Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Processing and Management 41 (2005) 665–690
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

The impact of webpage content characteristics on webpage visibility in search engine results (Part I) ☆

Jin Zhang *, Alexandra Dimitroff

School of Information Studies, University of Wisconsin Milwaukee, Milwaukee, WI 53211, USA

Received 12 August 2003; accepted 1 December 2003
Available online 21 January 2004

Abstract

Content characteristics of a webpage include factors such as keyword position in a webpage, keyword duplication, layout, and their combination. These factors may impact webpage visibility in a search engine. Four hypotheses are presented relating to the impact of selected content characteristics on webpage visibility in search engine results lists. Webpage visibility can be improved by increasing the frequency of keywords in the title, in the full-text and in both the title and full-text.

Keywords: Search engine optimization; Webpage placement; Webpage visibility

1. Introduction

The proliferation of computing and networking techniques has made it possible for users across the world to access internet sources and electronically publish information on the internet. The world of the internet was transformed with the development in the mid-1990s of search engines. These tools provided access to the overwhelming number of resources on the Web to not only academic users, but increasingly to the general public and commercial enterprises. It is estimated that more than 1.3 billion websites are available on the internet, and over 1 million new websites are added to it every year (Ambergreen, 2002). Over 30,000 search engines more than 95% of the internet search traffic and 80% of internet users search for information on the internet via search engines (Haltley, 2002). Most users usually examine only the top 10 websites in a search engine results list and only 1% of users check beyond the third page of a search engine results list

☆ The research was supported by an IMLS National Leadership Grant (#NR-10012-01).
* Corresponding author.
E-mail addresses: jzhang@uwm.edu (J. Zhang), dimitrof@uwm.edu (A. Dimitroff).

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2003.12.001

(Ambergreen, 2002). Not surprisingly, information science researchers as well as a growing group of information entrepreneurs began evaluating the performance of various search engines in the late 1990s. Reviews of the literature about search engine performance focus on one viewpoint: that of the user (Leighton, 2003; Oppenheim, Morris, McKnight, & Lowley, 2000; Schwartz, 1998). This focus addresses the needs of only half of the internet user community. The study described here looked at the other primary internet user community: webpage publishers.

Internet users can be categorized into two broad groups: end user searchers and webpage publishers. The first group's priority is to locate information on the internet conveniently and accurately. Information browsing and information searching are two primary means. The former relies on a well-organized subject directory system while the latter rests on a search engine. Most of the time, these users prefer to employ a search engine to do the job. The second group's focus is the creation of webpages and the publication of them on the internet. This group's priority is to maximize the probability that their published websites are indexed by search engines and that they appear high on searchers' search engine results lists. With the creation of digital access sources and services in all sorts of environments—libraries, businesses, government agencies, non-profit organizations, museums, to name a few—insuring that an end user searcher finds a particular website is becoming increasingly difficult. Information organizations—those institutions that have traditionally provided the organizational and access tools for information seekers—are now dealing with an increasingly complex digital world. While providing a discrete address to a collection of digitized information mimics traditional access to collections, the distributed nature of information retrieval requires that information institutions consider additional means of providing access to their resources.

Search Engine Optimization (SEO), or search engine positioning, is the process of identifying factors in a webpage which would impact search engine accessibility to it and fine-tuning the many elements of a website so it can achieve the highest possible visibility when a search engine responds to a relevant query. Search engine optimization aims at achieving good search engine accessibility for webpages, high visibility in a search engine result, and improvement of the chances the webpages are retrieved. Search engine optimization is a difficult task, far more intricate and complex than one would expect, particularly since different search engines have different indexing strategies and ranking algorithms.

There are various factors which can contribute to visibility of a webpage in a search engine results list, for example, webpage metadata structure, webpage content, hyperlink cited status, search query expansion, and other possible factors. A metadata system is a system used to describe a webpage for a variety of reasons. Webpage content is simply determined by words on the webpage itself. Hyperlink cited status of a webpage refers primarily to the number of webpages on the internet that hyperlink or cite to a particular webpage: the more pages hyperlink a webpage, the better the hyperlink cited status, and vice versa. Hyperlink cited status of a webpage is a variable that may affect its visibility in a search return list. Since a webpage with high hyperlink status usually is considered to be more important or influential than other pages with low hyperlink status, some search engine ranking algorithms take it into consideration, making result ranking appear to be more relevant. In other words, a returned webpage with a better hyperlink cited status would be ranked higher than other returned pages.

Query expansion also affects webpage visibility in a search engine from a quite different perspective. The internet search process is an interactive process between a human being and a search

engine. It is a complex process affected by multiple variables. During this interactive process, an initial query may be changed, modified, or revised, moving toward a more effective, and well-defined query. Some search engines monitor, analyze, and use users' query expansion information as a factor for webpage visibility calculation.

These factors can be grouped into two basic categories. The first group includes webpage metadata structure and webpage content. These factors are internal and are determined by the webpage itself. They can be obtained or parsed from a webpage. The second group includes hyperlink cited status, query expansion, and possible others. These factors are external to the webpage and cannot be obtained from the webpage itself. The factors in the first group can be controlled and manipulated by webpage designers or developers due to their internal nature. They should be primary factors in optimizing the visibility of a webpage in a search engine results list. The factors in the second group cannot be controlled and managed by the webpage designers or developers because of their external nature. That is, hyperlink cited status of a webpage totally depends on whether other websites cite or hyperlink to a webpage. Query expansion relies on users' search behavior.

Obviously, a webpage designer cannot control an internet searcher's behavior and cannot change webpage hyperlink cited status. He/she can only control the internal factors identified in the first group. For this reason, only variables in the first group are considered in this study; the external variables in the second group were excluded and isolated from the study to eliminate their possible interference. The intent was to strengthen the findings of the examination of the internal factors.

An increasing number of websites are turning to search engines as their primary marketing route (Centaur Communication, 2002). Driven by this trend, search engine optimization is a booming field for entrepreneurs. Hundreds of companies offer search engine optimization services to help enhance customers' online experiences by pushing relevant websites to the fore (Kanaley, 2002) (for example, Search Engine Optimization Free, 2003, (http://hotwired.lycos.com/webmonkey/01/23/index1a.html); Search Engine Optimization Tips (http://www.submit-it.com/subopt.htm); How search engines rank Webpages (http://searchenginewatch.com/webmasters/rank.html); Search Engine Submission & Search Engine Optimization (http://www.topseo.com/); Dynamic Web Ranking (http://www.hot-new.com/webrank.htm#dyn); Search engine optimization, search engine ranking, website ranking, website optimization (http://usasearchengineoptimization.info/); Search engine optimization (http://www.bruceclay.com/web_rank.htm); Search Engine Optimization Support Forums (http://www.supportforums.org/); and High Rankings Advisor (http://www.highrankings.com/advisor.htm)). These services range from free webpage optimization submission (Domain Name Express, 2003; Terra Lycos Network, 2003), to paid-optimization software (Web Position Gold, 2003; Website optimization tools, 2003), to fee webpage submission (Ihelpyou, 2002; Submit today, 2003). Free website submission cannot assure that the submitted website will end up in a good position. On the other hand, many web publishers, especially non-profit institutes or organizations, cannot afford to pay for optimization software and pricey website submission.

A growing industry has blossomed that offers advice (for a fee in most cases) on maximizing webpage placement. This advice about which techniques will provide optimal ranking results is hinted at on the internet itself but none of those offering advice provide details about any empirical research on which their recommendations might be based. While a common theme

among these advice givers is "location, location, location" the specific advice is fairly generic and based on conventional wisdom, not on tested hypotheses. This research will remedy that situation by focusing on exactly how webpage construction and posting effect ranking on results lists of various search engines.

Research on this emerging topic, on the other hand, has not been reported in research-oriented publications. Some websites offer search engine optimization tips based on their experiences (Search Engine Optimization—A 10 Step Program, 2003; Search engine optimizer, 2003) while others merely provide a basic introduction to the topic (Greenberg, 2000; Sullivan, 2003).

1.1. Purpose

The issue examined in the research described here is a universal one insofar as the use of the internet could potentially be used by anyone. The issue of site ranking within a results list is most obviously of interest to website publishers. Current "literature" (mostly prepared by commercial firms offering consulting services) focuses on the benefits to the private sector of high site ranking. However, this issue would be of equal interest to those in the non-profit sector, including libraries and museums, because of their inherent interest in disseminating information about their own institutions as well as increasing access to their various information seeking constituencies. The findings will enable institutions or organizations, particularly those involved in digital access activities or things like that nature, to better place their websites in end user searchers' results lists. The findings will help these institutions to disseminate their information products to more general searchers who use all-purpose search engines for their internet searching.

1.2. Objectives

The objectives of this research were threefold: (1) to identify webpage design factors that impact ranking in search engine results lists from the web publisher's perspective, (2) to compare the impact of those design factors in a webpage on different general search engines, and (3) to develop a practical strategy or approach to improve ranking of a webpage from an internet search engine. As mentioned earlier, only internal factors rather than external factors were considered in this study.

1.3. Research question and hypotheses

This study examined various webpage design factors and their relationship to search engine results list placement. The primary research question was: "How can the ranking of a website in a search engine's results list be improved from the webpage developer's perspective?" That is, how can the visibility of a webpage in a search engine's result be optimized? Visibility is defined as the ranking position in a search engine results list. The nearer to the top of the search results list, the better its visibility, and vice versa.

Based on the primary research question, four hypotheses were developed. These hypotheses tested the impact of webpage characteristics (not including metadata characteristics, which were examined separately and reported elsewhere) on webpage visibility. They are as follows:

(1) *Hypothesis 1 (H1)*

There are no differences in terms of search return performance among webpages with different keyword frequencies within a webpage title (i.e., <html> title), different search engines, and their interactions.

(2) *Hypothesis 2 (H2)*

There are no differences in terms of search return performance among webpages with different keyword frequencies within a webpage full-text, different search engines, and their interactions.

(3) *Hypothesis 3 (H3)*

There is no difference in terms of search return performance between webpages with keywords only in titles, the webpages with keywords only in full-texts, and those with keywords in both title and full-text.

(4) *Hypothesis 4 (H4)*

There is no difference in terms of search return performance among webpages with keyword font color, font size, webpage with keyword plural form, keyword case status, and keyword adjective form.

2. Experimental design

2.1. Creation of test webpages and posting of these webpages

2.1.1. Webpage content characteristics analysis

The key initial task was to identify webpage content characteristics of importance from the publishing point of view. In other words, any factor that might affect the return position of a search engine was identified. After they were identified, they were grouped and characterized. These webpage content characteristics factors are described below.

(a) Keyword position: We believed that position or location of a keyword within a webpage plays an important role in terms of its return visibility in a search engine. Keyword in a title and keyword in full-text were treated separately in this study.
(b) Keyword duplication: We hypothesized that keyword frequency within a webpage would make a significant contribution to its visibility in a search engine result. Duplication of a keyword can happen in a title, or a full-text. In this study, the maximum keyword frequencies for title and full-text were set to 4 and 5 respectively.
(c) Combination of these factors: Various meaningful combinations of title and full-text keywords were taken into consideration. This offered the opportunity to observe the impact of factor combinations on a hit list of a search engine.
(d) Layout: The study also attempted to investigate whether other minor factors in a webpage such as font color, font size, font case status, word plural form, and word adjective form would make a contribution to webpage return position of a search engine.

2.1.2. Creation of test webpages

Creation of content for a test webpage was the next step in the study. A public domain webpage was downloaded from the National Center for Complementary and Alternative Medicine. The

content was not copyrighted, thus avoiding any potential copyright concerns. We decided that the content of the webpage should be a manageable topic that was clear and explicit, not too narrow and not too broad in terms of content. The length of the webpage was relatively short (approximately 1100 words). The title of the webpage was "Major Domains of Complementary and Alternative Medicine".

The original webpage was processed based on the content characteristics and additional test webpages were derived from it for the study. Each derived webpage represents one of the webpage content characteristics. The content of the original webpage was slightly revised when some derived webpages were generated. These changes included discarding or adding some keywords to title, full-text, changing term forms, and inserting the investigator names into each of the test webpages.

Each derived webpage was given a unique HTML file name from which the investigators could easily trace its content characteristics during observation and data collection. It also facilitated later data analysis.

"Acupuncture" and "homeopathy" were identified as keywords representative of the webpage content. These two keywords were used as query words to search for the test webpages throughout the study.

The primary keyword "acupuncture" was used for the first three categories "combination", "duplication", and "position", and the secondary keyword "homeopathy" for the last category "layout".

2.1.3. Webpage posting

After the test webpages were prepared, they were posted in the public domain so that search engines could crawl and index them. The University of Wisconsin Milwaukee allocated a special domain for the study and we set up a special account on the University server.

2.2. Submission of posted webpage addresses

2.2.1. Identification of search engines with submission features

A pilot study was conducted to identify any unanticipated problems with the observation and data collection phases. The pilot study showed that if the test webpages were created and posted in a public domain that is available and open to all search engines, it does not necessarily mean that they would be crawled and indexed effectively by a search engine. Search engines do not treat all public domains equally. That is, not all public domains on the internet would be crawled and indexed by a search engine. In addition, the crawling frequency to each public domain a search engine crawls varies. A high profile public domain can attract more search engine and high-frequency patrons while a low profile public domain may get none. For instance, Microsoft's homepage would be more frequently visited by a search engine than other low profile personal homepages. This suggested that in order to maximize the probability of our test webpages being indexed within a reasonable time frame, a more aggressive strategy needed to be considered. This resulted in the test webpages being submitted to search engines directly rather than passively waiting for them to be visited. Fortunately, most search engines integrate a webpage submission mechanism in their main windows, which allows users to submit their webpages directly to their databases.

Note that not all search engines offer this feature. As a result we had to identify those search engines that did offer this feature. After a thorough search, 35 search engines with this feature were identified.

2.2.2. Separation of free submission search engines and fee submission search engines

The identified search engines were divided into two groups: free submission search engines and fee submission search engines. Fee submission search engines enable users to submit their http addresses to them only if they pay for the submission. These search engines were eliminated from the testing search engine list because the financial factor rather than webpage content characteristic factor would play a crucial role in the returned position of a submitted webpage. In this case (that is, search engines with submission fees), it is possible that a poorly organized webpage would pop up at the top of a search results list just because the webpage publisher had paid the search engine for that placement. The methods used by these search engines are not explained and the high cost of submitting the testing webpages to dozens of search engines led us not to consider them. Therefore, all fee submission search engines were eliminated and free submission search engines were kept for use in the study. After the elimination process, 19 search engines remained in the final list.

All test webpage http addresses were submitted to each of the 19 search engines. The submission time for each search engine and the search engine address were recorded.

2.3. Search and observation

One week after the webpage URLs were submitted, searching and observation began. The observation interval was set at one week.

2.3.1. Search

In order to get a satisfactory search result, an efficient and effective search strategy was needed. Because this study was not concerned with the search strategy, we developed two strategies that would most efficiently retrieve a relevant set, including our webpages. We could then focus on the study variables, namely the relative ranking in the results lists. Based on the content of the test webpages, two search strategies were used:

[1] "Acupuncture" + "Dimitroff" + "Jin"
[2] "Homeopathy" + "Dimitroff" + "Jin"

The first one was used to retrieve the webpages in the duplication and position categories while the second one was used to retrieve the webpages within the layout category.

Pilot study findings suggested that it was necessary to add the qualifiers "Dimitroff" and "Jin" to the two originally single term queries, eliminating irrelevant webpages and increasing the likelihood of retrieval of the test webpages within the first several hundred hits. Dimitroff and Jin were added to each of the derived webpages before they were submitted to search engines. These terms were selected because they are, obviously, very specific and could effectively exclude other webpages from the result set. It is important to point out that since the impact of the added qualifiers on each testing webpage was the same, it did not affect the final analysis of the study.

The intent was to examine the webpages' relative ranking in search results lists, not their absolute position.

Since some search engines are case sensitive to query terms or even query term order, a search strategy for each search engine that included these minor modifications was developed.

2.3.2. Search result observation

We found that there was a time lag from when a webpage URL was submitted to when that webpage was included in the database of a search engine. In other words, a webpage address submitted to a search engine does not necessarily mean that it would be immediately available in the database of the search engine. If a webpage was not included in the database of a search engine, it would be impossible to retrieve it through the search engine. The time lag ranged from about two weeks to several months, varying by search engine.

Observations were made on a weekly basis. Each observation consisted of a query being submitted to each of the search engines. Every item in the results list was checked up to 500 items per observation. If any retrieved item was identified as one of the test webpages, its position in the list and corresponding search engine was recorded. All selected search engines were searched each week.

We continued observations for several weeks after posted webpages appeared in all study search engines. Data were collected for a total of 21 weeks.

3. Data analysis

3.1. Examination of hypotheses

In order to examine and test the proposed hypotheses, three statistical techniques were used: one-way *ANOVA*, two-way *ANOVA*, and independent-sample *T*-test. For the one-way *ANOVA* and the two-way *ANOVA*, the assumptions were that the involved dependent variable was normally distributed, the population variances of the dependent variable were the same for all cells, and the case represents random samples and the values of dependent variable were independent of each other. For the *T*-test, the assumptions were that the variable was normally distributed in the population, the variances of the normally distributed test variable were equal, and the case represents random samples and the values of dependent variable were independent of each other.

The measurement for the study was the position of a retrieved webpage in a search engine. That is, the position of a webpage in a search engine results list was used to measure performance of the testing webpages. The retrieved webpages with a location at the beginning of a results list of a search engine have good visibility. In other words, the higher a position of a retrieved webpage in a search engine results list, the better its performance. A higher position of a retrieved webpage corresponds to lower value of the variable position and a lower position corresponds to a higher value of the variable position.

The significance level (p) or *sig* for tests is 0.05. Regardless of the specific statistic used, if p or the *sig* is smaller than 0.05, the finding is statistically significant and null hypotheses are rejected. Note that in the statistical result tables the software may have presented a *p value of* 0.000. That is because that the system can only produce approximate p values for calculation. In other words, it

is too small to be considered or it rounds down to 0.000. But it is not equal to zero. In fact, the three zeros (0 0 0) after the decimal point indicate the degree of accuracy for the p value. SPSS was used for data analysis.

3.1.1. Examination of hypothesis 1 (H1)

H1 states that there are no differences in terms of search return performance among webpages with different keyword frequencies within a webpage title, different search engines, and their interactions.

The two factors are keyword frequency within a webpage title and search engine. They are all independent variables. The dependent variable is the webpage return position in a search engine. Since two factors (search engine and keyword frequency in a title) were involved in the question, a two-way *ANOVA* method was used.

In Tables 1–4, SE stands for search engine. Within SE, column labels 1, 2, 3, 4, 5, 6, 7 and 8 refer to All Web, EntireWeb, Google, Lycos, AltaVista, Yahoo, Inforspace/Fast, and Netscape, respectively. FREQUENCY refers to keyword occurrence in the title of a webpage. Valid values for FREQUENCY are 1, 2, 3 and 4. POSITION refers to the webpage retrieval position in a search engine results list.

Table 4 shows that the effect of the keyword frequency (FREQUENCY) in a webpage title is statistically significant ($F = 4.208$, $p = 0.006$ (<0.05)). Search engine (SE) is also statistically significant ($F = 14.414$, $p = 0.000$ (<0.05)) but the interaction (SE * FREQUENCY) is not significant ($F = 1.348$, $p = 0.168$ (>0.05)).

Because overall F tests for keyword frequency ($p = 0.006$) and search engines ($p = 0.000$) are significant, follow up tests (*Tukey* method) were conducted to evaluate pairwise differences among the means. Tables 5 and 6 illustrate the detailed results for the two follow up tests. The abbreviations in these two tables are the same as those in the prior tables. Assuming that the higher a returned webpage position, the better its performance, good performance was achieved by the following search engines: Google (three negative significant mean differences), AltaVista (three negative significant mean differences), Yahoo (four negative significant mean differences and three negative mean differences), Inforspace/Fast (three negative significant mean differences and two negative mean differences), and Netscape (three negative significant mean differences and one negative mean difference). This means that the mean difference ($I - J$) is significant and negative, thus these search engines achieved good performance (see Table 5). Among them, Yahoo achieved the best performance (four negative significant mean differences and three negative mean differences) even though the difference between it and AltaVista, Inforspace/Fast, or Netscape is not statistically significant (but still stays negative).

Following the same principle, we can conclude that the webpages (one negative significant mean difference and two negative mean differences) where keyword frequency is 3 in a title

Table 1
Between-subject factors for H1

FREQUENCY				SE							
1	2	3	4	1	2	3	4	5	6	7	8
54	42	96	92	40	40	83	32	37	14	17	21

Table 2
Descriptive statistics for H1

SE	FREQUENCY	Mean	Std. deviation	*N*
1	1	6.6667	2.06559	6
	2	10.0000	–	1
	3	6.0000	2.47656	16
	4	6.1765	3.14713	17
	Total	6.2750	2.71735	40
2	1	6.6667	2.06559	6
	3	5.8824	4.29945	17
	4	7.2353	1.39326	17
	Total	6.5750	3.05411	40
3	1	4.0952	1.57812	21
	2	4.6190	1.59613	21
	3	3.0500	0.82558	20
	4	4.8095	0.74960	21
	Total	4.1566	1.40974	83
4	1	6.6667	2.06559	6
	3	5.6154	2.93083	13
	4	7.1538	1.62512	13
	Total	6.4375	2.35465	32
5	2	5.2727	0.64667	11
	3	2.6923	0.48038	13
	4	2.7692	1.01274	13
	Total	3.4865	1.38688	37
6	1	2.0000	0.00000	5
	2	2.6667	1.15470	3
	3	2.6667	0.51640	6
	Total	2.4286	0.64621	14
7	1	3.2000	1.09545	5
	3	3.6667	0.81650	6
	4	5.1667	0.40825	6
	Total	4.0588	1.14404	17
8	1	4.6000	1.94936	5
	2	3.3333	2.06559	6
	3	4.0000	0.70711	5
	4	4.8000	0.44721	5
	Total	4.1429	1.52597	21
Total	1	4.7222	2.20990	54
	2	4.5952	1.80864	42
	3	4.4062	2.72832	96
	4	5.5761	2.22490	92
	Total	4.8732	2.39363	284

Table 3
Levene's test of equality of error variances (a) for H1

F	df1	df2	Sig.
6.697	26	257	0.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
(a) Design: Intercept + SE + FREQUENCY + SE * FREQUENCY.

Table 4
Tests of between-subjects effects for H1

Source	Type III sum of squares	Df	Mean square	*F*	Sig.	Partial Eta squared
Corrected model	648.865[a]	26	24.956	6.595	0.000	0.400
Intercept	3468.287	1	3468.287	916.487	0.000	0.781
SE	381.820	7	54.546	14.414	0.000	0.282
FREQUENCY	47.769	3	15.923	4.208	0.006	0.047
SE * FREQUENCY	81.645	16	5.103	1.348	0.168	0.077
Error	972.572	257	3.784			
Total	8366.000	284				
Corrected total	1621.437	283				

[a] *R* squared = 0.400 (adjusted *R* squared = 0.339).

achieved the best performance, the webpages (one negative significant mean difference and one negative mean difference) where keyword frequency is 2, and last, webpages (one negative mean difference) where keyword frequency is 1 in Table 6. It is interesting that the webpages in which keyword frequency is 4 achieved the worst performance in this category. This suggests that as the number of keyword frequency increases the performance improves up to a frequency of 3. When the frequency is over 3 and the performance decreases dramatically. In other words, duplicating keywords in a title more than three times does not improve its visibility in a search engine results list.

Fig. 1 displays the profile plot of the cell means which may be useful in visualizing the differential effects.

3.1.2. Examination of hypothesis 2 (H2)

H2 states that there are no differences in terms of search return performance among webpages with different keyword frequencies within a webpage's full-text, different search engines, and their interactions.

The two factors and independent variables are keyword frequencies within a webpage's full-text and search engines. The dependent variable is the webpage return position in a search engine. Since two factors were involved in the hypothesis and the interaction between the two factors needed to be investigated, a two-way *ANOVA* method was used.

Tables 7–10 give detailed statistical data for H2. In these tables, definitions of POSITION, SE and the SE column labels are the same as described for H1, above. FREQUENCY refers to the frequency of keyword occurrence in the full-text of a webpage. It ranges from one occurrence to five occurrences.

Table 5
Multiple comparisons of search engines for H1

	(*I*) SE	(*J*) SE	Mean difference ($I-J$)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
Tukey HSD	1	2	−0.3000	0.43499	0.997	−1.6294	1.0294
		3	2.1184*	0.37444	0.000	0.9741	3.2627
		4	−0.1625	0.46138	1.000	−1.5725	1.2475
		5	2.7885*	0.44372	0.000	1.4325	4.1446
		6	3.8464*	0.60408	0.000	2.0003	5.6926
		7	2.2162*	0.56322	0.003	0.4949	3.9374
		8	2.1321*	0.52423	0.002	0.5301	3.7342
	2	1	0.3000	0.43499	0.997	−1.0294	1.6294
		3	2.4184*	0.37444	0.000	1.2741	3.5627
		4	0.1375	0.46138	1.000	−1.2725	1.5475
		5	3.0885*	0.44372	0.000	1.7325	4.4446
		6	4.1464*	0.60408	0.000	2.3003	5.9926
		7	2.5162*	0.56322	0.000	0.7949	4.2374
		8	2.4321*	0.52423	0.000	0.8301	4.0342
	3	1	−2.1184*	0.37444	0.000	−3.2627	−0.9741
		2	−2.4184*	0.37444	0.000	−3.5627	−1.2741
		4	−2.2809*	0.40479	0.000	−3.5179	−1.0438
		5	0.6701	0.38454	0.659	−0.5051	1.8453
		6	1.7281*	0.56205	0.047	0.0104	3.4457
		7	0.0978	0.51788	1.000	−1.4849	1.6805
		8	0.0138	0.47518	1.000	−1.4384	1.4660
	4	1	0.1625	0.46138	1.000	−1.2475	1.5725
		2	−0.1375	0.46138	1.000	−1.5475	1.2725
		3	2.2809*	0.40479	0.000	1.0438	3.5179
		5	2.9510*	0.46962	0.000	1.5158	4.3862
		6	4.0089*	0.62335	0.000	2.1039	5.9140
		7	2.3787*	0.58384	0.002	0.5944	4.1629
		8	2.2946*	0.54632	0.001	0.6250	3.9643
	5	1	−2.7885*	0.44372	0.000	−4.1446	−1.4325
		2	−3.0885*	0.44372	0.000	−4.4446	−1.7325
		3	−0.6701	0.38454	0.659	−1.8453	0.5051
		4	−2.9510*	0.46962	0.000	−4.3862	−1.5158
		6	1.0579	0.61040	0.666	−0.8075	2.9234
		7	−0.5723	0.56999	0.974	−2.3143	1.1696
		8	−0.6564	0.53149	0.921	−2.2807	0.9679
	6	1	−3.8464*	0.60408	0.000	−5.6926	−2.0003
		2	−4.1464*	0.60408	0.000	−5.9926	−2.3003
		3	−1.7281*	0.56205	0.047	−3.4457	−0.0104
		4	−4.0089*	0.62335	0.000	−5.9140	−2.1039
		5	−1.0579	0.61040	0.666	−2.9234	0.8075
		7	−1.6303	0.70208	0.286	−3.7759	0.5154
		8	−1.7143	0.67120	0.178	−3.7656	0.3370

Table 5 (*continued*)

	(*I*) SE	(*J*) SE	Mean difference ($I-J$)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
	7	1	−2.2162*	0.56322	0.003	−3.9374	−0.4949
		2	−2.5162*	0.56322	0.000	−4.2374	−0.7949
		3	−0.0978	0.51788	1.000	−1.6805	1.4849
		4	−2.3787*	0.58384	0.002	−4.1629	−0.5944
		5	0.5723	0.56999	0.974	−1.1696	2.3143
		6	1.6303	0.70208	0.286	−0.5154	3.7759
		8	−0.0840	0.63468	1.000	−2.0237	1.8556
	8	1	−2.1321*	0.52423	0.002	−3.7342	−0.5301
		2	−2.4321*	0.52423	0.000	−4.0342	−0.8301
		3	−0.0138	0.47518	1.000	−1.4660	1.4384
		4	−2.2946*	0.54632	0.001	−3.9643	−0.6250
		5	0.6564	0.53149	0.921	−0.9679	2.2807
		6	1.7143	0.67120	0.178	−0.3370	3.7656
		7	0.0840	0.63468	1.000	−1.8556	2.0237

Based on observed means.
*The mean difference is significant at the 0.05 level.

Table 6
Multiple comparisons of FREQUENCY for H1

	(*I*) FREQUENCY	(*J*) FREQUENCY	Mean difference ($I-J$)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
Tukey HSD	1	2	0.1270	0.40023	0.989	−0.9080	1.1620
		3	0.3160	0.33091	0.775	−0.5397	1.1717
		4	−0.8539	0.33349	0.053	−1.7162	0.0085
	2	1	−0.1270	0.40023	0.989	−1.1620	0.9080
		3	0.1890	0.35989	0.953	−0.7417	1.1197
		4	−0.9808*	0.36227	0.036	−1.9176	−0.0440
	3	1	−0.3160	0.33091	0.775	−1.1717	0.5397
		2	−0.1890	0.35989	0.953	−1.1197	0.7417
		4	−1.1698*	0.28382	0.000	−1.9038	−0.4359
	4	1	0.8539	0.33349	0.053	−0.0085	1.7162
		2	0.9808*	0.36227	0.036	0.0440	1.9176
		3	1.1698*	0.28382	0.000	0.4359	1.9038

Based on observed means.
*The mean difference is significant at the 0.05 level.

Table 10 illustrates that the main effect of keyword frequency (FREQUENCY) in the full-text of a webpage is significantly different ($F = 61.184$, $p = 0.000$ (<0.05)). Another effect (SE), search

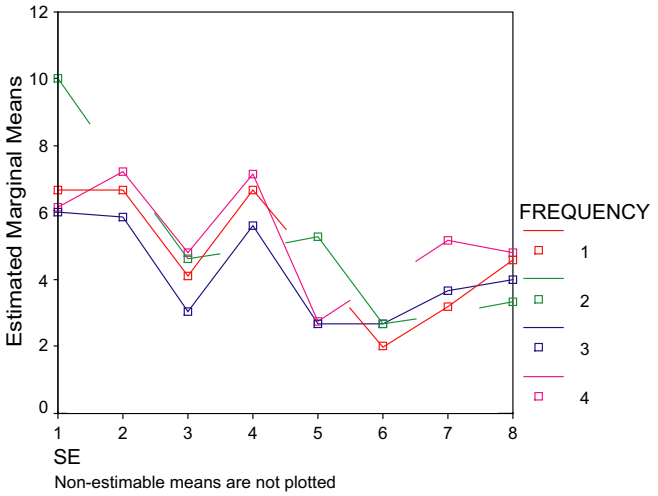

Fig. 1. Estimated marginal means of POSITION for H1.

Table 7
Between-subject factors for H2

FREQUENCY					SE							
1	2	3	4	5	1	2	3	4	5	6	7	8
30	76	79	91	97	78	67	88	58	15	17	29	21

engine, is also statistically significant ($F = 30.516$, $p = 0.000$ (<0.05)). In addition, the interaction of the two factors (SE * FREQUENCY) is also statistically significant ($F = 5.856$, $p = 0.000$ (<0.05)).

Because overall F tests for both keyword frequency ($p = 0.000$) and search engine ($p = 0.000$) are significant, the *Tukey* method was used to investigate pairwise differences among the means. Tables 11 and 12 show the detailed analysis results for the two follow-up tests. The same abbreviations are used in these two tables and the same analysis method is applied to this hypothesis. It is clear that the search engines Yahoo (seven negative significant mean differences), AltaVista (five negative significant mean differences and one negative mean difference), and Inforspace/Fast (five negative significant mean differences) achieved good performance (see Table 11). Once again, Yahoo achieved the best performance among all search engines.

After examining Table 12, it is apparent that the webpages (four negative significant mean differences) where keyword frequency is 5 achieved the best performance, the webpages (three negative significant mean differences) where keyword frequency is 4, then the webpages (two negative significant mean differences) where keyword frequency is 3. Finally the webpages where keyword frequencies are 1 and 2 achieved the worst performance. In this case it is clear that when the number of keywords in a full-text increases, webpage performance gets better. Unlike the performance of keywords in titles, there is no restriction on the number of keywords in the full-text in terms of the visibility improvement.

Table 8
Descriptive statistics for H2

SE	FREQUENCY	Mean	Std. deviation	*N*
1	1	12.5714	0.53452	7
	2	12.8125	1.47054	16
	3	9.6316	3.21819	19
	4	10.3889	1.41998	18
	5	9.2778	1.31978	18
	Total	10.6410	2.39032	78
2	1	12.6667	0.51640	6
	2	14.5000	2.40992	14
	3	12.6154	2.53438	13
	4	11.9412	2.60937	17
	5	10.8235	2.42990	17
	Total	12.3881	2.65692	67
3	1	16.0000	0.00000	4
	2	14.9524	0.21822	21
	3	11.4762	1.93956	21
	4	7.9524	0.21822	21
	5	6.9524	0.21822	21
	Total	10.5909	3.44959	88
4	1	12.5714	0.53452	7
	2	12.6364	2.50091	11
	3	9.7143	4.06540	14
	4	10.3077	2.01596	13
	5	9.3077	2.01596	13
	Total	10.6552	2.91127	58
5	4	7.0000	0.77460	11
	5	6.7500	0.50000	4
	Total	6.9333	0.70373	15
6	2	4.3333	0.57735	3
	3	5.0000	–	1
	5	3.4615	0.66023	13
	Total	3.7059	0.77174	17
7	1	9.4000	0.54772	5
	2	9.6667	0.51640	6
	3	8.1667	0.40825	6
	4	7.1667	0.40825	6
	5	6.1667	0.40825	6
	Total	8.0690	1.41247	29
8	1	16.0000	–	1
	2	15.0000	0.00000	5
	3	12.6000	1.51658	5

(*continued on next page*)

Table 8 (*continued*)

SE	FREQUENCY	Mean	Std. deviation	*N*
	4	8.0000	0.00000	5
	5	7.0000	0.00000	5
	Total	10.9048	3.54831	21
Total	1	12.6333	1.99107	30
	2	13.2500	2.81484	76
	3	10.6456	3.09689	79
	4	9.3516	2.32079	91
	5	7.8557	2.67702	97
	Total	10.2949	3.33020	373

Table 9
Levene's test of equality of error variances (a) for H2

F	df1	df2	Sig.
7.898	34	338	0.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
(a) Design: Intercept + SE + FREQUENCY + SE * FREQUENCY.

Table 10
Tests of between-subjects effects for H2

Source	Type III sum of squares	Df	Mean square	*F*	Sig.	Partial Eta squared
Corrected model	3002.145[a]	34	88.298	26.566	0.000	0.728
Intercept	14917.692	1	14917.692	4488.261	0.000	0.930
SE	709.993	7	101.428	30.516	0.000	0.387
FREQUENCY	813.438	4	203.359	61.184	0.000	0.420
SE * FREQUENCY	447.687	23	19.465	5.856	0.000	0.285
Error	1123.415	338	3.324			
Total	43658.000	373				
Corrected total	4125.560	372				

[a] R squared = 0.728 (adjusted R squared = 0.700).

Compare the two average standard deviations in Tables 8 and 2, the average standard deviation of this test (3.33020) is larger than that of the previous test (2.39363). This suggests that the impact of keyword frequency in a full-text on visibility is much stronger than that of keyword frequency in a title.

A visual display of hypothesis 2 results is presented in Fig. 2.

3.1.3. Examination of hypothesis 3 (H3)

H3 states that there is no difference with respect to search return performance between webpages with keywords only in the title, webpages with keywords only in full-text, and those with keywords in both title and full-text.

Table 11
Multiple comparisons for SE for H2

	(*I*) SE	(*J*) SE	Mean difference (*I* − *J*)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
Tukey HSD	1	2	−1.7470*	0.30368	0.000	−2.6733	−0.8208
		3	0.0501	0.28352	1.000	−0.8146	0.9149
		4	−0.0141	0.31610	1.000	−0.9783	0.9500
		5	3.7077*	0.51400	0.000	2.1400	5.2754
		6	6.9351*	0.48798	0.000	5.4468	8.4235
		7	2.5721*	0.39651	0.000	1.3627	3.7814
		8	−0.2637	0.44820	0.999	−1.6308	1.1033
	2	1	1.7470*	0.30368	0.000	0.8208	2.6733
		3	1.7972*	0.29560	0.000	0.8956	2.6987
		4	1.7329*	0.32698	0.000	0.7356	2.7302
		5	5.4547*	0.52076	0.000	3.8664	7.0431
		6	8.6822*	0.49510	0.000	7.1721	10.1922
		7	4.3191*	0.40524	0.000	3.0831	5.5551
		8	1.4833*	0.45594	0.027	0.0927	2.8739
	3	1	−0.0501	0.28352	1.000	−0.9149	0.8146
		2	−1.7972*	0.29560	0.000	−2.6987	−0.8956
		4	−0.0643	0.30834	1.000	−1.0047	0.8762
		5	3.6576*	0.50926	0.000	2.1043	5.2109
		6	6.8850*	0.48299	0.000	5.4119	8.3582
		7	2.5219*	0.39036	0.000	1.3313	3.7126
		8	−0.3139	0.44277	0.997	−1.6643	1.0366
	4	1	0.0141	0.31610	1.000	−0.9500	0.9783
		2	−1.7329*	0.32698	0.000	−2.7302	−0.7356
		3	0.0643	0.30834	1.000	−0.8762	1.0047
		5	3.7218*	0.52810	0.000	2.1111	5.3326
		6	6.9493*	0.50281	0.000	5.4157	8.4829
		7	2.5862*	0.41463	0.000	1.3216	3.8508
		8	−0.2496	0.46430	0.999	−1.6657	1.1666
	5	1	−3.7077*	0.51400	0.000	−5.2754	−2.1400
		2	−5.4547*	0.52076	0.000	−7.0431	−3.8664
		3	−3.6576*	0.50926	0.000	−5.2109	−2.1043
		4	−3.7218*	0.52810	0.000	−5.3326	−2.1111
		6	3.2275*	0.64583	0.000	1.2577	5.1973
		7	−1.1356	0.57982	0.512	−2.9041	0.6328
		8	−3.9714*	0.61632	0.000	−5.8512	−2.0916
	6	1	−6.9351*	0.48798	0.000	−8.4235	−5.4468
		2	−8.6822*	0.49510	0.000	−10.1922	−7.1721
		3	−6.8850*	0.48299	0.000	−8.3582	−5.4119
		4	−6.9493*	0.50281	0.000	−8.4829	−5.4157
		5	−3.2275*	0.64583	0.000	−5.1973	−1.2577
		7	−4.3631*	0.55689	0.000	−6.0616	−2.6646
		8	−7.1989*	0.59480	0.000	−9.0130	−5.3847

(*continued on next page*)

Table 11 (*continued*)

(*I*) SE	(*J*) SE	Mean difference ($I-J$)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
7	1	−2.5721*	0.39651	0.000	−3.7814	−1.3627
	2	−4.3191*	0.40524	0.000	−5.5551	−3.0831
	3	−2.5219*	0.39036	0.000	−3.7126	−1.3313
	4	−2.5862*	0.41463	0.000	−3.8508	−1.3216
	5	1.1356	0.57982	0.512	−0.6328	2.9041
	6	4.3631*	0.55689	0.000	2.6646	6.0616
	8	−2.8358*	0.52238	0.000	−4.4291	−1.2425
8	1	0.2637	0.44820	0.999	−1.1033	1.6308
	2	−1.4833*	0.45594	0.027	−2.8739	−0.0927
	3	0.3139	0.44277	0.997	−1.0366	1.6643
	4	0.2496	0.46430	0.999	−1.1666	1.6657
	5	3.9714*	0.61632	0.000	2.0916	5.8512
	6	7.1989*	0.59480	0.000	5.3847	9.0130
	7	2.8358*	0.52238	0.000	1.2425	4.4291

Multiple comparisons—dependent variable: POSITION.
Based on observed means.
*The mean difference is significant at the 0.05 level.

The independent variables were the webpages with keywords only in a title, the webpages with keywords only in a full-text, and those with keywords in both the title and full-text. The dependent variable was the webpage return position in a search engine results list. A one-way *ANOVA* was used for this test because of involvement of multiple independent variables.

Tables 13 and 14 display the generated data for H3. Numbers 1, 2, and 3 in the tables represent webpages with keywords only in title, webpages with keywords only in full-text, and those with keywords in both title and full-text, respectively. Since the p-value is 0.000 (<0.05) ($F = 445.688$), this hypothesis was rejected (see Table 14).

Due to the rejection of H3, post-hoc multiple comparisons (*Tukey* honestly significant differences (HSD)) were conducted to evaluate pairwise differences among the means. From data displayed in Table 15, we found that the mean differences ($I-J$) for group 3 (webpages with keywords in both title and full-text) are −3.735 and −9.1570 against group 2 and group 1 respectively, and the mean difference ($I-J$) for group 2 against group 1 is −5.412. The differences are negative and significant. This indicates that the webpages with keywords in both title and full-text achieved the best performance across the groups, and the webpages with keywords only in full-texts achieved better performance than the webpages with keywords only in titles.

As shown in Table 13, the standard deviation of group 3 is very small (0.51122), its corresponding lower bound (1.0035) and upper bound (1.2723) are very close. This suggests that performance of all search engines is quite consistent when keywords appear in both a title and a full-text of a webpage.

Table 16 presents the data in a different way by showing sets of means that do not differ significantly from each other. In this case, no more than one group forms a homogeneous subset.

Table 12
Multiple comparisons for FREQUENCY for H2

	(*I*) FREQUENCY	(*J*) FREQUENCY	Mean difference (*I* − *J*)	Std. error	Sig.	95% Confidence interval	
						Lower bound	Upper bound
Tukey HSD	1	2	−0.6167	0.39309	0.519	−1.6948	0.4614
		3	1.9878*	0.39098	0.000	0.9155	3.0600
		4	3.2817*	0.38382	0.000	2.2290	4.3343
		5	4.7777*	0.38086	0.000	3.7331	5.8222
	2	1	0.6167	0.39309	0.519	−0.4614	1.6948
		3	2.6044*	0.29293	0.000	1.8011	3.4078
		4	3.8984*	0.28330	0.000	3.1214	4.6753
		5	5.3943*	0.27928	0.000	4.6284	6.1603
	3	1	−1.9878*	0.39098	0.000	−3.0600	−0.9155
		2	−2.6044*	0.29293	0.000	−3.4078	−1.8011
		4	1.2939*	0.28035	0.000	0.5250	2.0628
		5	2.7899*	0.27629	0.000	2.0321	3.5477
	4	1	−3.2817*	0.38382	0.000	−4.3343	−2.2290
		2	−3.8984*	0.28330	0.000	−4.6753	−3.1214
		3	−1.2939*	0.28035	0.000	−2.0628	−0.5250
		5	1.4960*	0.26606	0.000	0.7663	2.2257
	5	1	−4.7777*	0.38086	0.000	−5.8222	−3.7331
		2	−5.3943*	0.27928	0.000	−6.1603	−4.6284
		3	−2.7899*	0.27629	0.000	−3.5477	−2.0321
		4	−1.4960*	0.26606	0.000	−2.2257	−0.7663

Based on observed means.
*The mean difference is significant at the 0.05 level.

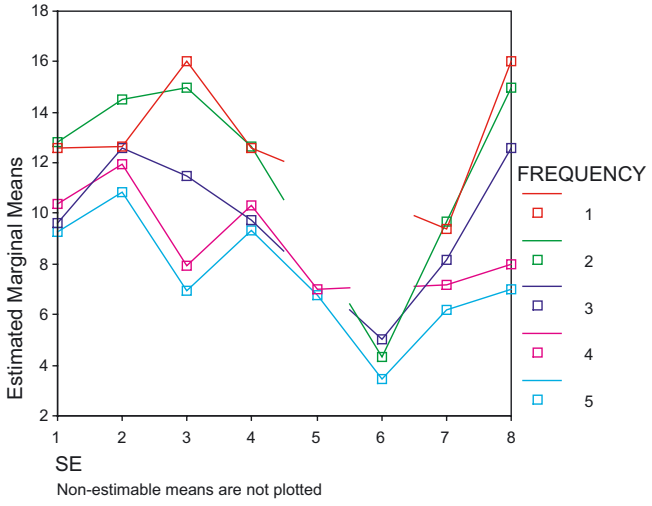

Fig. 2. Estimated Marginal Means of POSITION for H2.

Table 13
Descriptive for H3

	N	Mean	Std. deviation	Std. error	95% Confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
1	373	10.2949	3.33020	0.17243	9.9558	10.6340	3.00	18.00
2	284	4.8732	2.39363	0.14204	4.5937	5.1528	1.00	14.00
3	58	1.1379	0.51122	0.06713	1.0035	1.2723	1.00	3.00
Total	715	7.3986	4.26296	0.15943	7.0856	7.7116	1.00	18.00

Table 14
ANOVA for H3

	Sum of squares	Df	Mean square	F	Sig.
Between groups	7213.505	2	3606.753	445.688	0.000
Within groups	5761.893	712	8.093		
Total	12975.399	714			

Table 15
Tukey HSD multiple comparisons for H3

(I) TYPE	(J) TYPE	Mean difference (I − J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
1	2	5.4217*	0.22403	0.000	4.8955	5.9478
	3	9.1570*	0.40153	0.000	8.2139	10.1000
2	1	−5.4217*	0.22403	0.000	−5.9478	−4.8955
	3	3.7353*	0.40990	0.000	2.7726	4.6980
3	1	−9.1570*	0.40153	0.000	−10.1000	−8.2139
	2	−3.7353*	0.40990	0.000	−4.6980	−2.7726

*The mean difference is significant at the 0.05 level.

Table 16
Tukey HSD Homogeneous subsets for H3

TYPE	N	Subset for alpha = 0.05		
		1	2	3
3	58	1.1379		
2	284		4.8732	
1	373			10.2949
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.
(a) Uses mean sample size = 127.967.
(b) The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

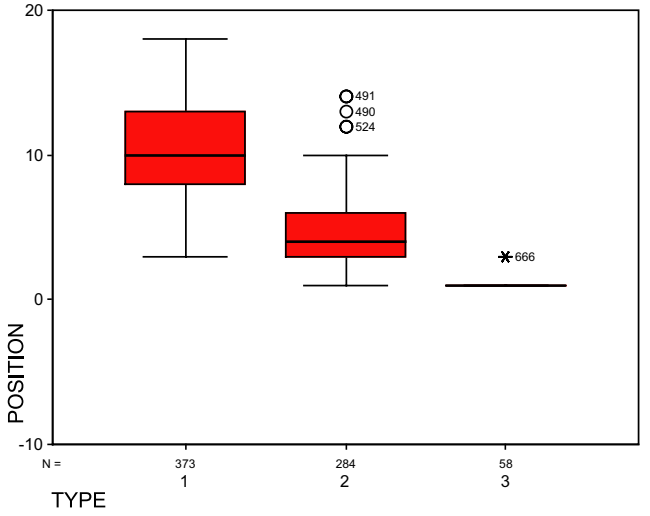

Fig. 3. Profile plots for H3.

The *ANOVA* test results are depicted using boxplots to show the distribution of the dependent variable across the groups (see Fig. 3).

3.1.4. Examination of hypothesis 4 (H4)

H4 states that there is no difference with respect to search return performance among webpages with different keyword font color, font size, keyword plural form, keyword case status, or keyword adjectival form.

The independent variables were webpages with different keyword font colors, font sizes, keyword plural forms, keyword case status, and keyword adjectival forms. The dependent variable was the webpage return position in a search engine result. Due to the multiple independent variables, a one-way *ANOVA* was used for this hypothesis.

In order to effectively compare impact of font color, font size, case status, etc. on their visibility in a search result, we also posted the original webpage with no font color change, no font size change, no case change, or no plural or adjective changes.

Tables 17 and 18 show the detailed results of the *ANOVA* test analysis. The *TYPE* column values 1, 2, 3, 4, 5 and 6 represent the original webpage, a test webpage with different keyword font color, font size, plural form, case status, and adjectival form, respectively. Due to the significant overall F in Table 18 ($F = 5.346$, $p = 0.000$ (<0.05)), the hypothesis was rejected. The results of a follow-up test show that the mean differences or the original webpage against the webpages with different keyword font color, font size, plural form, case status, and adjectival form are −2.5366, 1.3158, −0.8923, 0.0000, and 2.0000, respectively (see Table 19). None is significant. This suggests that there is no significant difference between the original webpage and webpages with different font color, between the original webpage and webpages with different font cases, between the original webpage and webpages with different font size, between the original webpage and webpages with different plural forms, or between the original webpage and webpages with different adjectival forms in terms of their visibility performance. Although the hypothesis was

Table 17
Descriptive for H4

Type	*N*	Mean	Standard deviation	Std. error	95% Confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
1.00	5	12.0000	1.73205	0.77460	9.8494	14.1506	9.00	13.00
2.00	41	14.5366	1.81793	0.28391	13.9628	15.1104	8.00	19.00
3.00	38	10.6842	4.64470	0.75347	9.1575	12.2109	4.00	15.00
4.00	65	12.8923	3.32191	0.41203	12.0692	13.7154	4.00	17.00
5.00	3	12.0000	7.00000	4.04145	−5.3890	29.3890	7.00	20.00
6.00	4	10.0000	5.59762	2.79881	1.0929	18.9071	5.00	17.00
Total	156	12.6667	3.74051	0.29948	12.0751	13.2583	4.00	20.00

Table 18
ANOVA for H4

	Sum of squares	Df	Mean square	*F*	Sig.
Between groups	328.015	5	65.603	5.346	0.000
Within groups	1840.652	150	12.271		
Total	2168.667	155			

rejected, the reason was not because of performance between the original webpage and the other modified webpages but because of the performance differences among webpages with different font color, webpages with different font cases, webpages with different font size, webpages with different plural forms, and webpages with different adjectival forms that result in the rejection. If a *T*-test had been conducted between the original webpages and each of other modified webpages, the same conclusions would have been drawn based on the same collected data. In fact, Table 20 also confirms the above conclusion because all six involved groups are within a homogenous subset.

Fig. 4 is the visual display of the distribution of dependent variable across the groups.

4. Conclusion

This research aimed to (1) identify which webpage factors effect webpage placement in search engine results lists, (2) analyze the effects of these factors on major search engines on the internet, and (3) recommend practical methods for improving webpage visibility search engine results lists in based on the findings.

Toward these aims, test webpages were derived and modified from a selected original webpage and were posted on the internet. The addresses of these derived webpages were submitted to 19 search engines so that the posted webpages could be indexed in their databases. One week later the investigators began searching the 19 search engines weekly. The returned results from each of the search engines were monitored and recorded. After 21 weeks of observation, eight search engines responded to the submissions positively. All collected data were tabulated and classified. Three

Table 19
Tukey HSD multiple comparisons for H4

(*I*) TYPE	(*J*) TYPE	Mean difference (*I* − *J*)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
1.00	2.00	−2.5366	1.65937	0.646	−7.3271	2.2539
	3.00	1.3158	1.66647	0.969	−3.4952	6.1268
	4.00	−0.8923	1.62573	0.994	−5.5857	3.8011
	5.00	0.0000	2.55823	1.000	−7.3854	7.3854
	6.00	2.0000	2.34988	0.957	−4.7840	8.7840
2.00	1.00	2.5366	1.65937	0.646	−2.2539	7.3271
	3.00	3.8524*	0.78881	0.000	1.5751	6.1296
	4.00	1.6443	0.69863	0.180	−0.3726	3.6612
	5.00	2.5366	2.09514	0.831	−3.5120	8.5851
	6.00	4.5366	1.83495	0.139	−0.7608	9.8340
3.00	1.00	−1.3158	1.66647	0.969	−6.1268	3.4952
	2.00	−3.8524*	0.78881	0.000	−6.1296	−1.5751
	4.00	−2.2081*	0.71534	0.029	−4.2732	−0.1430
	5.00	−1.3158	2.10078	0.989	−7.3806	4.7490
	6.00	0.6842	1.84138	0.999	−4.6317	6.0001
4.00	1.00	0.8923	1.62573	0.994	−3.8011	5.5857
	2.00	−1.6443	0.69863	0.180	−3.6612	0.3726
	3.00	2.2081*	0.71534	0.029	0.1430	4.2732
	5.00	0.8923	2.06860	0.998	−5.0796	6.8642
	6.00	2.8923	1.80459	0.598	−2.3174	8.1020
5.00	1.00	0.0000	2.55823	1.000	−7.3854	7.3854
	2.00	−2.5366	2.09514	0.831	−8.5851	3.5120
	3.00	1.3158	2.10078	0.989	−4.7490	7.3806
	4.00	−0.8923	2.06860	0.998	−6.8642	5.0796
	6.00	2.0000	2.67546	0.976	−5.7239	9.7239
6.00	1.00	−2.0000	2.34988	0.957	−8.7840	4.7840
	2.00	−4.5366	1.83495	0.139	−9.8340	0.7608
	3.00	−0.6842	1.84138	0.999	−6.0001	4.6317
	4.00	−2.8923	1.80459	0.598	−8.1020	2.3174
	5.00	−2.0000	2.67546	0.976	−9.7239	5.7239

*The mean difference is significant at the 0.05 level.

different statistical techniques were employed to examine the four proposed hypotheses. Although all proposed hypotheses were rejected, the findings are nevertheless very positive and suggest several options to optimize webpage visibility in a search engine. Based on the statistical analysis presented in the previous section, some highlighted findings and suggestions are summarized below:

(1) When the number of duplicated keywords in a webpage title increases, its visibility in a search engine results list increases up to three duplications. When the duplications exceed three, there is a downturn in terms of visibility performance in a search engine results list. Therefore, a point of diminishing returns has been identified at four duplicated keywords.

Table 20
Tukey HSD homogeneous subsets for H4

TYPE	*N*	Subset for alpha = 0.05 1
6.00	4	10.0000
3.00	38	10.6842
1.00	5	12.0000
5.00	3	12.0000
4.00	65	12.8923
2.00	41	14.5366
Sig.		0.151

Means for groups in homogeneous subsets are displayed.
(a) Uses harmonic mean sample size = 7.064.
(b) The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

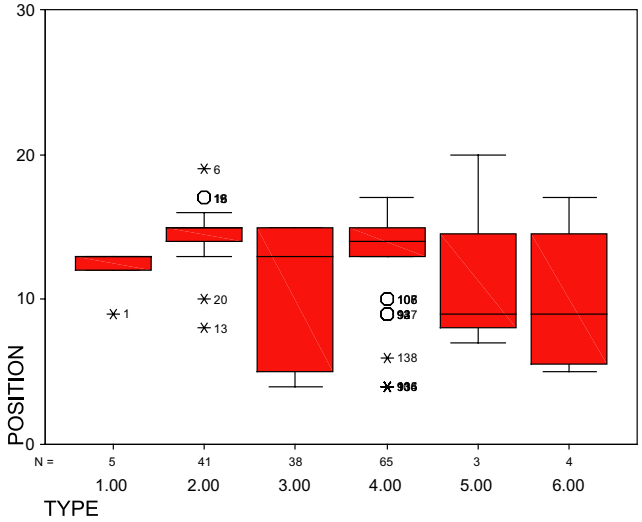

Fig. 4. Visual display of results for H4.

(2) As the number of duplicated keywords in the full-text of a webpage increases, the visibility in the results list of a search engine increases. No diminishing returns were found with full-text keywords.
(3) Webpages with keywords in both title and full-text achieved better visibility performance than the webpages with keywords only in full-texts and the webpages with keywords only in titles in light of returned position in a search engine results list. Webpages with keywords only in full-texts achieved better performance than webpages with keywords only in titles.
(4) There is no significant difference between the original webpage and webpages with font color changes, font case changes, font size changes, plural form changes, or adjectival changes in terms of their visibility performance. Search engines are apparently blind to design features

that, while not important in terms of retrieval, are important in terms of positive affective response to webpage design.

The findings of this research can benefit web publishers, search engine designers and web information seekers through application of the simple principles noted above.

Future research directions might include, but are not limited to, an investigation of the factors beyond a webpage that affect webpage visibility in search engine results. In the study described here, the focus was on investigating the factors with a webpage and their impact on visibility. We found that factors beyond a website such as the profile status of the host website where the webpage is posted, whether the host website is linked by other websites, depth of a directory where a webpage is posted, and so on, may also play a role in its visibility. An investigation of the impact of keyword frequency in metadata fields on website visibility is also an area for future examination. The impact of factors beyond a website on webpage visibility, especially webpage hyperlink cited status, and their combination with the factors within a webpage, will be a future study direction.

References

Ambergreen (2002). M2 Presswire.

Centaur Communication (2002). E-volve: Helping the needle out of the Haystack. Marketing Week, p. 47.

Domain Name Express (2003). Available: http://www.submitexpress.com/ [1 February].

Dynamic Web Ranking (2003). Available: http://www.hot-new.com/webrank.htm#dyn [28 January].

Greenberg, K. (2000). Search patterns. *Brandweek, 41*(35), 72–73.

Haltley, R. (2002). *Making sure your Website is top*. UK Newsquest Regional Press.

High Rankings Advisor [online] (2003). Available: http://www.highrankings.com/advisor.htm [27 January].

How search engines rank Webpages (2003). Available: http://searchenginewatch.com/webmasters/rank.html [28 January].

Ihelpyou (2002). Available: http://www.freemoneyservices.com/.

Kanaley, R. (2002). Firms use search engine optimization for high ranking in search results. Business and Financial News.

Leighton, H. V. (2003). *Precision among World Wide Web search services (search engines): AltaVista, Excite, HotBot, Infoseek, Lycos* [online]. Available: http://www.winona.msus.edu/library/webind2/webind2.htm [28 January].

Oppenheim, C., Morris, A., McKnight, C., & Lowley, S. (2000). The evaluation of WWW search engines. *Journal of Documentation, 56*(2), 1990–2111.

Schwartz, C. (1998). Web search engines. *Journal of the American Society for Information Science, 49*(11), 973–982.

Search Engine Optimization—A 10 Step Program, 2003. Available: http://www.netmechanic.com/news/vol5/promo_no8.htm [28 January].

Search engine optimization [online] (2003). Available: http://www.bruceclay.com/web_rank.htm [27 January].

Search engine optimization, search engine ranking, website ranking, website optimization [online] (2003). Available: http://usasearchengineoptimization.info/ [27 January].

Search Engine Optimization Tips [online] (2003). Available: http://www.submit-it.com/subopt.htm [27 January].

Search Engine Optimization Free [online] (2003). Available: http://hotwired.lycos.com/webmonkey/01/23/index1a.html [27 January].

Search engine optimizer [online] (2003). Available: http://www.se-optimizer.com/ [1 February].

Search Engine Submission & Search Engine Optimization [online] (2003). Available: http://www.topseo.com/ [27 January].

Search Engine Optimization Support Forums. Available: http://www.supportforums.org/ [27 January].

Submit today [online] (2003). Available: http://www.submittoday.com/ [27 January].
Sullivan, D. (2003). Intro to Search Engine Optimization [online]. Available: http://searchenginewatch.com/webmasters/intro.html [27 January].
Terra Lycos Network [online] (2003). Available: http://hotwired.lycos.com/webmonkey/01/23/index1a.html [27 January].
Web Position Gold [online] (2003). Available: http://www.submitcorner.com/Software/Webposition/ [27 January].
Website optimization tools [online] (2003). Available: http://www.pandia.com/optimization/index.html#tools [27 January].