

THE ROLE KEYWORD LOCATION PLAYS IN WEBSITE VISIBILITY TO SEARCH ENGINES: AN EMPIRICAL STUDY

Mr WT Kritzinger
Faculty of Business Informatics
Cape Technikon
Cape Town
South Africa
kritzingerw@ctech.ac.za

Dr M Weideman
Faculty of Business Informatics
Cape Technikon
Cape Town
South Africa
meliusw@yahoo.com

Keywords:
Website visibility, Keyword location, Search engines

ABSTRACT

The primary objective of this research project is to report on the location of keywords as a ranking factor of e-Commerce websites.

Many authors claim that the amount of data available on the Internet cannot be measured. New and existing authors constantly add more data by uploading new and revised webpages to web servers, some on an hourly basis. Also, there is no central body responsible for categorising, validating or censoring data on the Internet. It is these factors that contribute to the rather chaotic situation Internet users face when attempting to retrieve relevant information from the Internet.

An e-Commerce website needs to attract visitors and the website designer needs to ensure the website is visible to search engines. Being listed in a search engine index is, however, no guarantee that a user will be able to find the website, even if the website qualifies as a candidate for the user's search. Websites that are not ranked highly by search engines are less likely to be visited by potential customers. Users tend to examine only the first page of search results and once they find a good match for their search, they normally do not look further down the list. Most search engines display only 10 to 20 of the most relevant results on the first page. Thus, exclusion from these top results means that only a small number of search engine users will actually see a link to the website.

It has been found that one of the main rules in ranking algorithms of search engines involve the location and frequency of keywords on a webpage. Some authors refer to it as the “location/frequency method”. Pages with the search terms appearing in the HTML title tag are often assumed to be more relevant than others to the topic. Search engines will also check to see if the search keywords appear near the top of a webpage, such as in the headline or in the first few paragraphs of text. Search engines assume that any page relevant to the topic will mention those words right from the beginning. However, no empirical results were found to confirm that there is a measurable relationship between keyword location and website ranking.

The method employed in this project was to inspect webpages and the location of their keywords. The authors identified four search engines and used the keyword “books” to search for the top 10 results of the respective search engines. These four data sets were then used to compare the ranking degree to the keyword location of the websites listed.

The correlation (or lack thereof) will indicate the relation between webpage ranking and keyword location. It is believed that this research project will provide website designers with guidance to work towards achieving a higher rank with search engines by simply considering the location of webpage keywords.

1. INTRODUCTION

According to Weideman and Kritzing, the amount of data available on the Internet cannot be measured. New and existing webpage authors constantly add more by uploading new and revised webpages to web servers, some on an hourly basis.

Also, there is no central body responsible for categorising, validating or censoring data on the Internet. It is these factors that contribute to the rather chaotic situation Internet users face when attempting to retrieve relevant information from the Internet (Weideman & Kritzing, 2003).

2. BUSINESS ASPECT

Commerce on the Internet is growing at a head-spinning rate according to Podesta (2000), and those companies that cannot or will not capitalize on this interconnected electronic marketplace could be destroyed by it. Simeon (1999) claims that the Internet has become a powerful business tool. This new approach to the communication and distribution of information and services has transformed the fundamental dynamics behind many social and business interactions. The barriers and obstacles, which often accompanied traditional commerce, are giving way to new business approaches. Consumers, producers and distributors now all have flexible, fast and inexpensive ways of participating in the market for products and services around the world (Simeon, 1999).

According to Podesta (2000), what really matters about the Internet can be summarised in one word – speed – and speed is the lifeblood of business success in the 21st century. This author made the statement:

“Product development cycles are measured in months; product ramp-up is measured in days; and product lifetimes are approaching that of the mosquito. The Internet can help us keep up, and that is why our industry needs to use the Internet to keep pace with the new demands for speed” (Podesta, 2000).

The commercial potential of the web is a subject of widespread discussion, with many predictions of continuing rapid growth in the future. A web presence for a company should mean that potential clients using the web, to search for a product or service that they provide, will see the company’s site appear whichever search engine the client uses. If there are a lot of competitors on the web, the company would like their name to appear at the top of the list.

Of great concern, therefore, is the question of how to produce a successful web intervention for a business. One useful model summarising the findings of previous research is Simeon’s attracting, informing, positioning and delivering (AIPD) approach to evaluating websites (Simeon, 1999).

According to Thelwall, the first hurdle for any website is the first aspect: attracting visitors. “If the site does not get visited then its content is irrelevant” (Thelwall, 2000b). Certain steps need to be taken to make websites visible.

3. SEARCH ENGINES

Most web users employ search engines as part of at least one of their strategies to find new websites (CyberAtlas, 1999). Any website owner who wants to attract new visitors should therefore be concerned whether or not his site is registered with the major search engines. While there are literally thousands of search engines on the web, only seven are worth paying attention to, according to Nielsen (2004). Google is the major player in the game, accounting for 72% of all website searches in 2003. Nielsen stated that 97% of all web searches are performed on Google, Yahoo!, MSN, Ask Jeeves, Lycos, AltaVista and AOL (Nielsen, 2004).

Having a page indexed, the essential first stage of being recognized by search engines, is extremely important according to Introna and Nissenbaum. “Without much exaggeration one could say that to exist is to be indexed by a search engine”. If a webpage is not in the index of a search engine, a person wishing to access it must know the complete Uniform Resource Locator (URL) – also known as the webpage address (Introna and Nissenbaum, 2000).

There are a number of design and other issues that need to be addressed in order to get the page registered with search engines, states Thelwall (2000a). Some search engines such as Yahoo! have a large number of human site reviewers, and the business needs to submit the address of the site in order to have it reviewed and registered so that it can be found in a search. Other search engines such as HotBot use automated programs known as “spiders” to trawl the web and index pages (Thelwall, 2000a).

Search engines create a map of the web by indexing webpages according to keywords and then create enormous databases that link page content to keywords to URLs. Keywords are not determined a priori by the designers of the search engines’ databases nor, explicitly, by some other authority, but rather they are “deduced” from webpages themselves in the process of indexing (Introna and Nissenbaum, 2000). In a particular webpage a keyword can be any of the following:

- Words that has a strong relation to the actual content of the webpage.
- Actual keywords indicated by the webpage designer in an HTML meta tag.
- All or some of the words appearing in the title that is indicated by the HTML title tag.
- The first x number of words in a webpage.
- All the words in the webpage.

Being listed in a search engine index is, however, no guarantee that a user will be able to find the website, even if the website qualifies as a candidate for the user’s search. Websites that are not ranked highly are less likely to be visited (Courtois & Berry, 1999; Notess, 1999). Users tend to examine only the first page of search results and once they find a good match for their search, they tend not to look further down the list. Most search engines display only 10 of the most relevant results on the first page. Thus, exclusion from the top 10 results means that only a small number of search engine users will actually see a link to the website (Introna & Nissenbaum, 2000; Henzinger et al, 2002). There is serious competition for those top 10 seats.

4. USING KEYWORDS

The issue of visibility is one that is easy for the inexperienced to ignore. An otherwise excellent website may be completely ignored because few potential customers ever find it (Thelwall, 2000a)

The foundation of an effective search engine marketing campaign is selecting the best keywords that your potential customers use to find your site. Thus, selecting the right keywords requires research (Thurow, 2003). Some search engines are case sensitive, which means that searching for Chinese might yield different search results than searching for chinese (Thurow, 2003).

When creating a webpage, one should make sure the page contains between 200 to 600 words of text. Density of keywords in the regular text should be about 1 – 5%, so in a piece of text of 200 words, a designer should repeat a given keyword (or key phrase) 2 to 10 times (Galon, 1999).

According to Sullivan, one of the main rules in a ranking algorithm involves the location and frequency of keywords on a webpage (Sullivan, 2003). Sullivan refers to it as the “location/frequency method”. Pages with the search terms appearing in the HTML title tag, according to Sullivan, are often assumed to be more relevant than others to the topic. Search engines will also check to see if the search keywords appear near the top of a webpage, such as in the headline or in the first few paragraphs of text. Search engines assume that any page relevant to the topic will mention those words right from the beginning. Accordingly, a document with a high frequency of keywords in the beginning of a document is seen as more relevant (relative to the keyword entered) than one with a low frequency lower down in the document.

All the search engines consider the words at the top of a webpage more important than the words on the rest of the webpage. A keyword which appears early in a webpage has a high keyword prominence (Thurrow, 2003).

Keyword prominence is a measure of how close to the beginning of a given page area your targeted keyphrases appear. In general, according to Konia (2002), higher prominence is better, but the designer should not assume that beating your competitors’ prominence will result in a higher ranking for his/her page (Konia, 2002). To increase keyword prominence, a designer should try to move the keyphrase closer to the beginning of the area he/she is working with. If there are several occurrences of the keyphrase in the area, the designer can also remove one or more occurrences of the keyphrase that appear toward the end of the area (Konia, 2002).

Konia further states that designers can try to balance out the keyphrases in their page area in order to avoid alerting the search engine that they are crowding all of these important words into the beginning of the area. One can achieve this balance while still increasing the prominence of the area by adding keyphrases to the middle and/or end of the area, while adding more occurrence of the keyphrase at the beginning of the area (Konia, 2002). For example, adding one keyphrase at the end of the area and two at the beginning will increase the prominence of the area while avoiding an imbalanced appearance.

Here are some examples of keyword prominence calculations according to Konia (2002):

- If a keyword appears at the beginning of an area, its prominence will be 100%
- If a keyword appears exactly in the middle of an area, its prominence will be 50%
- If the keyword appears at the end of the area, the prominence will be 0%
- If a keyword appears at the beginning of the area, with a repetition of the keyword at the end of the area, the prominence will be 50%.
- If an area consists of multiple parts (such as having three headings tags on the page), all three areas are treated as a single adjoining area when prominence is calculated.

5. METHODOLOGY AND RESULTS

The authors decided to conduct a simple experiment to test whether the prominence of a keyword has a measurable effect on a website's visibility to search engines. As mentioned above, 97% of all web searches are performed on Google, Yahoo!, MSN, Ask Jeeves, Lycos, AltaVista and AOL (Nielsen, 2004). The authors chose Google, Yahoo!, MSN and Ask Jeeves to conduct the experiment on.

The authors then visited these search engines one by one and typed in one keyword, namely "books". The top ten search results from the respective search engines were then visited and inspected. The webpages were divided into three equal areas, namely Top, Middle and Bottom. The keyword "books" was then counted in each of the three areas and a percentage was calculated of the occurrence of the keyword in each of the three areas.

Furthermore, the authors used WebPosition Gold 2 to calculate the visibility percentage of each of the top ten search results in the respective search engines using the keyword "books". As indicated in Table 1 – 4, WebPosition Gold 2 visibility percentage decreases as each respective search engine's ranking decreases. See Tables 1 – 4 for a summary:

GOOGLE					
Website	Rank	Keyword Position			Visibility Percentage
		Top	Mid	Bot	
www.barnesandnoble.com/	1	50%	25%	25%	100.00%
www.amazon.com	2	100%	0%	0%	96.67%
digital.library.upenn.edu/books	3	40%	20%	40%	93.33%
www.powells.com	4	46%	37%	17%	90.00%
www.bartleby.com	5	0%	0%	100%	86.67%
www.abebooks.com	6	36%	21%	43%	83.33%
www.borders.com	7	66%	34%	0%	80.00%
www.scholastic.com	8	0%	100%	0%	76.67%
www.oreilly.com	9	50%	17%	33%	73.33%
www.cs.cmu.edu/Web/books.html	10	40%	20%	40%	70.00%

Table 1: Google summary

Yahoo!					
Website	Rank	Keyword Position			Visibility Percentage
		T	M	B	
www.amazon.com	1	100%	0%	0%	100.00%
www.barnesandnoble.com/	2	50%	25%	25%	96.67%
www.borders.com	3	66%	34%	0%	93.33%
www.bookfinder.com	4	43%	43%	14%	90.00%
www.booksense.com	5	43%	0%	57%	86.67%
onlinebooks.library.upenn.edu	6	40%	20%	40%	83.33%
www.allbookstores.com	7	45%	28%	27%	80.00%
www.barnesandnoble.com/bookbrowser	8	57%	14%	29%	76.67%
www.powells.com	9	46%	37%	17%	73.33%
www.bartleby.com	10	0%	0%	100%	70.00%

Table 2: Yahoo! summary

MSN					
Website	Rank	Keyword Position			Visibility Percentage
		T	M	B	
www.amazon.com	1	100%	0%	0%	100.00%
www.abebooks.com	2	36%	21%	43%	96.67%
www.books.com	3	50%	25%	25%	93.33%
digital.library.upenn.edu/books	4	40%	20%	40%	90.00%
www.nytimes.com/pages/books	5	56%	12%	32%	86.67%
www.powells.com	6	46%	37%	17%	83.33%
www.addall.com	7	50%	25%	25%	80.00%
www.promo.net/pg	8	64%	27%	9%	76.67%
www.alibris.com	9	86%	14%	0%	73.33%
www.ivillage.com/books	10	61%	33%	6%	70.00%

Table 3: MSN summary

AskJeeves					
Website	Rank	Keyword Position			Visibility Percentage
		T	M	B	
www.amazon.com/	1	100%	0%	0%	100.00%
www.bookfinder.com/	2	43%	43%	14%	96.67%
www.barnesandnoble.com	3	50%	25%	25%	93.33%
www.borders.com	4	66%	34%	0%	90.00%
www.abebooks.com	5	36%	21%	43%	86.67%
www.powells.com	6	46%	37%	17%	83.33%
www.amazon.co.uk/	7	48%	26%	26%	80.00%
www.allbookstores.com	8	45%	28%	27%	76.67%
www.freebooknotes.com/	9	50%	0%	50%	73.33%
www.bythebooks.com/	10	23%	54%	23%	70.00%

Table 4: AskJeeves summary

Using the figures above a Spearman Correlation was calculated for each of the four search engine result sets. The authors repeated the calculation on groups of two and three search engines. The Spearman Correlation calculated will indicate whether there is a significant relationship between the Visibility percentage of the website and the proximity of the keywords. Lastly the authors grouped all four search engines together which provided a large enough sample to calculate a Pearson Correlation, which will give an overall result.

6. CONCLUSIONS

The statistical results were summarised in table 5.

Search Engine	Top	Middle	Bottom
Yahoo!	-	-	Negative
Google	-	-	-
MSN	-	-	-
AskJeeves	-	-	Negative
Yahoo! and Google	Positive	-	-
Yahoo! and MSN	-	-	-
Yahoo! and AskJeeves	Positive	-	Negative
Google and MSN	-	-	-
Google and AskJeeves	-	-	-
MSN and AskJeeves	-	-	-
Yahoo!, Google and MSN	Positive	-	-
Yahoo!, Google and AskJeeves	Positive	-	Negative
Yahoo!, MSN and AskJeeves	Positive	-	Negative
Google, MSN and AskJeeves	-	-	-
Yahoo!, Google, MSN and AskJeeves	Positive	-	-

Table 5: Statistical summary

When the statistical results were viewed individually it was found that with Yahoo! and AskJeeves there was a negative significant relationship between the visibility percentage and keywords listed at the bottom area of the webpage. This indicates that the ranking of websites listed at Yahoo! or AskJeeves will decrease if there keywords are listed in the bottom area of the webpage.

- 6.1 It was also noted that when Yahoo! and Google are grouped there was a positive significant relationship between the visibility percentage and keywords listed at the top area of the webpage. This will indicate that a website's ranking will increase with Yahoo! and Google when the keywords are more densely grouped in the top area of the webpage. See Table 5.
- 6.2 When Yahoo! and AskJeeves where grouped it was found the there was a positive significant relationship between the visibility percentage and keywords listed at the top of the webpage and a negative significant relationship between the visibility percentage and keywords listed at the bottom of the webpage. This will indicate that the ranking of websites listed with Yahoo! and AskJeeves will increase when the keywords are listed at the top area of the webpage and that ranking will decrease when the keywords are listed at the bottom area of the webpage. See Table 5.

- 6.3 When investigating Yahoo!, Google and MSN, a positive significant relationship exists between the visibility percentage and keywords placed at the top area of a webpage. Thus, ranking will increase at these three search engines when the keywords appear at the top area of the webpage. See Table 5.
- 6.4 When investigating Yahoo!, Google and AskJeeves as a group and Yahoo!, MSN and AskJeeves as a group, it was found that there is a positive significant relationship between the visibility percentage and keywords placed at the top area of the webpage and a negative significant relationship between the visibility percentage and keywords placed at the bottom of the webpage. Thus, it is true to say that ranking will increase when keywords are placed at the top of the webpage and ranking will decrease when keywords are placed at the bottom of the webpage. See Table 5.
- 6.5 Lastly a Pearson Correlation was calculated with all four of these search engines grouped to get an overall view. It was found that there exists a positive significant relationship between the visibility percentage and keywords placed at the top of the webpage. Thus, ranking/visibility will increase when the keywords are placed in the top area of a webpage.

In general it therefore appears as if the concentration of keywords should be at the top rather than the bottom of a webpage to enhance visibility to search engine crawlers. Future research could include an investigation into how search engine crawlers will view this higher concentration of keywords, and the possibility of blacklisting as a result of keyword crowding at the top. A larger sample could be taken to confirm the trends identified in this project.

7. REFERENCES

Courtois, M.P. & Berry, M.W. 1999. Results ranking in web search engines. *Online*, 39-46: 39-46

CyberAtlas. 1999. <http://www.cyberatlas.com> [accessed 13 November 2003].

Galon, D. 1999. The savvy way to successful website promotion. Victoria, Canada: Trafford Publishing.

Henzinger, M.R., Motwani, R. & Silverstein, C. 2002. Challenges in web search engines. *SIGIR Forum*, 36(2).

Introna, L. & Nissenbaum, H. 2000. Shaping the Web: Why the Politics of Search Engines Matters. *Information Society*, 16(3): 169-186

Konia, B.S. 2002. Search Engine Optimization with WebPosition Gold 2. Plano, Texas: Wordware Publishing, Inc.

Mouton, J. 2003. How to succeed in your Master's & Doctoral studies: a South African guide and resource book. Pretoria, South Africa: Van Schaik Publishers

Nielsen, J. 2004. Statistics for Traffic Referred by Search Engines and Navigation Directories to Useit. <http://www.useit.com> [accessed 29 June 2004].

Notess, G.R. 1999. Rising relevance in search engines. *Online*, May/June, 23(3): 84-86.

Podesta, G. 2000. e-Commerce: Helping Customers Gain the Competitive Edge. *Plastics Engineering*. July. 73-74.

Simeon, R. 1999. Evaluating domestic and international Web-site strategies. *Internet Research: Electronic Networking Applications and Policy*. 9(4): 297-308.

Sullivan, D. 2003. www.searchenginewatch.com [accessed on 13 November 2003].

Thelwall, M. 2000a. Effective websites for small and medium-sized enterprises. *Journal of Small Business and Enterprise Development*. 7(2): 149-159.

Thelwall, M. 2000b. Commercial web sites: lost in cyberspace? *Internet Research: Electronic Networking Applications and Policy*. 10(2): 150-159.

Thurrow, S. 2003. Search engine visibility. Indianapolis, IND: New Riders.

Weideman, M. & Kritzing, W. 2003. Search engine information retrieval: empirical research on the usage of metatags to enhance website visibility and ranking of e-commerce websites. *Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI2003), Information Systems, Technologies and Applications*, Orlando, Florida, USA, 27-30 July 2003: 231-236.