment spam.

## 7.4 The War Against Search Engine Spam

Almost all of the dubious practices listed in the previous section can easily be detected – Google’s algorithm is particularly adept at spotting spam, and Yahoo’s is improving all the time.

A technique used for filtering spam from email – called the Bayesian method – is also gaining popularity among search engines. This involves “teaching” an AI program to recognize spam by showing it existing examples. The program sees patterns that are not usually discernable by humans. The Bayesian method is used as a safeguard to catch spam attempts that are not caught by the regular algorithms.

The most effective – but most costly and time-consuming – way to counter spam is through human editorial intervention. Google tries to avoid this solution as much as possible; Yahoo relies a little more than its rival on editorial input. Still, even Google will undertake practices such as reviewing sites with high positions for the most popular search queries.

Even if a site manages to pass through all of these nets, spam can be reported by users who are frustrated with the poor quality of results for a given search query, or by those who believe their competitors use spam techniques.

If you wish to report a search spammer to Google, fill out the form at:  
**<http://www.google.com/contact/spamreport.html>**

To report results spam on Yahoo, go to:  
**[http://add.yahoo.com/fast/help/us/ysearch/cgi\\_reportsearchspam](http://add.yahoo.com/fast/help/us/ysearch/cgi_reportsearchspam)**

## A Final Word of Thanks

Thanks for buying this e-book – and for reading it to the end! I hope you have found it useful.

If you have enjoyed *Website Findability* or – better yet – if it has helped increase traffic to your website, I would love to get your feedback. You can reach me by email: [mike@mediajunk.com](mailto:mike@mediajunk.com).

Of course, you may differ with some of the guidelines I offer here, or you may receive different advice elsewhere. In my experience, no two SEO consultants ever completely agree. But what I have revealed in this book are the techniques that work for me.

If you feel I have got something wrong, then by all means let me know. Similarly, if you feel I have left something out, or have any other suggestions as to how I may improve this book, drop me a line.

Maybe I'll include your suggestion in the upcoming version of the book. **Remember, you've already paid for the next five updates.**

Thanks again – and good luck with your findability strategy!