Casting your Bread upon the Waters: Sundry Ways of Maximizing Search Engine Response for a Web Page

Matt Melchert
Information Technology Department
The Waikato Polytechnic
Hamilton, New Zealand

itmpm@twp.ac.nz

ABSTRACT

Anyone who puts a website up on the web naturally wants to get a high ranking on search engines. Being near the front of a search listing greatly increases the chances that a website will be visited.

This paper describes an attempt to evaluate five techniques for improving a site's search engine performance. Five different versions of a website—each with identical text but using a different technique for maximizing search engine response— were submitted to five search engines. The results so far have been rather strange.

First an overview of search engines is given, highlighting specific features of different search engine sites. This is followed by a description of the experiment and an evaluation of the results so far.

1. INTRODUCTION

Search engines have been around since the advent of the world wide web. They provide a means of finding information that would probably be otherwise unobtainable.

However, the sheer number of responses to a search often numbers in the thousands or even millions. Most viewers won't look past the first page.

In the face of such competition, how do you ensure that your page will appear in the first page of a search result? A little knowledge of how search engines rank pages according to relevance is useful.

Search engines typically look for:

- Unusual words that appear often in the text of the document
- Unusual words that appear in the document title
- Keywords listed in META tags, which allow keywords to be included that may not appear explicitly in the text. (Submission Tips, 2001).

However, a little knowledge can become a dangerous thing. Some webmasters have used a technique called "spamming"-a term which refers to massive duplication of information for the sake of increasing visibility-to increase the likelihood that their page will get a high listing. This could include:

- Duplicating keywords dozens or hundreds of times in META tags.
- Submitting multiple copies of a website in the hope that at least one will get a good listing.

To defeat such artificial techniques, search engines increasingly check for spamming and eliminate spammed websites from the database as a punishment (Search Engine Features, 2001).

Knowledge of how search engines work can enable a webmaster to increase the likelihood that their page will get a high ranking on search engines. Therefore, this experiment was devised to determine which search optimisation techniques work best with which search engines.

2. SEARCH ENGINES

There are a number of ways to find specific information on the Internet (Ackerman & Hartman, 2000). The ones which are relevant to our topic are:

- Directories or subject catalogs. These are hierarchies of information that are constructed by hand by a team of editors.
- Search engines. These are suites of programs that search the web for websites and assemble web pages in databases. These databases can then be searched for specific content. Note: this is done normally without human intervention.
- Meta-search tools. These pass a query on to a set of search engines and return the results for each.

A trend in the last year or so has been towards a new type of page called a *portal*. A portal is a kind of "gateway to the Internet", incorporating search engines, directories, and specific links to topics of interest which are often customisable by members. They also include advertising, which may be tuned to members' stated interests. Thus a site that was formerly specifically a search engine may now include any or all of the other features mentioned.

In order for a site to be found by a search engine, the search engine must be aware of the site's existence. This is generally done in two ways:

- An element of the search engine explores links in the web, recording each page it finds and associated keywords in a database. This program is variously called a crawler (AltaVista, 2001) or spider (Lycos, 2001).
- The webmaster specifically submits a site to the search engine.

The first method requires that there be a link to your site from some other site. Then, the crawler needs to find that link and follow it to your site. This is beyond the control of the webmaster. The second method requires someone to actually submit the site to the search engine. The site is then verified by the crawler and added to the database if valid.

3. THE EXPERIMENT

Given that search engines use a variety of techniques for ranking sites as mentioned above, it was decided to run an experiment to determine which techniques were the most effective.

A simple but non-trivial web page was duplicated several times. The topic chosen was a brief treatise on the humour of Spike Milligan. Each version has identical text, but differs in the techniques used to increase the ranking of the page in a search engine. Each version of the page includes one search engine technique to be tested. The aim is to see which techniques are most effective in getting a response from the search engine.

The techniques tested were as follows:

- No techniques used (I). The page is submitted in "bare" form. A paragraph containing the significant keywords will appear near the end of the document. This will test how important a search engine considers the whole text of the document when ranking.
- No techniques used (II). Same as above, but the paragraph with the significant keywords appears at the very top of the document. This will test how

important a search engine considers the opening text in ranking a document.

- Meaningful title. This will test how strongly each search engine rates the title in its ranking system. The keywords will appear ONLY in the title.
- META tags. A list of significant keywords is stored in a META tag in the head of the document. The keywords appear ONLY in the META tag. This will assess the importance attached to keywords in META tags when ranking a document.
- Spammed keywords in META tags. The significant keywords are repeated 100 times in the META tag. Unscrupulous webmasters sometimes do this in hopes of increasing the page's ranking, but some search engines disallow spammed keywords and will refuse to include the document.

These techniques are taken from Danny Sullivan's Search Engine Watch site (searchenginewatch.com, 2001).

The test search string to be used is: "yiddle +sapristi". These are nonsense words which were used frequently in The Goon Show, for which Milligan wrote the scripts. These being fairly unusual words, they can be moved around the document easily and they won't appear surreptitiously and repeatedly.

These five files were uploaded to the www.crosswinds.net hosting service, and stored in the same directory. To keep the project manageable, only five search engines were chosen. These were:

AltaVista, Excite, Google, HotBot, and Lycos. Directories were avoided, as the human element in selecting and ranking the pages would tend to cloud the result.

The five versions of the web page were all submitted to each of the search engines. These search engines were checked regularly after submission using AgentWebRankingSuite 2.0.3 (, 2001).

4. RESULTS

The results differed wildly between search engines. Search results started appearing within a week of submission from AltaVista. The version that originally appeared was the one that had the keywords in the title, and it was the first page listed. The other search engines never reported any of the pages at all. The other search engine sites were contacted to find out what may have happened. The only response I got was to wait for two months, as it might take that long for a submission to register.

Two months went by. None of the other search engines registered any of the pages. But then, a strange thing happened: the AltaVista listing at first disappeared for a week or so, but then started returning different versions of the page on different days! Danny Sullivan of the Search Engine Watch site was then contacted to ask for his ideas on this strange behaviour. He suggested that the pages may not have appeared on the other search engines because of anti-spam policies, i.e. pages with similar names and similar content at the same hosting site are disallowed.

It's a mystery why AltaVista would have returned different versions on different days. This is probably due to details of the ranking algorithm.

5. CONCLUSIONS AND FURTHER WORK

This has been a rather difficult study to undertake. A lot of things were hard to control and difficult to determine, such as why pages did not show up on four of the five search engines and why the AltaVista results kept changing.

There is some evidence that because of the astronomical number of pages on the web and the number of search engine submissions that are useless junk pages, sites discovered by a crawler are ranked more highly than pages that were submitted (Inktomipenalty, 2001). Further, there have been reports that Inktomi (a search engine program used by a number of search sites) penalizes sites that have been submitted for free as opposed to a paid submission (ibid.). These trends will make it increasingly difficult to run a test of this type.

To avoid the anti-spam pitfall when similar sites are submitted to a search engine, it was decided to upload five versions to five different hosting services.

Accordingly, the author attempted to find five other hosting services and upload the files. A discovery was made: free services on the internet are difficult to use, unreliable, and increasingly hard to find. Many hours were wasted signing up with various services only to find that they weren't free, weren't accessible, or didn't provide an upload service. At the time of writing this process is still underway.

Part of the reason this study proved so difficult and inconclusive is that it was done on the cheap. No software or services were purchased in doing this study. It is increasingly clear that to obtain meaningful results some money will need to be spent.

Further developments of this study could include:

- Completion of the dispersed hosts scheme.
- Comparing the ranking of a paid submission with one that was done for free.
- Purchasing several search engines and installing them on an intranet. The experiment could then be run in a more controlled manner.

Fortunately, the experiment has not been a total loss. Two principles were suggested by the results:

- Make sure the document title incorporates all meaningful keywords that describe your website. This is borne out by the author's and other people's experience with other websites. Although it wasn't touched on in the experiment, this will also work well with web pages that contain framesets and no text.
- Avoid spamming. Search engines increasingly disallow pages that are spammed.

And finally, the bad news:

- In order to have a high quality website that will be well trafficked, you will probably need to spend some money.

REFERENCES

Ackermann, Ernest, and Hartman, Karen. *Searching and Researching on the Internet and the World Wide Web*. **Franklin, Beedle & Associates, Inc, 2000.**

Add your site to Lycos. Accessed 14 May 2001. <http://home.lycos.com/addasite.html>.

Search engine submission tips. Accessed 14 May 2001.<http://www.searchenginewatch.com/webmasters/index.html>.

AgentWebRankingSuite freeware. Accessed 17 May 2001. <http://www.aadsoft.com/agentwebranking/intro.htm>.

How AltaVista works. Accessed 14 May 2001. <http://doc.altavista.com/adv_search/ast_haw_index.html>.

Inktomi's Free Add URL Penalty. Accessed 17 May 2001. <http://www.searchenginewatch.com/sereport/01/02-inktomi.html>.

Search Engine Features for Webmasters. Accessed 18 May 2001.