The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1066-2243.htm

Analysing Google rankings through search engine optimization data

Michael P. Evans
Information Systems, University of Reading, Reading, UK

Abstract

Purpose – The purpose of this paper is to identify the most popular techniques used to rank a web page highly in Google.

Design/methodology/approach – The paper presents the results of a study into 50 highly optimized web pages that were created as part of a Search Engine Optimization competition. The study focuses on the most popular techniques that were used to rank highest in this competition, and includes an analysis on the use of PageRank, number of pages, number of in-links, domain age and the use of third party sites such as directories and social bookmarking sites. A separate study was made into 50 non-optimized web pages for comparison.

Findings – The paper provides insight into the techniques that successful Search Engine Optimizers use to ensure a page ranks highly in Google. Recognizes the importance of PageRank and links as well as directories and social bookmarking sites.

Research limitations/implications – Only the top 50 web sites for a specific query were analyzed. Analysing more web sites and comparing with similar studies in different competition would provide more concrete results.

Practical implications – The paper offers a revealing insight into the techniques used by industry experts to rank highly in Google, and the success or otherwise of those techniques.

Originality/value – This paper fulfils an identified need for web sites and e-commerce sites keen to attract a wider web audience.

Keywords Worldwide web, Search engines, Directories

Paper type Research paper

Introduction

As the world wide web has matured, search engines have occupied an increasingly powerful position, by both channelling the attention of millions of users, and generating revenue for web sites through contextual advertising programmes, such as Google's AdSense (Google, 2005). The search engine companies are in a powerful position in the online world. Indeed, such is the popularity of the search engine that over half of all visitors to a web site now come from a search engine rather than from a direct link on another web page (McCarthy, 2006). With search engines collectively handling over 4.5 billion user queries a month (Nielsen-NetRatings, 2005), there is fierce competition amongst competing web sites to attract those users to their site at the expense of their competitors.

However, the competition is made even more ferocious by the searching behaviour of the user. Search engines may return many millions of documents for each user query, but the user only looks at a select few. Indeed, according to Jansen and Spink (2006), 73 percent of search engine users never look beyond the first page of returned

Internet Research
Vol. 17 No. 1, 2007
pp. 21-37

1066-2243
DOI 10.1108/10662240710730470

results. Accordingly, the competition for a high ranking for popular user queries is now extremely intense.

Understanding which factors can influence a page's ranking in a search engine is therefore crucial for any web site that wishes to attract large numbers of users (in particular, e-commerce sites). This paper therefore sets out to identify the most effective techniques that can be used.

In order to do this, the paper presents the results from an analysis of the most successful pages that were created as part of a Search Engine Optimization (SEO) competition (SEO is the process of trying to rank highly a given web page or domain for specific keywords). Because all of these pages are highly optimized, the resultant set of data represents an aggregation of the most popular (and thus implicitly, the most effective) techniques used by the most successful Search Engine Optimizers in operation today.

The paper is presented as follows. Section 2 discusses previous research that has been conducted in the area, both from academia and from industry. Section 3 discusses the problems faced in identifying the factors that are used by search engines when determining a specific web page's rank. Section 4 sets out the design and methodology of the analysis, and lists the search engine factors that will be analyzed. Finally, section 5 presents and discusses the results, and summarizes the most effective techniques identified during the study before the paper concludes with potential issues and a summary of the conclusions.

Related studies

Academic studies

Pringle *et al.* (1998) examined the responses given by the InfoSeek, Excite, AltaVista and Lycos search engines to 50 single-word queries. Using decision trees and regression analysis, they concluded that a high ranking required "... informative title, headings, meta fields ... text ... important keywords in the title, headings and meta fields, but do not use excessive repetition which will be caught out" (Pringle *et al.*, 1998). However, their research is now eight years old, and three of the search engines listed no longer provide their own search results (Clay, 2005).

Khaki-Sedigh and Roudaki (2003) used a simple linear regression model to approximate the dynamics underlying Google, and thus predict the absolute PageRank of a web page. However, their model does not indicate which of the factors included are important.

Fortunato *et al.* (2006) performed a similar experiment in which they also attempt to approximate the dynamics of Google's PageRank algorithm, this time through the number of in-links (i.e. URLs that reference a particular web page). However, although they were able to show that the number of in-links is a good approximation of PageRank for popular sites, PageRank is not the only determining factor used by Google in ranking its results (Moran and Hunt, 2006). Indeed, Google now claims to be using more than 200 "signals" when determining the rank of a page, with thousands of machines involved in the ranking process for every query (Eustace, 2006).

Finally, Bifet *et al.* (2005) used many different factors in an estimation function they derived for the ranking function of a search engine. They then used their function to compare their own predicted rankings with the actual rankings of Google. They found a variety of factors affected the rankings, depending, seemingly, on the subject being

searched for. For example, queries classified as "Art" ranked well if the content had a low fraction of non-English words, whereas queries containing only the name of US states ranked well if they had many in-links. However, the Support Vector Machine they used to obtain their function did not work as well as they would have liked, leading to inconclusive results. Furthermore, the queries they chose were somewhat arbitrary, and do not reflect the typical user query.

Commercial studies

Away from the research establishment, a new industry has emerged called Search Engine Optimization (SEO), which seeks to determine the most important factors to be used to get a high ranking, and then apply those factors to a client's web site for a fee. However, despite the large proliferation of such companies (e.g. Bruce Clay Inc., HighRankings.com, SearchEngineWorld.com, SearchEngineWatch.com, MarketLeap.com, SEOMoz.org, etc.), they only have partial information of a search engine's heuristics, based largely on trial and error (Fortunato *et al.*, 2006).

Although undoubtedly some SEO companies have inferred the heuristics accurately, there is too much conflicting information on the Web to determine the accuracy of the claims of any one individual company without further evidence.

Issues with inferring search engine-ranking factors

From the literature (see, for example, Moran and Hunt (2006), Fortunato *et al.* (2006), Bifet *et al.* (2005)), the web factors that could potentially influence a search engine's ranking of a web site can be classified according to two distinct categories: Query-Factors, which rely on the content of a web page, such as the existence and frequency of keywords; and Query-Independent Factors, which rely on information from external web pages that link to a web page under consideration.

However, both types of factor are notoriously difficult to enumerate as the search engines do not reveal which particular ones they use when determining a web site's ranking. Worse, the problem is compounded by the following issues:

- There are over 200 different factors (or signals) used by Google to calculate a page's rank.
- What these factors are is unknown, as is the weighting of each factor towards the final rank.
- The weighting of each factor used to determine the top ten results may be different from the weighting used for the remainder.
- Different query terms may employ different ranking factors and/or different weights (Bifet *et al.*, 2005).
- Google has multiple data-centres distributed across the world, not all of which are in sync at any one time. Thus the ranking algorithm used in one data-centre may change subtly from the ranking algorithm used in another (Cutts, 2006).

This makes identifying the factors involved in a search engine's ranking algorithm extremely difficult without a large dataset of millions of Search Engine Results Pages (SERPs) and extremely sophisticated data-mining techniques.

A novel approach to identifying ranking factors

Although identifying the ranking factors is extremely difficult to infer, and the claims made by individual SEO companies difficult to verify, an understanding of the most effective techniques can be achieved by analyzing a set of highly optimized web pages created by a host of the leading SEO companies and individuals.

These web pages can easily be found by entering a specially constructed query into any search engine. This query contains the keywords *V7ndotcom Elursrebmem,* which was defined by the industry-leading SEO web site: www.v7n.com in a SEO competition it ran between January 15 2006 and May 15 2006 (Scott, 2005). The aim of the competition was to see who could rank highest for this particular query by noon on May 15 2006.

The keywords in the query were constructed in such a way as to ensure there were no existing pages that would rank for this query before the competition began, and the only pages that would ever rank for it would be those that would be competing in the competition.

The participants were leading SEO companies and individuals. Winning the contest meant not only receiving a cash prize, but also the personal and commercial kudos that comes from being seen as the best SEO in the business. Consequently, the competition was fierce, and every page returned for *V7ndotcom Elursrebmem* is highly optimized. Those pages that are returned in the top 50 results, therefore, should reveal the techniques used by the top SEOs in the field.

Although this approach will not elicit a definitive list of the factors used in a search engine's ranking algorithm, it will provide an insight into the most popular and effective techniques used by today's leading SEOs. As these techniques clearly work, the end result should be just as effective as knowing the set of ranking factors used by the search engines.

For comparison with this set of results, a separate analysis of the results returned for a different, regular query will also be conducted. For this second analysis, the query mobile phones will be used. This query was chosen as mobile phones are extremely popular devices, with over 2 billion in existence (Wireless Intelligence, 2005); are complex enough to require the user to seek information before purchasing them; have acquired a large number of both corporate and independent web sites; and are one of the more popular items searched for by users. According to Yahoo's Overture Keyword Selector Tool (Yahoo, 2006), which lists the number of searches for a specific query submitted to Yahoo in a month, the query was used 4,666,114 times in September 2006. As such, mobile phones are an extremely popular query, and should serve this analysis well.

Experimental design and methodology

Defining the factors to analyze

The search engine that will be used in this analysis is Google, as it is by far the most popular, handling 46.2 percent of all search queries (Nielsen-NetRatings, 2005). Of the 200 or so factors that Google claim they use when determining a page's rank, the following have been chosen as representing the factors that most likely exert the greatest influence on a page's rank:

- *Number of web pages in a site indexed by search engine.* Some web sites are bigger than others by several orders of magnitude. Bigger may be better as far as rankings are concerned.
- *PageRank of a web site.* Google's PageRank (Brin and Page, 1998) algorithm helps rank web sites according to the number of in-links, and the calculated authority of each site providing the in-link. Generally, the higher a site's PageRank, the higher its ranking (and the more authority it can confer to other sites it links to).
- *Number of in-links to a web site.* Fortunato *et al.* (2006) speculate that PageRank can be substituted by in-links as a good approximation of rank.
- *Age of the web site's domain name.* The SEO community currently speculates that older domain names will rank more highly than newer domain names for the same content (WebConfs, 2006).
- *Listing in Yahoo and DMoz directories.* Both Yahoo and DMoz.org (the Open Directory) are human-edited directories whose results feed into directories from other search engine companies such as Yahoo and Google, respectively. Because of the high quality control of these directories, the sites they list are deemed to be of high authority, which the search engines may use as one of their ranking factors.
- *Number of pages listed in Del.icio.us.* Del.icio.us is a social bookmarking site that enables anyone to bookmark a page. Because of its popularity and the fact that a bookmark can be interpreted as an implicit recommendation of a page, the number of different people who have bookmarked a specific page may add to that page's ranking.

Capturing the data

The data was captured using a utility called SEO for Firefox from SEOBook.com (SEOBook, 2006). This extremely useful tool captures a variety of SEO-specific information for a particular query, and inserts that information into the results page returned from either Google or Yahoo.

It should be noted that the link analysis component of SEO for Firefox (and thus of this study) is based on links from Yahoo. This is because the link: operator used by Google to return the number of links pointing to a page is notoriously unreliable, whereas the Yahoo equivalent is much more accurate. However, although the exact number of links to a particular page recorded by Google and Yahoo may differ, the relative difference will remain the same for all links to pages. Thus as long Yahoo is used consistently for all link analysis, the results will remain valid.

Results and analysis

The following sections present the results of the study. For clarity, the results from the v7n query will be called the v7n set, while the results from the mobile phones query will be called the mobile phones *set*.

Analysis of the top ten results for the query V7ndotcom Elursrebmem

Table I lists the results of the analysis of the top 10 results for the query *V7ndotcom Elursrebmem*. The winner was Scott Jones's site called, unsurprisingly, "V7ndotcom

Table I.
Top ten results for the query *V7ndotcom Elursrebmem*

Google rank	Pages indexed	PR	Age	Del.icio.us	Yahoo domain links	Yahoo page links	Alexa	Dmoz
1	280	6	02-2005	13	51,800	13,700	27,686	2
2	8	5	–	2	1,190	1,160	1,168,618	0
3	27	4	12-2003	7	89,700	418	104,168	2
4	2	5	–	0	8,770	8,770	2,768,345	0
5	51	7	–	5	11,300	14,600	510,945	2
6	303	5	–	2	2,930	2,870	663,232	0
7	106	4	–	2	18,100	4,870	221,113	0
8	21,600	6	03-2002	70	157,000	32,700	2,065	1
9	160	5	–	0	11,100	2,960	663,232	0
10	74	4	02-1998	2	9,280	292	351,597	0

Elursrebmem" (www.v7ndotcomelursrebmem.net). However, as if to confirm the uncertainty of search engine rankings, this site was number 1 on very few Google datacenters, which had been slow to update. The majority of the datacenters showed Jim Westergen's site (www.jimwestergren.com/v7ndotcom-elursrebmem/) appearing at number 1 at the time the competition ended. However, when the search was made at noon on May 15 2006, it was one of the older datacenters that served the results, giving Scott Jones the win. Just 15 minutes later, this datacenter had updated, placing Jim Westergen's site at no. 1 (v7n, 2006), where it still ranks today. SEO is anything but a precise science.

At first glance, the results presented in Table I, and indeed of the top 50, show a wide variance for each individual factor. However, the techniques used by each SEO competitor become clearer, as the following analysis shows.

Number of pages indexed

For the top ten results, the number of pages of individual sites indexed by Google range from two to 21,600. Widening the result set to the top 50, this range increases from two to 334,000, with some SEO competitors clearly attempting to influence the rankings through sheer volume. However, with the second placed competitor having only eight pages indexed, a high volume of pages is clearly not needed to rank highly – quality seemingly counts over quality.

That said, an analysis of the top 50 shows that creating a large number of pages is a technique used by many SEOs, with some success. Figure 1 shows the number of pages indexed for the top 50 (number of pages shown on a logarithmic scale). The majority of pages ranked in the top 27 clearly have more pages indexed than those that rank between 28 and 50.

For comparison, Figure 1 shows the number of pages indexed for the mobile phones set (again, *y* axis uses logarithmic scale). This reveals a much more uniform distribution of pages. The outliers on this graph are for pages from wikipedia (55,900,000 pages, ranked no. 2), Google (9,450,000 pages for its SMS service, ranked number 18), Amazon (116,000,000 pages indexed, ranked number 37) and Google again at rank 44.

Such enormous sites tend to appear highly in the rankings for many different terms due to their extreme size and the large number of pages containing high quality

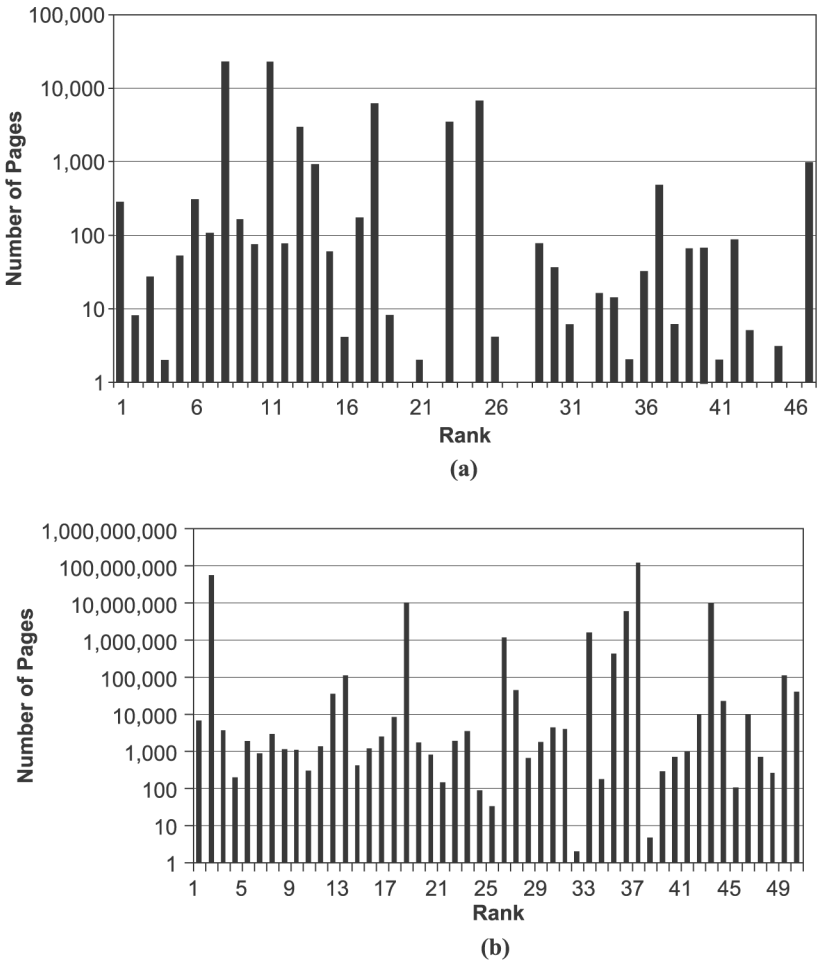

Figure 1.
Number of indexed pages (a) v7n set; (b) mobile phones set

content, and thus can skew the data. It is therefore important to identify exactly which sites represent the outliers in the data before determining whether or not they are representative of the results being studied.

In this case, the outliers can mostly be ignored, as it is the general trend under observation, but other researchers should analyze carefully the properties of the sites they are analyzing, as it is difficult to compare equivalent sites without first identifying the type of site being analyzed. Evans (2006) describes a method for identifying similar sites for comparative analysis.

Conclusion: volume of pages is a factor employed by many SEOs, but with limited results. Google's claim that high quality content beats low quality seems to be borne out (although spammers can still get high results if they are not too obvious in other areas).

PageRank of a web site

The PageRank of the top ten from the v7n set ranged from PageRank 4 (PR4) through to PR7. Figure 2 shows the frequency distribution of PageRank, which clearly shows how important PageRank is to a page's ranking. For example, no page with a PageRank less than 4 ranked at all within the top 40. However, despite the obvious importance of PageRank, it is impossible to state that a specific page with a certain PageRank will rank higher than other pages with a lower PageRank; only that high PageRank pages tend to rank higher than lower PageRank pages.

Comparing the PageRank distribution for the v7n set (Figure 2) with the mobile phones set reveals a broader distribution of the PageRank values for the mobile phones set. This is due to different types of web site all ranking highly for the query mobile phones, each with its own individual properties that will impact upon a search engine's ranking algorithm. For example, some web sites may rank highly without the whole site being specific to mobile phones (e.g. Wikipedia, ranked no. 2 with PR6, but over

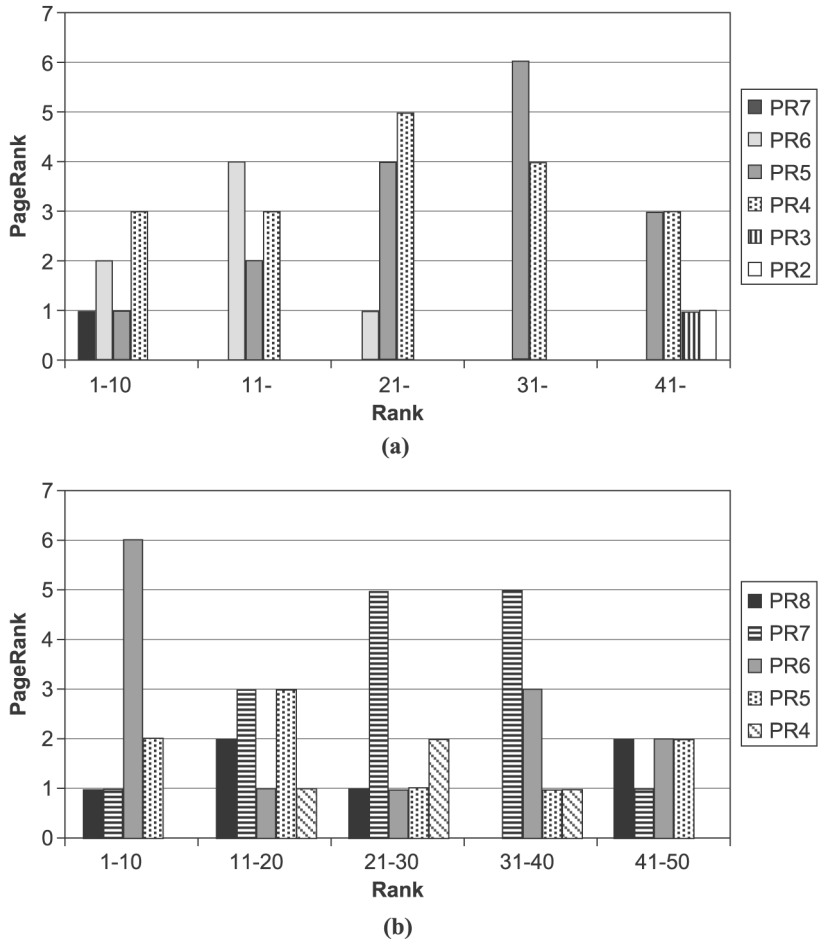

Figure 2.
PageRank frequency distribution (a) v7n set; (b) mobile phones set

55,000,000 pages indexed, only a very small proportion of which are relevant to the mobile phones query).

In contrast, the v7n set is more focused, with every page specific to the v7n competition, and thus to the v7n query (see Figure 3).

Analyzing the percentage of pages having a specific PageRank shows a disparity between the two sets of results. The mobile phones set has an average PageRank 5.84, compared with 4.5 for v7n, reflecting the longer amount of time the mobile phone pages have had to earn their high PageRanks.

More interesting is the broader distribution of PageRank values for the mobile phone set. With the v7n set, the standard deviation is just 0.16, with 78 percent of the pages having a PR4 or PR5. This compares with the mobile phone set, which has a standard deviation of 0.89. It's clear from this that attaining a high PageRank was obviously a tactic employed by the SEOs, but given the short amount of time they had to accumulate their PageRank value, PR5 was the highest most of them could achieve

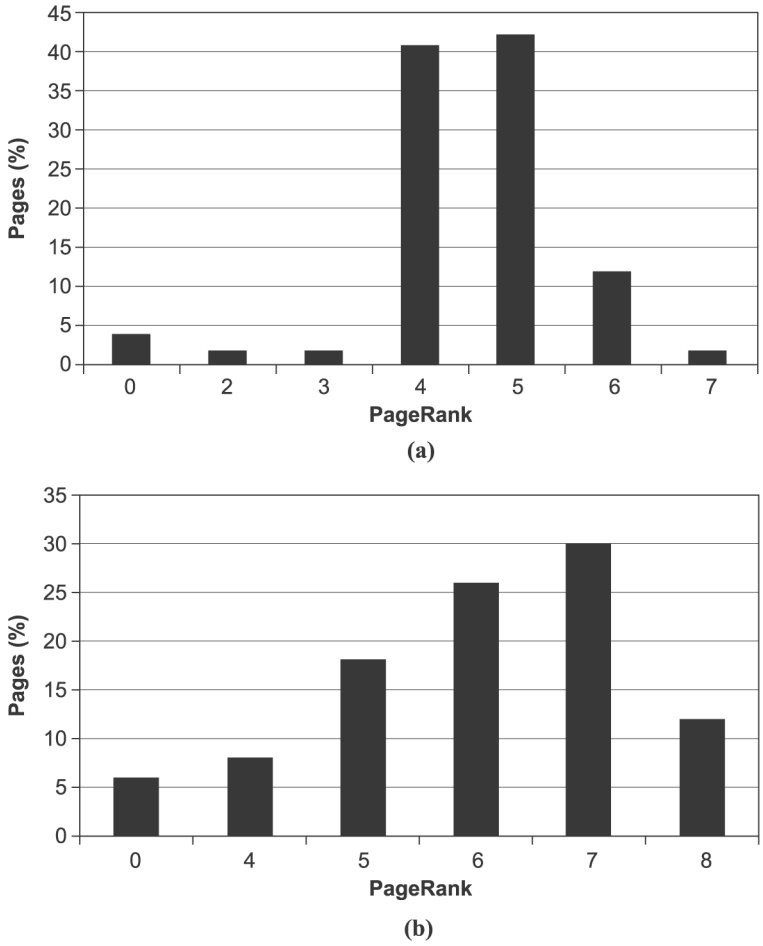

Figure 3.
Percentage of pages with specific page rank (a) v7n set; (b) mobile phones set

(only one achieved the highest value in the set of PR7, which was ranked at no. 5). In contrast, PR accumulated naturally for the mobile phone set over a long period of time (ten years, in some cases), hence the broader distribution of values.

Conclusion: PageRank is still extremely important in ranking highly, but a high PageRank will only make it probable that your page will rank highly. Other factors play a role that may negate a high PageRank.

Number of in-links

Figure 4 shows the number of page in-links for the v7n set (with the removal of the outlier at rank 26, which, with 173,000 in-links, is some five times that of its nearest competitor). Note that these figures reflect the number of in-links to a specific page, rather than to the whole web site. The trend clearly shows a decline in the number of in-links as the rankings fall.

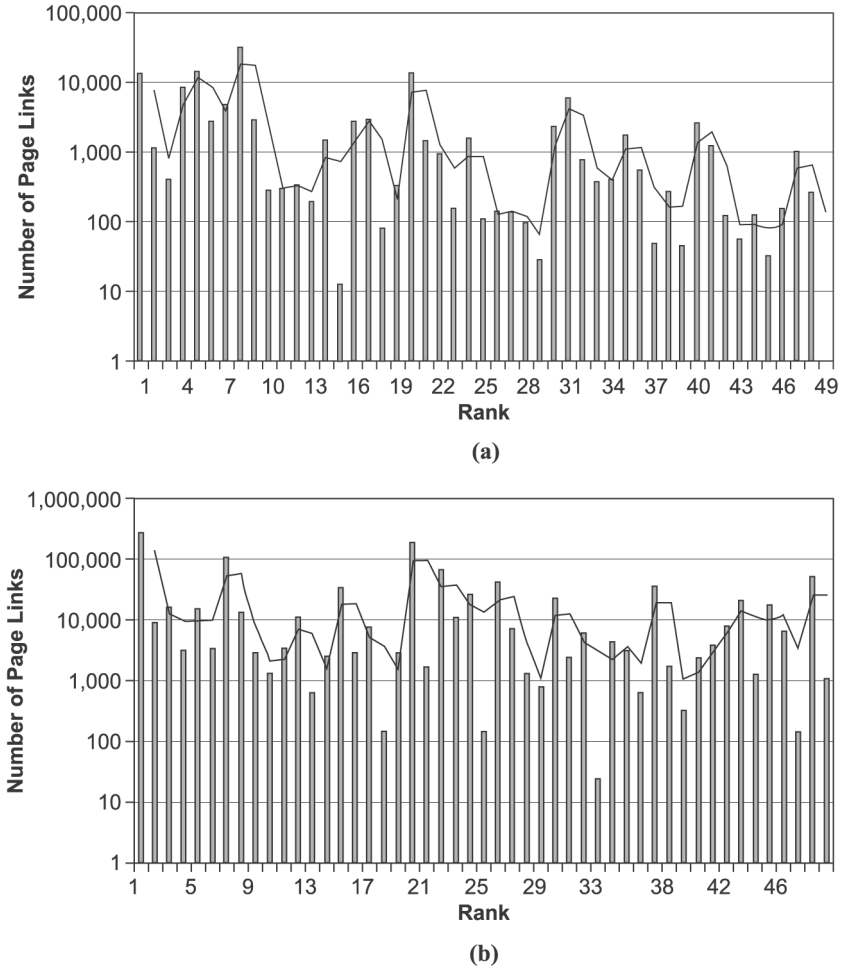

Figure 4.
Number of in links to page (a) v7n set; (b) mobile phones set

In contrast, there is no apparent declining trend for mobile phones (Figure 4). This can be explained by the number of pages returned by Google for the mobile phones query (83.7 million), and the length of time the top 50 mobile phone sites have had to accumulate in-links (up to ten years). The trend would surely be downwards as the rank number increases, and it would be interesting to see at which point in the rankings the downward trend becomes apparent for such a mature query.

Domain age

Domain age (i.e. the date at which each domain was registered) has been posited as an important factor in the ranking of a site, as older domain names are said to be inferred by Google's ranking algorithm as conveying more trust, and therefore should rank higher than newer domains.

Analysing the domain age for both the v7n and mobile phone sets reveals a marked difference between the two, in terms of the modal average age. For the v7n set, the modal average of first registration is 2004 (Figure 5), corresponding to an age of two years, whilst for the mobile phones set, the average is 2000, or six years (Figure 5). Note that the data for the v7n set is limited to 17 domains out of the 50 web sites analyzed, as no data could be found for the remaining 33.

This discrepancy is clearly to be expected, given that the v7n contest only began on January 15 2006, and most of the web sites in the top 50 were created specifically for the contest. What is interesting, however, is that despite this, some of the web sites' ages are up to ten years old! This occurs because the SEOs used an old domain that had been registered many years before the contest began, and populated it with content specific to the v7n query. As such, they were relying on the domain already having an element of trust within Google's ranking algorithm, in order to get a higher ranking than a brand new domain.

The number of SEO pages using this technique shows how popular it is amongst the SEO community. However, there is no discernible trend in the domain age results, with the number one ranked site being only a year old at the time the contest ended, while the oldest site, at ten years old, is ranked number 36.

However, the data in this particular sample is limited, so no firm conclusion can be made about the success of this technique. What can be concluded, however, is that it is a technique widely used amongst the SEO community.

Turning to the mobile phones set (Figure 5), there is a much broader distribution of ages, centred around those that are six years old. Interestingly, no web site appears in the top 100 that is less than two years old. Equally, despite there being no obvious correlation between age in the top 20, as Figure 5 shows, there is a clear trend for the average web site age for those sites ranked from 21-80. Although not conclusive proof, it does lend credence to the idea that domain age plays an important role in a page's ranking.

Conclusion: Domain age is perceived as an important factor by SEOs, and initial results suggest there may be some truth to this.

DMoz directory submissions

Figure 6 shows the number of sites listed in The Open Directory for the v7n set and the mobile phones set. The mobile phones set is consistently high, with 80 percent of sites included in the directory. In contrast, the v7n set shows a marked difference between

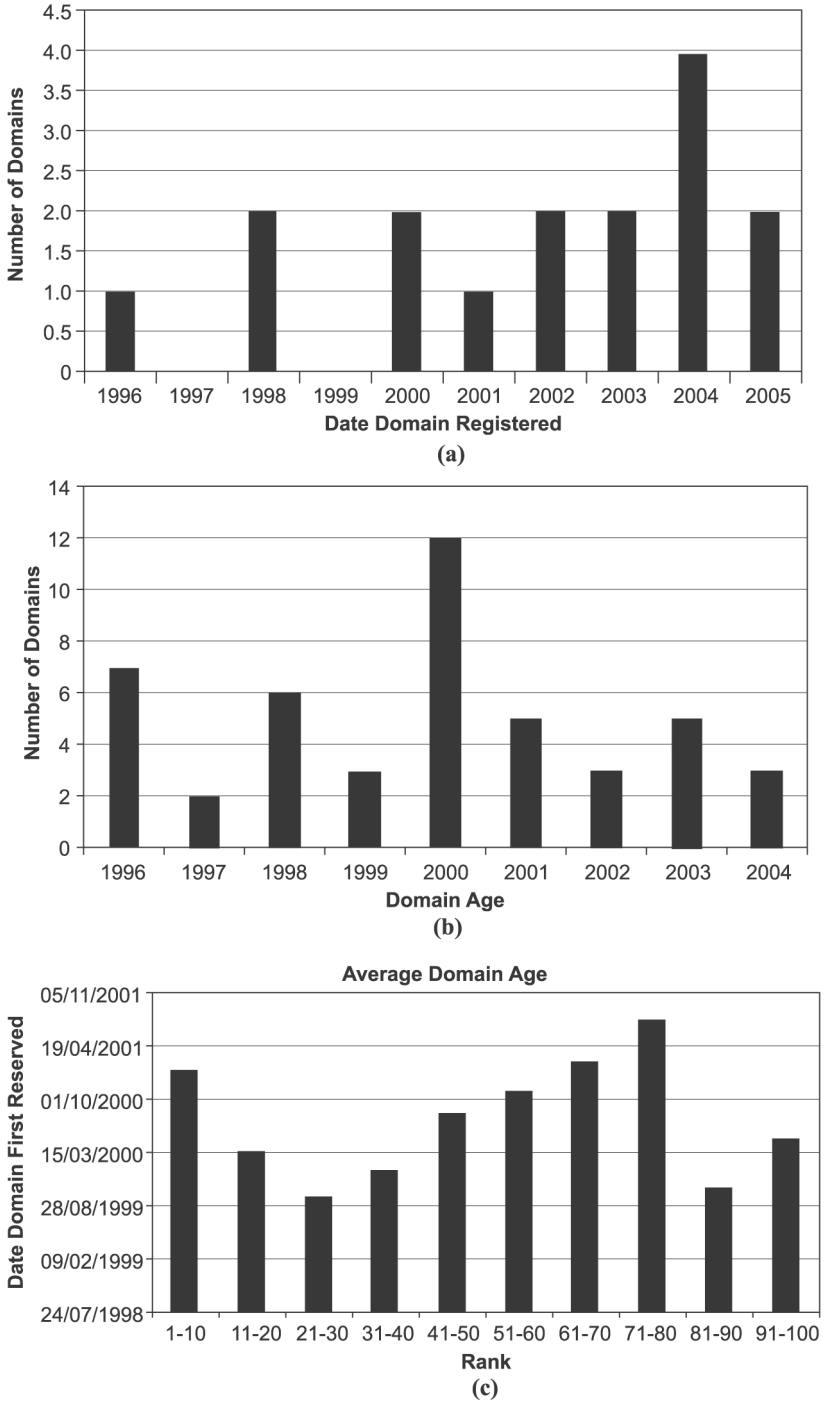

Figure 5.
Domain age frequency (a) v7n set; (b) mobile phones set. Average domain age, mobile phones set

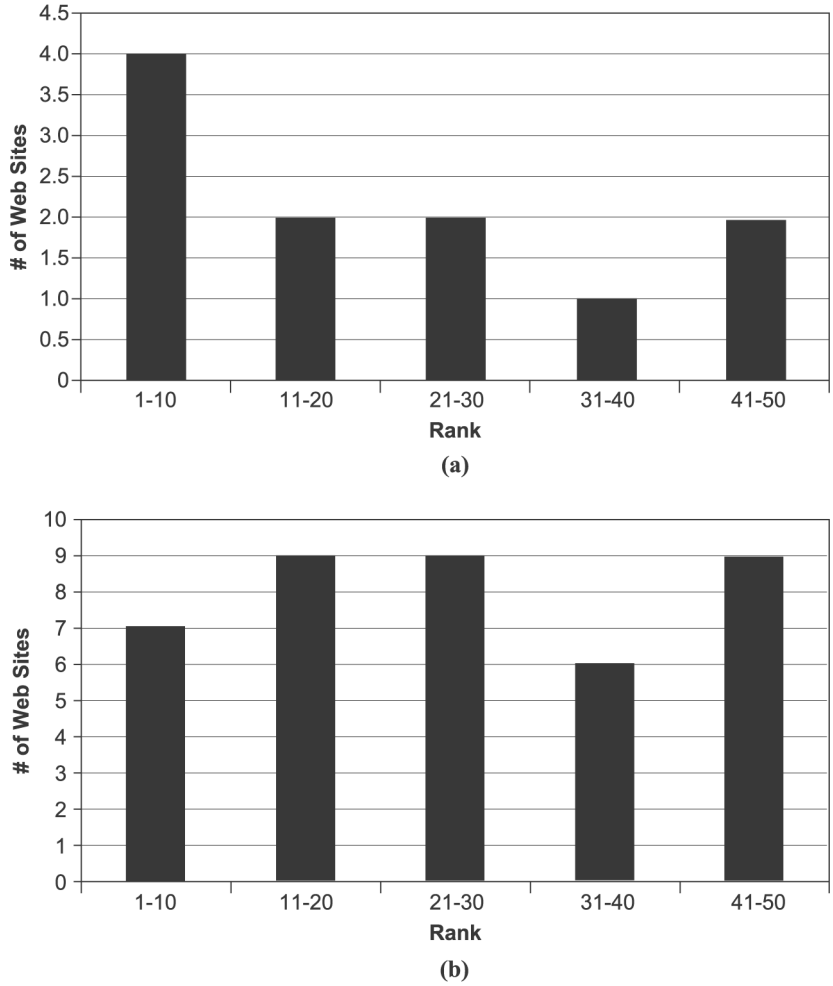

Figure 6.
Number of web sites listed in DMoz (a) v7n set; (b) mobile phones

the top 10 sites and the remaining sites, and only 22 percent of sites in the whole set being included.

Being listed in DMoz is notoriously difficult, however, with lead times of six months to a year before an entry submission is actually included, due to the fact that human volunteers must judge each and every entry. With the v7n contest only running for four months, it is to be expected that so few sites were listed. It is interesting to note, however, that of those listed, over twice as many rank in the top ten as for the remaining ranks.

Conclusion: Being listed in DMoz is a technique employed by the successful SEOs.

Yahoo directory submission

The results of the Yahoo directory submission analysis were less conclusive, as so few of the v7n set had a Yahoo directory entry. Only 14 percent of sites were listed,

compared with 90 percent of the mobile phones set. The reason for this is presumably the fact that entry into the Yahoo directory costs $300 and again may take several months for a site to be listed. Consequently, with a high initial outlay and no guarantee that an entry will even appear in the Yahoo directory in time for the contest's closure date, it appears that very few SEOs attempted this option. As such, no conclusion can be drawn on this result.

Del.icio.us bookmarks

The number of pages bookmarked in Del.icio.us for both the v7n set and the mobile phone set also shows a notable disparity. Del.icio.us links will only appear if people choose to bookmark them. Although this appears to be a fail-safe way of determining a page's popularity, a bookmark does not, of course, give any indication of the intention of the bookmarker. As such, a page can appear popular simply by the page's author encouraging as many people as possible to bookmark it for reasons other than popularity. One example might be a plea to "add this page to del.icio.us to help me win this SEO contest!"

As can be seen in Figure 7, there does appear to be a trend in the number of sites being bookmarked in Del.icio.us for the v7n set, with more of the high ranking sites appearing in del.icio.us than the lower ranking sites. This does not imply that a high ranking in del.icio.us necessarily affects the ranking; just that the successful SEOs chose to use del.icio.us as a high-ranking technique more frequently than the less successful SEOs.

That said, the number of bookmarks gained from del.icio.us for each site was low, with only nine out of the 50 web sites receiving more than ten bookmarks. However, one site stands out with 1,562 bookmarks (rank number 25), although the site itself (Google Blogoscoped (Lenssen, 2003)) is a blog that has been running since 2003 and which has a large readership, rather than a site designed specifically for the contest.

In contrast, the number of pages with del.icio.us bookmarks in the mobile phones set is very high, with only four out of the top 50 sites not having any links at all. The number of bookmarks is also much greater per page, with only 16 having fewer than ten bookmarks, and the range varying from 0 to 7,940.

Conclusion: 92 percent of the top 50 pages in the mobile phones set have del.icio.us links, while only 54 percent of the v7n set do. However, of the v7n set, there is a clear trend showing more del.icio.us bookmarks the higher the ranking. As such, attracting del.icio.us bookmarks would appear to be a technique used by the more successful SEOs, but it cannot be said that del.icio.us bookmarks confer high ranking.

Issues and further work

This study has focused primarily on Query Independent Factors and ignored Query Factors. It would be interesting to conduct a similar study for both Query Factors and Query Independent Factors. However, with over 200 potential factors that could be studied, the time and effort for such a study was beyond the current resources available.

It would also be interesting to compare the results of the v7n set with a similar set of results from other competitions to see if similar patterns emerge. As such, the author plans to capture data from future SEO contests in real time, both to compare the

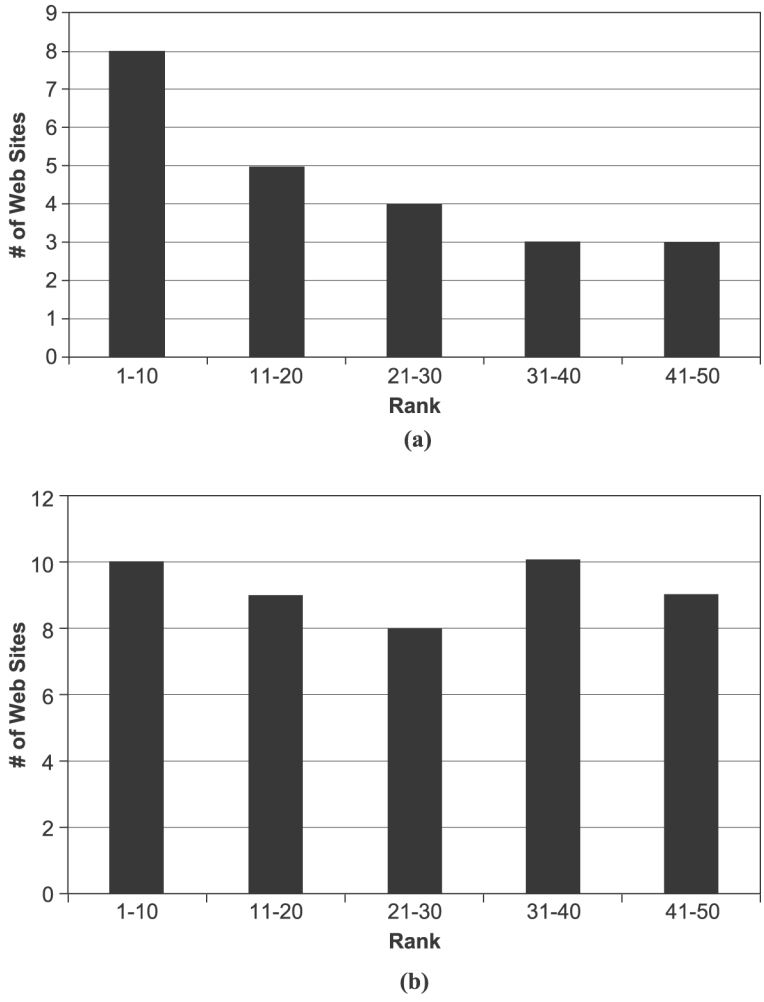

Figure 7.
Number of web sites with Del.icio.us bookmarks (a) v7n set; mobile phones set

techniques used with the current v7n set, and also to see how and when the techniques are deployed.

Conclusion

This paper has presented the results of a study into the techniques used by top SEOs to rank their web pages no. 1 in a SEO competition. After describing the experimental design and methodology used, the results of the study were as follows:

- Many SEOs generated many pages to influence rankings, which proved a partial, if limited, success.
- High PageRank in Google clearly plays a major part in a page's rankings, and attaining a high PageRank was a goal of most of the SEOs. However a PR of a particular rank will not necessarily rank higher than a PR of a lower rank.

- The more successful SEOs attracted many in-links to their page, with a clear trend showing declining in-links for lower rankings. Accordingly, attracting many in-links is another technique used by SEOs that would appear to have a good deal of success.
- A listing in DMoz is a technique favoured by the more successful SEOs.
- Many SEOs use older domains for higher rankings, and there may be truth that this is a successful technique.
- The more successful pages had more del.icio.us bookmarks.

References

Bifet, A., Castillo, C., Chirita, P.-A. and Weber, I. (2005), "An analysis of factors used in search engine ranking", *Proceedings of the Workshop on Adversarial IR on the Web, Chiba, 10-14 May*.

Brin, S. and Page, L. (1998), "The anatomy of a large-scale hypertextual (web) search engine", *Computer Networks and ISDN Systems*, Vol. 30 Nos 1-7, pp. 107-17.

Clay, B. (2005), *Search Engine Relationship Chart*, Bruce Clay Inc., Moorpark, CA, available at: www.bruceclay.com/searchenginerelationshipchart.htm

Cutts, M. (2006), *More Info on Page Rank*, MattCutts.com, available at: www.mattcutts.com/blog/more-info-on-pagerank/

Eustace, A. (2006), *Search Technology Overview*, Google Press Day, May, available at: http://blog.searchenginewatch.com/blog/060510-123802

Evans, M.P. (2006), "Analyzing Google's rankings using web site query-independent factor analysis", *Proceedings of the 6th International Network Conference (INC 2006), Plymouth, 11-14 July*.

Fortunato, S., Boguna, M., Flammini, A. and Menczer, F. (2006), "How to make the top ten: approximating PageRank from In-degree", paper presented at the 14th International World Wide Conference, Edinburgh, May 22-26, available at: http://xxx.lanl.gov/PS_cache/cs/pdf/0511/0511016.pdf

Google (2005), *AdSense Contextual Advertising Programme*, available at: www.google.com/ads.

Jansen, B.J. and Spink, A. (2006), "How are we searching the world wide web? A comparison of nine search engine transaction logs", *Information Processing and Management*, No. 42, pp. 248-63.

Khaki-Sedigh, A. and Roudaki, M. (2003), "Identification of the dynamics of the Google ranking algorithm", paper presented at the 13th IFAC Symposium On System Identification, available at: www.iranseo.com/ studies/google_ranking_algorithm.pdf

Lenssen, P. (2003), *Google Blogoscoped*, available at: http://blog.outer-court.com/archive/2006-01-15-n84.html

McCarthy, J. (2006), *User Navigation Behavior to Affect Link Popularity*, quoted in Search Engine Roundtable, available at: www.seroundtable.com/archives/001901.html

Moran, M. and Hunt, B. (2006), *Search Engine Marketing, Inc. – Driving Search Traffic to your Company's Web Site*, IBM Press, Armonk, NY.

Nielsen-NetRatings (2005), *Nielsen NetRatings Search Engine Ratings*, provided to SearchEngineWatch, July, available at: http://searchenginewatch.com/reports/article.php/2156451

Pringle, G., Allison, L. and Dowe, D.L. (1998), "What is a tall poppy among web pages?", *Proceedings of the 7th International World Wide Web Conference, Brisbane, April*,

pp. 369-377, available at: www.csse.monash.edu.au/~ lloyd/tilde/InterNet/Search/1998_WWW7.html

Scott (2005), *SEo Contest, V7n Forum*, available at: www.v7n.com/forums/seo-forum/22836-seo-contest.html

SEOBook (2006), *SEO for Firefox Tool*, available at: http://tools.seobook.com/firefox/seo-for-firefox.html

v7n (2006), *v7n Forums*, available at: www.v7n.com/forums/326411-post10.htm

WebConfs (2006), *The Age of a Domain*, available at: www.webconfs.com/age-of-domain-and-serps-article-6.php

Wireless Intelligence (2005), *Mobile Phone Industry Report*, available at: www.wirelessintelligence.com

Yahoo (2006), *Keyword Selector Tool*, available at: http://inventory.overture.com/d/searchinventory/suggestion/

Corresponding author
Michael P. Evans can be contacted at: michael.evans@reading.ac.uk