Part of Omega Digital Media. Optimising since 1997

Page 1 of 14

How Search Engine Optimisation Works

Dr Brian Clifton PhD
www.search-engine-professionals.co.uk

Updated August 2005

(original whitepaper written 2002)

<u>Table of Contents</u>

Introduction

Just ten years ago, when the web first emerged as a commercial entity, there was little need for the immensely complicated search algorithms that power Search Engines today. But the phenomenal growth of web content brought a pressing need to provide consumers (visitors) with an efficient way of navigating content to find relevant information quickly and easily. Without the immense developments in search technology, the web would have drowned under its sea of over 10 billion pages.

As you would expect, adjacent to the improvements in Search Engine technology, Search Engine Optimisation is the art and science of helping web sites get found on the major Search. The process is often complicated, timely and manually intensive. It can even be a waste of time if your pages are not structured in a 'Search Engine friendly' manner. However get it right, and your business can rocket!

This white paper explains how Search Engines work and how Search Engine Optimisation (SEO) can help you achieve high Search Engine visibility.

Independent research shows that over 90% of people searching for a product or service do so using one of the major Search Engines or directories.

What is a Search Engine?

For our purposes there are two primary kinds of search services on the Internet:

- **Organic Search Engines** use automated programs called "spiders" or "robots" to index web sites into a large database – effectively an Internet roadmap. You submit your page to an organic Search Engine and the "spider" will index your entire site. Examples include Google, Yahoo, MSN, Teoma (AskJeeves) and others.

- **Search Directories** are human reviewed category listings that rely on submissions from site owners. There is no guarantee your site will be included and which category and what text is used to describe your site is the responsibility of the reviewer. Examples are Yahoo Directory, Open Directory (Google Directory), Looksmart and others.

There is also what is known as Pay-Per-Click (PPC) Search Engines. However, these are not considered Search Engines themselves, but tools for advertisers to bid on keyword placement. Bids are then resold as "featured listings" to the major Search Engines. Examples include Overture, Google Adwords and FindWhat/Espotting. The PPC model is discussed in our white paper Understanding Pay-Per-Click on Overture.

Search Engine Professionals

Search Engine Relationship Chart (as of Jan 2006*)

Because of their sophistication and vast database size, there are only a few major Search Engines and almost all have relationships with another. An interactive version of this chart is available at: www.omegadm.co.uk/relationship-chart.htm
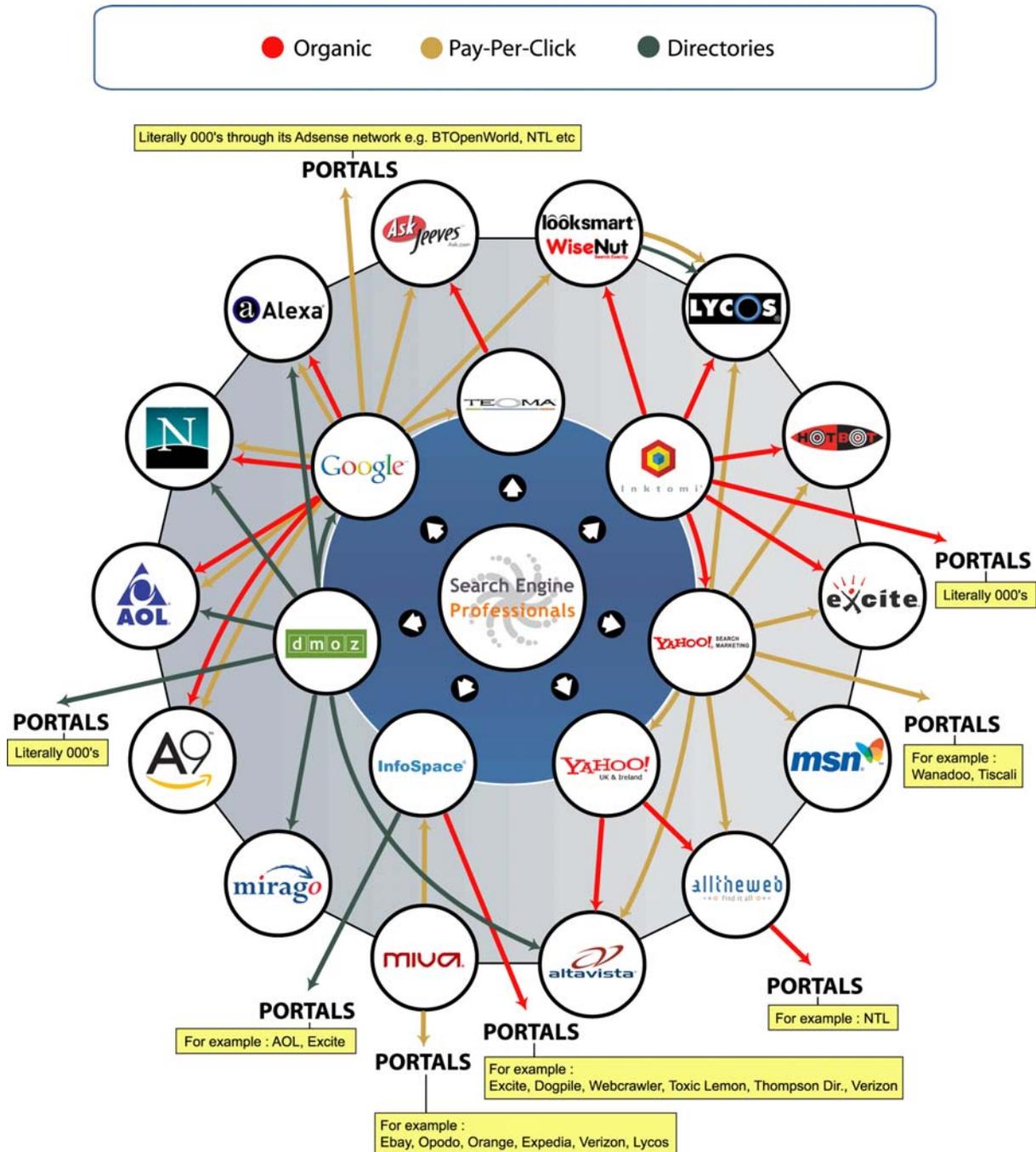
*This chart is from a UK perspective i.e. in the US, some of the PPC relationships are different. For example, Excite.com, Lycos.com and HotBot.com receive PPC results from Google Adwords. For further information on PPC, see Understanding Pay-Per-Click on Overture.

Why Not Submit to Thousands of Search Engines?

A common myth on the Web is that submitting to thousands of "engines" will gain you all kinds of new traffic. In reality, there are not even thousands of "engines" to submit to. The bulk of those submissions go to what are called FFA (Free for All) sites that are not Search Engines at all. Instead, they are simply a list of links to the last fifty or so URLs that were submitted to the service. Nobody ever uses these lists to find anything because they are just a random list of web sites, generally with no organisation to the list.

FFA sites have two purposes:

The primary purpose of these sites is to gather e-mail addresses from web marketers so they can be sent offers for other products and services. By submitting, you are unknowingly giving your consent for them to e-mail you. Once on these 'mailing lists', it is virtually impossible to be removed.

In addition, FFA sites are used in an attempt to artificially inflate the Link Popularity of a site (the number of other web sites linking to a given web site), as a high Link Popularity gives you a better Search Engine ranking – particularly with Google.

However, artificial links from un-related web sites is a Search Engine spam technique. Similar in principal to email spam, Search Engine spam is unwanted and annoying to the software developers of the major Search Engines. When your site gets caught using such techniques, it will be banned from the Search Engines – and they almost always get caught over time. Once banned, it is almost impossible to get un-banned. The Internet is just too big for anyone to pay attention to a known spammer.

Page Ranking

As you will have realised, submitting alone is not enough. Where your site ranks on is what makes the difference. For example, when someone queries a Search Engine for a keyword related to your site, does your page appear in the top 10 matches? If you're not within the first two or three pages of results i.e. Top 30, you lose no matter how many engines you have submitted to. The key to improving your web sites' position in the different Search Engines, is understanding the basic criteria by which organic Search Engines index and then retrieve documents...

How Page Ranking Works

An organic Search Engine spider is a piece of software that acts like an electronic librarian who cuts out the contents pages of each book in every library in the world, sorts them into a extremely large master index, and then builds an electronic bibliography that stores this information as a database index. Some software spiders can index over a million documents a day. The technology involved is actually enormous. For example, Google has over 30 data centres world wide, with 2,000+ computers per centre! However, the basic principal is that the index is built from the actual _content_ of each site.

So, organic Search Engines contain large databases (indexes) that web users query to find Web sites. When you visit Google, AOL Search etc. and type in a keyword, you're simply performing a database query. To determine which documents or web sites to return for a particular keyword search, each Search Engine must have its own method of ranking documents within its database. This is done using a ranking (algorithm) method.

You can ethically influence the ranking of your web site by utilising "optimisation" techniques. These work _with_ the ranking algorithms to ensure your site correctly ranks high if a searcher types in keywords relevant to your web site.

I have highlighted 'work with ranking algorithms' and 'correctly ranks… for relevant' searches. This is important, as trying to trick the Search Engines into ranking your site highly will get you banned.

Where your web site appears in the returned matches depends on **5** key factors:

- Which Keywords the user entered
- Your web page Keyword Density for these keywords
- Your web page Keyword Prominence (page location) for these keywords
- Your Link Popularity – the number of other sites that link to you
- Your link and keyword Relevancy

Choose the Right Keywords

Choose the right keywords, and you'll see your traffic rocket. Choose the wrong keywords and all your efforts will have been wasted. Therefore, think long and hard on what keywords people are likely to use to find you. Make lists of keywords and then combine them into phrases.

For example, you rarely want to target a single keyword since with the billions of words indexed on the web now, generally one word simply won't work. People learn quickly that if they type in "estate agents" or "properties" that they get listings for properties from all over the world. Now it would be nice if you were positioned well on those extremely broad keywords, however, a better use of your time is to PAIR the generic keyword with something more specific. You might get lucky and rank well on just "estate agents" since its in your page too, but if not, at least you'll rank well on "Islington estate agents", which will be far less difficult to achieve a top 10 listing for. A more targeted, specific keyword will also bring you much better qualified prospects.

Target each page for 3-4 well focused key phrases. Avoid general terms. For example, a list of keywords for a company selling "cheap cars" may look like this:

- ❑ cheap
- ❑ cars
- ❑ cheap cars
- ❑ quote
- ❑ discounted
- ❑ used cars

However, the word "cheap" alone does nothing to target their audience. It has to be paired with other words like "Ford Focus" or "Sussex used car dealership" to have any meaning. It's the same with "quote" and "discounted" - too broad. Pair them with one or two other words. Statistically, most people now search for three to four word phrases to avoid getting back too many unrelated matches. Keep this in mind when you design your pages.

Don't forgit mispelled keywhords!
Vodafone produces 8,823 key word searches on Overture (UK) whilst Vodaphone delivers 7,386.

Keyword Density

For a particular keyword/phrase it is relatively straightforward to calculate your corresponding keyword density (KWD). Simply count the number of occurrences of your keyword/phrase that occur and divide by the total number of words (ignore the html source tags):

$$KWD = (\text{No. of occurrences} / \text{Total number of words}) \times 100$$

For example, on this page, the KWD for the phrase 'keyword' is:

$$KWD = (13 / 341) \ x100$$
$$KWD = 3.8\%$$

This initial calculation is obviously very low and would not rank highly on any major Search Engine. However the actual calculation the Search Engine performs is slightly more complicated because each Search Engine has its own 'skip words'. That is, common words (e.g. and, it, to, the etc.) that are not counted in the calculation. Also partial matches may or may not be included (I included 'keywords' in the above example) and sometimes misspellings are taken into account (Google does this). A better estimate for the above example is ~8% which is pretty good. SEO companies usually aim for a KWD of 10% plus.

Keyword Prominence

Keyword prominence, as the name suggests, is the location of the keywords within a page. This is much more complicated to estimate than KWD and again each Search Engine applies its own set of rules. As a guide, keywords placed near the top of the page (including the html source tags) carry more weight than those towards the bottom of the page.

Another instance of the importance of keyword prominence, is keywords *within* your page link text. For example, consider the following two snippets from two different web pages:

Example 1:

      Click here to see our price list.

The link is on the word 'here'.

Example 2:

      Follow this link for price details on our range of widget spare parts.

The link is on the words 'widget spare parts'

The second example is much more Search Engine visible for anyone searching the web for 'widget spare parts' because it has a link with those keywords in it. As an aside, it is also next to the keywords 'range' and 'price' so the sentence as a whole is well constructed for searches such as 'widget part prices', 'widget part range'.

Link Popularity

Link Popularity is a measure of the number of web pages that link to another web page. This includes all directory listings, affiliate linkings and any other web site that mentions you by including a link to your web site on their web site (known as referrals). Search Engines like Link Popularity as a ranking indicator because in essence, a link from page A to page B in considered as a vote, by page A, for page B. Web sites 'voting' for another web site is an indicator of relevance.

Sergy Brin and Larry Page invented* the concept of Link Popularity (originally called BackRub), while studying for their PhD's at Stanford University. It is now the foundation of their Search Engine – Google and is recognised as an important factor by other engines in determining ranking. Read the original paper.

* It should be noted that around the same time that BackRub was being developed in 1998, Professor Jon Kleinberg of Cornell University was also developing algorithms to analyse link structures on the web. Often both reference each other as the founding fathers of this technique.

Google uses Link Popularity as an indication of visitor popularity, and will then alter your ranking because of this factor. Obviously the higher your Link Popularity the better. However it is not just the number of links pointing to your web site that matter - the "quality" (or relevancy) of those links/sites play an important factor here.

Link Popularity was introduced to the Search Engine world in 1998 when Google was launched. The proof that Link Popularity actually works in delivering better quality results for the Search Engine visitor, comes from the phenomenal success of Google in such a short space of time. By 2002, Google had become the No.1 recognised brand on the Internet. Try the following example:
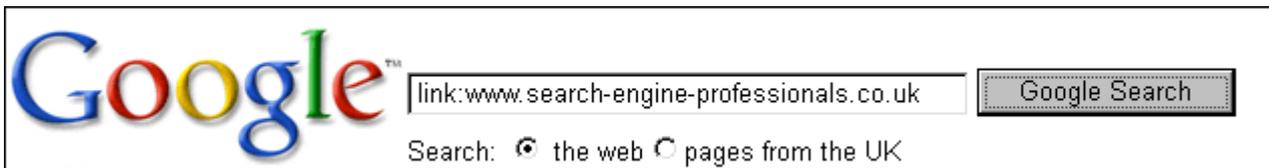
This will return the number of pages linking to the home page of this address - as seen by Google. It will include sub pages from the same web site, as well as pages from other web sites. The latter is much more important than the former.

Using the cool java tool from TouchGraph, the images below are 3D representations of link structures as viewed by Google. Each line represents a link to a web site.
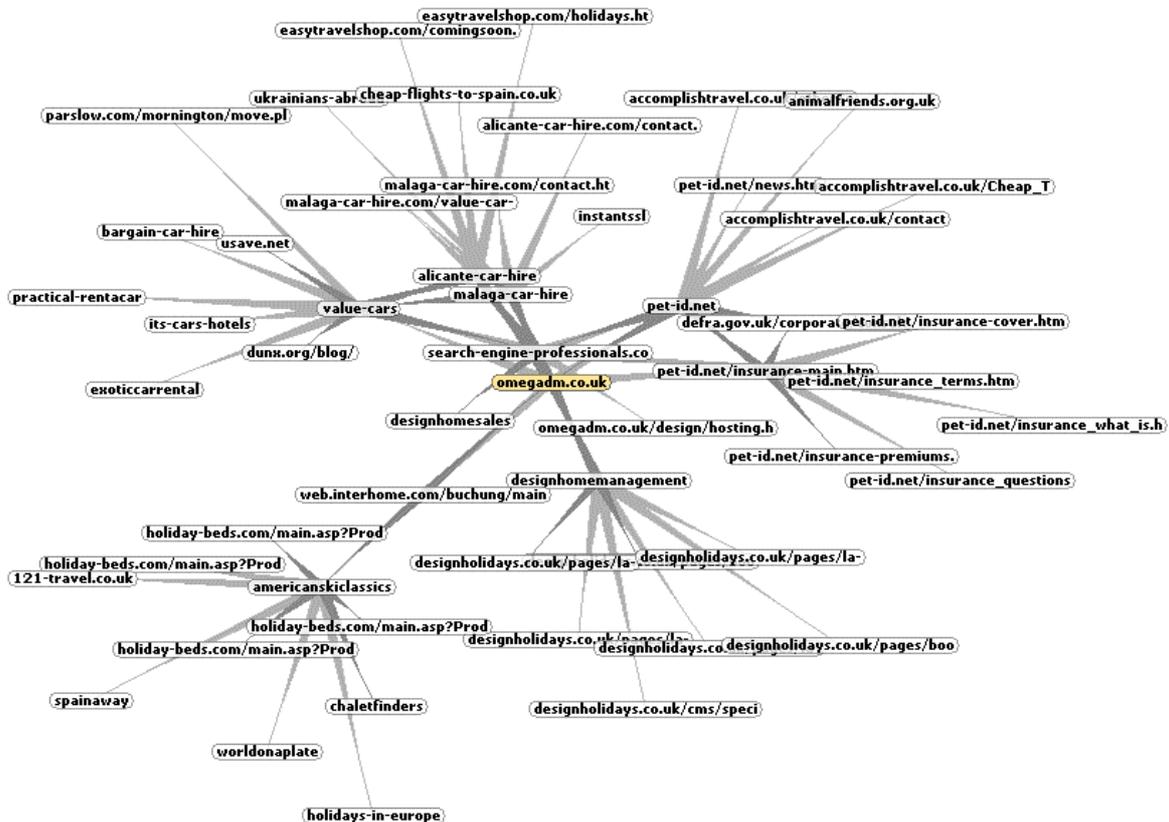
3D Link structure of different web sites
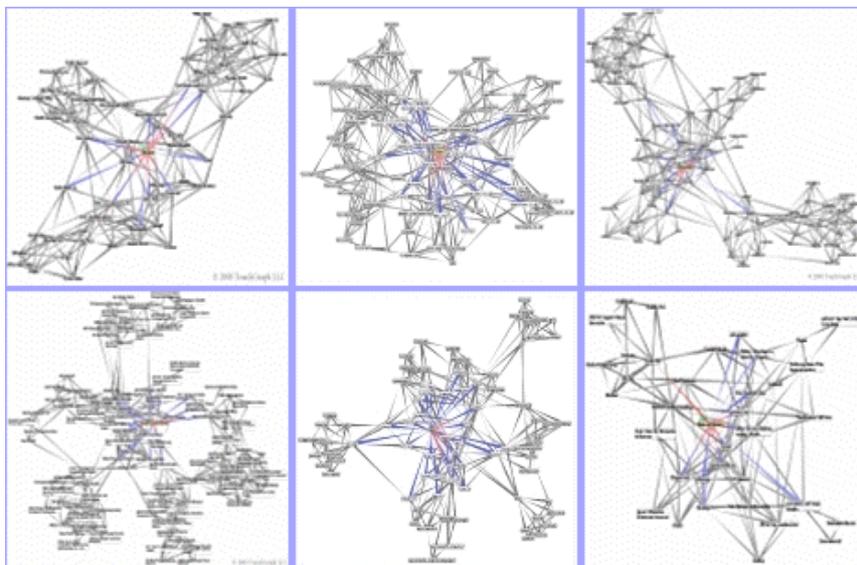
Notice how link associations have formed clusters or 'communities'. The Search Engine Teoma (the technology behind Ask Jeeves) is known to analyse these associations, both within and between these communities, in its own version of Link Popularity.

Achieve Top Search
Engine Results

Search Engine Professionals

Part of Omega Digital Media. Optimising since 1997

3D Link structure of www.omegadm.co.uk (centre)

Link & Keyword Relevancy

Relevancy is the most recent development to improve a Search Engines' ability to provide accurate results. It is also the most complex and "fuzziest" part of the page ranking process.

Relevancy is determined by looking at the language and words used within a web page (density, location), word context, discovering associations between them, checking where the site is referenced elsewhere (Link Popularity), spidering beyond the page going to other linked sites, then going back to the original site and checking the association.

As you can see, Relevancy is actually a combination of all the issues discussed so far and is a complex science – verging on artificial intelligence. The technology leaders in this field are Google, Yahoo and Teoma (owned by Ask). It is becoming more and more important as a ranking method for the Search Engines because it is next to impossible to hack/spam.

What is Google's Page Rank?

Google's PageRank (PR) relies on the unique democratic nature of the web by using its vast link structure as an indicator of an individual page's value i.e. Link Popularity. However it is not just the volume of votes or links a page receives that improves Page Rank, Google also analyses the page that 'casts' the vote. Votes cast by pages that are themselves 'important' (re: Search Engine visible) weigh more heavily and help to make other pages 'important'.

Part of Omega Digital Media. Optimising since 1997

Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search. Of course, important pages mean nothing to you if they don't match your query. So, Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search i.e. using the other methods described above.

The recent patent application (March 2005) that Google applied for has shed more light on Google's ranking system - you can read the 14,000+ word application at the US Patent and Trademark Office web site. Google's application is for a system, which Google calls *search engine 125*, designed to generate a "ranking score … that attempts to quantify the quality of documents" i.e. web sites.

One thing it means, is that Google's *search engine 125* takes into account a vast number of factors in determining how high or low to rank your web site in response to a Search Engine query relating to your web site. Not only the points mentioned so far, but also names, dates, visitor patterns, the number of years you registered your domain for, the legitimacy of other sites hosted on the same web server as yours, previous Search Engine rankings, how often your pages are accessed, how long visitors stay on each page, and signs of unusual activity and even inactivity.

For links, *search engine 125* will scrutinise not just their content, but also information on how they got to your site, how relevant they are to the other links and content on your site, how fresh or stale they seem to be, how compatible they are with the layout and design of your site and when and at what rate they were added to your pages.

A number of e-articles have been written about the Google patent application. However, be careful about what you read. A lot has been written saying: Urgent: do this, don't do that, change this, change that or risk disappearing from the Google index. Quite simply the patent application is not a blue print for how Google works. We do not even know whether Google is actually using *search engine 125*. We only know that they applied for a patent on it. However, based upon what is already known about Google's ranking system, it is reasonable to assume that most, if not all, of *search engine 125*'s elements are already in place.

Comparison of Organic versus Pay-Per-Click costs

As mentioned at the beginning of this document, you can buy placement (advertisements) on the major Search Engines by using what is called Pay-Per-Click portals. We mention it here for those of you who are familiar with the concept and are wondering if SEO is worth all the hassle described so far. Consider results obtained from analysing visitor behaviour for a site during 16-27 April 2005:

Visitor source v campaign cost

	Organic visitors	PPC visitors
Number of visitors	1007	1200
Cost	£107.00	£746.10
Cost per visitor	**£0.10**	**£0.62**

These figures speak for themselves – Yes SEO is worth it because long term it can reduce your marketing budget by 86%! In a very competitive market such as Travel, the savings would be even greater.

In fact we recommend PPC campaigns be used to compliment a Search Engine optimisation strategy. To find out why, read our white paper Understanding Pay-Per-Click on Overture.

A Final Note

All techniques described in this whitepaper are entirely <u>legitimate</u>. We are **not** intending to fool or 'cloak' Search Engines into displaying irrelevant content. Such cloaking methods are highly discouraged in the professional SEO world, as would any dubious marketing techniques in the non-web world. Our best practise approach is analogous to a newspaper editor structuring their headlines to catch a reader's attention and engage them, without compromising the quality of the news article - otherwise the reader would never come back!

Search Engines themselves invest huge resources looking for evidence of what could be called "jury tampering" i.e. attempts by web site owners to use deceit to artificially inflate their ranking. This is commonly referred to as spamming - analogous to email spam. And in the Internet Age, spam is like horse thieving in the Wild West - it is not a misdemeanor, but a hanging offence!

In their recent patent application, Google state quite bluntly their view on Search Engine spamming:

> *This is our game. It may be possible to beat us at it -- for a brief while -- by cheating, but the odds against it are long.*

Achieve Top Search
Engine Results

Search Engine Professionals

Part of Omega Digital Media. Optimising since 1997

SEO Methodology Flow Chart

```
  ┌──────────────────────┐              ┌──────────────────────┐
  │  Competitor Analysis │              │   Keyword Research   │
  └──────────────────────┘              └──────────────────────┘
                    ↘                      ↙
                  ┌──────────────────────────┐
                  │  Keyword Identification   │
                  └──────────────────────────┘
                              ↓
                    ┌──────────────────┐
                    │   Audit Report   │
                    └──────────────────┘
                              ↓
         ┌──────────────────────────────────────────────┐
         │  Making your website Search Engine Friendly   │
         └──────────────────────────────────────────────┘
         ↙               ↓              ↓              ↘
┌──────────────────┐ ┌──────────────┐ ┌──────────────┐ ┌──────────────────┐
│Content Optimisation│ │Page Re-naming│ │Link Re-naming│ │ Tag Composition │
└──────────────────┘ └──────────────┘ └──────────────┘ └──────────────────┘
```

Content Optimisation

Page Re-naming

Link Re-naming

Tag Composition

Creation of Site Map

Re-submissions if required

Link Popularity Building

Manual Submissions to TOP Search Engines

Paid Inclusion

Manual submissions to Regional Search Engines

Evaluation

Competitor Monitoring

Ranking Monitoring

Visitor Reporting

Further Information

Other white papers in this series:
- Understanding Pay-Per-Click on Overture
- How Search Engine Optimisation Works – **this one**
- Why not Submit to Hundreds or Even Thousands of Search Engines?
- Web Analytics Data Sources

If you would like to discuss your specific business requirements with a respected professional, please complete the form below and fax back to:

Brian Clifton
Fax: 01444 456814

What is your website address? _____

What keywords apply to your website?

_____

What is your current monthly traffic?
- ☐ Less than 1,000 page views
- ☐ 1,001 – 10,000 page views
- ☐ 10,001 – 25,000 page views
- ☐ 25,001 – 100,000 page views
- ☐ 100,001 – 500,000 page views
- ☐ More than 500,000 page views
- ☐ We have no visitor statistics available
- ☐ Site not yet launched

What is your current timeframe?
- ☐ In the next 30 days
- ☐ In the next 60 days
- ☐ Unsure of timeframe

Indicate your business needs:
- ☐ Search Engine (organic) optimisation
- ☐ Bid optimisation (pay-per-click)
- ☐ Web site audit/consultancy
- ☐ Visitor reporting/analysis

Have you used a provider for any of these before? ☐ Yes   ☐ No
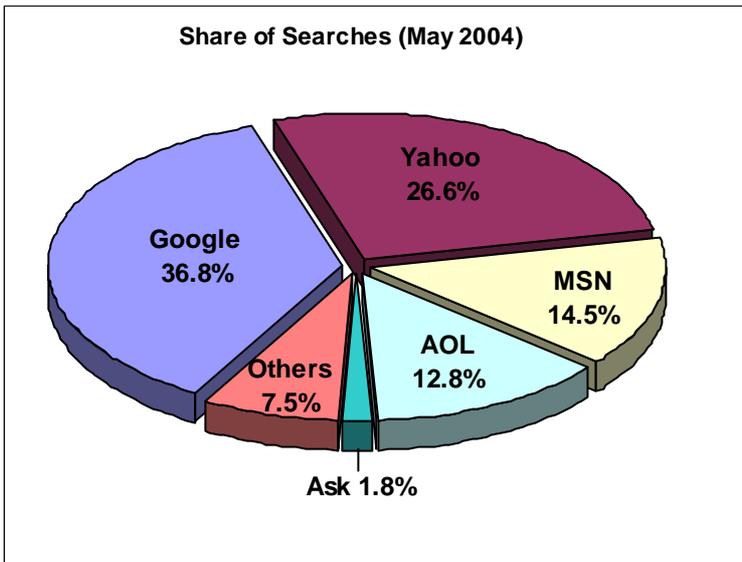
Your Name: _____

Telephone: _____

Email : _____

Appendix I: Search Engine Rankings

Which are the most popular Search Engines? The chart below is taken from comScore Media Metrix data for US 'search specific' traffic.

Share of Searches (May 2004)

Yahoo 26.6%
Google 36.8%
MSN 14.5%
AOL 12.8%
Others 7.5%
Ask 1.8%

The numbers are search-specific but not necessarily web-search specific. For example, a search performed at Yahoo Sports would count toward Yahoo's overall total. This is important as such "channel" driven searches were reported as making up 58% of Yahoo's total for January 2003. The same could be true of other non-pure search sites, such as MSN and AOL.

Also note that some search companies own more than one site. This means searches at different sites may be combined into one overall figure. For example, Yahoo data also includes AltaVista, AllTheWeb and Overture searches; AOL includes searches at Netscape and the Open Directory; Ask also includes Teoma searches; Others combines searches from InfoSpace, Lycos, HotBot, and Excite.

A significant finding from a recent study by Keynote (formerly Vividence) in Jan 2005, is that searchers are very fickle. They found that though 75% of users say they have a favourite Search Engine, 50% will go to another Search Engine as an alternative if they don't get satisfactory results. Also, as many as 20% of users were found to regularly use different Search Engines for different types of searches.

Search Providers
Not all search engines use their own technology. Some outsource to others (see our Search Engine relation chart), so a few Search Engines have a much greater reach than the above chart reflects.

Organic Search Providers (May 2004)

Google 49.6%
Yahoo 41.1%
Others 7.5%
Ask 1.8%
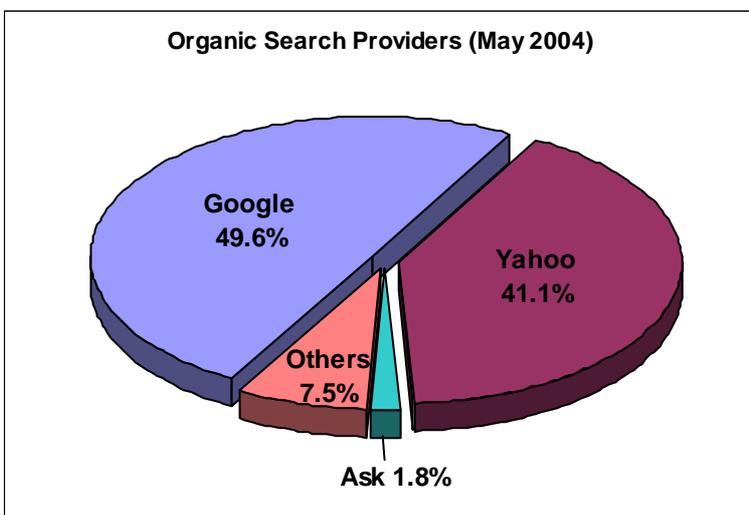
For example, Google supplies results to AOL, Yahoo supplies results to MSN via Inktomi (though this will change in early 2005)

The numbers are further complicated by the insertion of Pay-Per-Click results and that the provider of these varies from country to country. Note, PPC advertisements are excluded from this data.

Further information on PPC can be found in our whitepaper Understanding Pay-Per-Click on Overture.