1-1-2007

Key word placing in Web page body text to increase visibility to search engines

Wouter Thomas Kritzinger  
*Cape Peninsula University of Technology*, kritzingerw@gmail.com

Melius Weideman  
*Cape Peninsula University of Technology*, weidemanm@cput.ac.za

*Recommended Citation*  
Kritzinger, Wouter Thomas and Weideman, Melius, "Key word placing in Web page body text to increase visibility to search engines"  
http://dk.cput.ac.za/inf_papers/46
Key word placing in Web page body text to increase visibility to search engines

W.T. Kritzinger  
e-Innovation Academy  
Cape Peninsula University of Technology  
Cape Town, South Africa  
kritzingerw@cput.ac.za

M. Weideman  
e-Innovation Academy  
Cape Peninsula University of Technology  
Cape Town, South Africa  
weidemannm@cput.ac.za

The growth of the World Wide Web has spawned a wide variety of new information sources, which has also left users with the daunting task of determining which sources are valid. Many users rely on the Web as an information source because of the low cost of information retrieval. It is also claimed that the Web has evolved into a powerful business tool. Examples include highly popular business services such as Amazon.com and Kalahari.net. It is estimated that around 80% of users utilize search engines to locate information on the Internet. This, by implication, places emphasis on the underlying importance of Web pages being listed on search engines indices. Empirical evidence that the placement of key words in certain areas of the body text will have an influence on the Web sites' visibility to search engines could not be found in the literature. The result of two experiments indicated that key words should be concentrated towards the top, and diluted towards the bottom of a Web page to increase visibility. However, care should be taken in terms of key word density, to prevent search engine algorithms from raising the spam alarm.

Key words: Search engine optimization, body text, key words, location, prominence

Received 23 May 2006; accepted January 2007

Contents

1. Introduction
2. Statement of research problem
3. World Wide-Web and business
1 Introduction

The growth of the World-Wide Web (WWW) has spawned a wide collection of new information sources, which culminated in users being faced with the daunting task of determining which sources are valid, and which are not (Wagner, Cheung, Rachael and Böttcher 2006:40). Most users rely on the Web owing to the low cost of information retrieval, as opposed to the expense of having to buy a book or make use of a library. Other advantages of the Web include the convenience in terms of time and access as well as the ability to record results easily. According to Green (2000:124–137), the Web is not just about promoting one's work, but is also the interactive exchange of information, which has now evolved into a powerful business tool.

Amazon.com and Kalahari.net serve as two representative examples of the Web being used as such a powerful business tool. With regard to scientific communication, the Web is a key medium with access not only to e-journals and conference proceedings, but also to information on specific research topics (Aguillo 2000). Some tools that the Internet offers are commercial search engines and subject gateways – seen as two of the most important ones for the locating/retrieval of information (Thelwall 2002).

The Web has undergone a significant transformation in the mid-1990s with the development of search engines. Some authors claim that there are an estimated 1,3 billion sites on the Web, each providing publicly accessible information – over 1 million new Web sites are being added annually (Zhang and Dimitroff 2005). These search engines provide access to this overwhelmingly complex information resource.

Thelwall (2001) estimates that around 80% of users utilize search engines to search for information on the Web. This places emphasis on the underlying importance of Web pages being listed with search engines. An important strategy for any Web site owner is planning how visitors could find their way to their particular site (Thelwall 2001). An example of this is the business strategy that a number of large companies are using to ensure that their Web site obtains a high ranking in Google searches. Certain companies opt for paid placement to have their sites ranked higher than they normally would have (Smith 2003). This is supported by Zhang and Dimitroff (2004: 310) who are of the opinion that:

'Basically, every Internet Web publisher wants good webpage visibility in search engine results so as to increase accessibility of their webpages. Unfortunately, many websites have poor visibility in search engine rankings or may not be listed at all due to various reasons.'
According to Podesta (2000), Web site developers have to ensure that they have good rankings with search engines. Normally only 10 Web sites can appear on the first of the results page of search engines, with most of them vying for the first position. However, being listed in a search engine index serves as no guarantee that a user will be able to find the Web site (Weideman and Kritzinger 2003).

2 Statement of research problem

During this research, no empirical evidence could be found in the literature on the effect of key word location on a Web page on Web site visibility and, furthermore, if the lack of the correct use of key words plays a part in the high rate of Web site failures.

It was the intention of the authors to investigate whether there is a link between the correct usage of key words and the visibility of a Web site to search engines. The closer to the top of the search results list, the better its visibility, and vice versa.

3 World-Wide Web and business

The Internet has created an unparalleled opportunity for users to access knowledge products on a 'just-in-time' and 'on-demand' approach according to Hämäläinen, Whinston and Vishik (1996:51–58). The Internet is the fastest growing technology in the world and it has taken just seven years to reach a 25% market share, as opposed to the telephone that took 35 years, and the television, which took 26 years (Singh 2002). According to Moodley as cited in Singh (2002), the traditional 'bricks and mortar' business is rapidly being replaced by 'clicks and mortar' businesses, the latter referring to business being conducted on the Internet. Commerce on the Internet is growing at a high rate, according to Podesta (2000:73), and those companies that cannot or will not capitalize on this interconnected electronic marketplace, could be left behind.

Simeon (1999:287) further 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 underpinning 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: 297).

The Internet and, more specifically, the WWW are attracting businesses in their thousands. The main application areas seem to be publicity, marketing and advertising, direct online selling, research and development and communication and collaboration.

4 Search engines

4.1 Introduction

According to Weideman (2004), a search engine is a program that offers users interaction with the Internet through a front end, where the user can type in a search term, or make successive selections from relevant directories. The search engine then compares the search
term against an index file, which contains information about Web pages. Matches found are then returned to the user via the front end.

Weideman and Kritzinger (2003:231–236) states that search engines provide the average Internet user with a (mostly) free, apparently easy way to find general information on the Internet. Search engines are used to locate information on the Web, whether relevant or not (Alimohammadi 2003:238–242).

Search engines have been praised because of their ability to quickly locate a vast array of information on an extraordinary range of topics (Rowland 1998:222). Despite this fact, they have fallen victim to extensive criticism. Many users feel that search engines tend to retrieve information that is totally irrelevant and contrary to what the users are looking for. In addition, they have been criticized for the tendency to retrieve duplicates (Green 2000:124–137). In view of the fact that there are many different types of search engines, the issue surfaces of what ranking criteria they use to decide which Web sites are good enough to be included in their database. Each has its own rules for searching and of establishing which Web sites to include in their database (Synder and Rosenbaum 1999:375–384).

It is claimed that search engines normally use two methods to determine which Web pages to index. They either follow links from previously registered Web pages, or they allow users to register the URL addresses of unregistered pages (Thelwall 2000:149–159).

4.2 Indexing

Instead of a central catalogue, the Web offers a choice of dozens of different search tools, each with its own database, search capabilities and method of displaying results (Tyner 2001). As a result, searching for relevant information on the Web can be a time consuming and frustrating process.

With the estimated 1.3 billion Web sites available (Zhang and Dimitroff 2004:310–320), search engines are intended to assist searchers in sorting through the large amount of information that is available on the Internet. This plays an important role in the process of information retrieval. Green (2000:125) provides a very basic definition of a search engine, namely 'a database that contains massive amounts of data about websites'.

This author further states that the database is compiled using a program known as a 'crawler', which visits sites and indexes them in a predetermined way. The index is updated regularly, either by human editors or by crawlers. Both humans and crawlers simply collect information on new Web sites by visiting as many Web sites as possible, and then building them into the index (Weideman 2005).

5 Web sites (HTML coding)

The coding of a typical HTML Web page consists of a header and a body section. The header section contains instructions and information, which is not displayed as part of the Web page, while the body coding determines what and how the user will see the Web page content on the screen. A simplistic version of the HTML code of an imaginary Web page could read as follows:

<HTML>
It is this HTML coding which search engine crawlers evaluate and index when they visit a Web page. Metatags are located inside the HTML coding.

The body area is the main area of an HTML page, which contains all the visible text and images that appear in the browser window. The body area begins with the body tag, which is coded <body>, with a closing tag of </body>, and is divided into a number of sub-areas. This research focused on the location of key words used in the body section of a Web page. A simplistic version of the HTML code of the body section of an imaginary Web page could read as follows:

```html
<body>
  <div id='down'>
    This is the first sentence a search engine crawler will see.
  </div>
  <div id='top'>
    This is the first sentence a human viewer will see.
  </div>
</body>
```

In this example, the first sentence will be moved down and the second sentence will be moved to the top of the screen when the code is viewed in an Internet browser. This highlights the fact that the crawlers and humans do not view Web pages in the same way.

Search engines constantly change the tags on which they place emphasis. One aspect of crawlers which seldom changes, however, is the fact that they index title tags and body text, hence the importance of placing key words throughout the visible body text on Web pages. This would guarantee that search engines could find and record all the relevant key words (Thurow 2003:21, 70–86).

### 6 Key words

Pages with the search terms appearing in the HTML title tag are often assumed to be more relevant than others to the topic. The focus of this project was on the location of key words in the body text of a Web page (Sullivan 2003).

The term 'search query' refers to the word(s) that a user will enter as input parameters to a search engine. Most of these engines' algorithms try to match the terms contained within the query to key words (as found on Web pages) it has stored within its index repository. According to Konia (2002), multi-word key word phrases could produce higher rankings in search engine results.
'Key word prominence' refers to how often and where key words appear on a Web page. One recommendation is to place important key words at or near the top of a Web page and in the 'title' or 'description' meta tag (Wong 2004). The authors, however, did not find any empirical results to confirm or refute the recommendation made by this author.

One of the foundations of an effective search engine marketing campaign is to select the best key words that potential customers could use to find the site (Weideman and Kritzinger 2003:233). However, ranking in search engines is determined by how relevant the algorithms consider a Web page to be for a particular key word (Nobles and O'Neil 2000). According to Meadhra (2004), the following list orders the possible key word locations from the highest to the lowest priority:

- Domain name
- Page title
- Headings (enclosed within HTML H1, H2, H3 tags)
- Body text
- Meta tags
- Links
- ALT text.

The same author further claims that a page containing multiple instances of a key word will generally rank higher than a page on which the key word appears only once. Furthermore, the interaction of key word placement and repetition is one of the areas where it is almost impossible to second-guess the ranking algorithms.

Many authors have stressed the importance of placing carefully chosen key words in HTML body text in the past. Bowman (2004a) lists the following as being important issues regarding the use of key words:

- How many times they are repeated
- Where they appear
- How they are positioned relative to each other
- Which tags surround the key words.

The same author also lists areas where key words should be included as being domain names, Web page titles, meta tags, header content, body content, text links, filenames, ALT text, named anchors, ordered lists, directory names and cascading style sheets (CSS) classes (Bowman 2004b).

Some references to the location and frequency of key words inside the body text of a Web page were found. Sullivan (2002a; 2003) refers to it as the 'location/frequency method'. Sullivan further claims that key words in the search query that also appear in the title tag and close to the top of a Web page (such as in the headline or in the first few paragraphs of text), have a positive effect on the ranking of this Web page.

Pages with the search terms appearing in the HTML title tag, according to Sullivan (2002a; 2003), are often assumed to be more relevant than others to the topic. Search engines assume that any page that is relevant to the topic will mention those words right from the beginning. Accordingly, a document with a high frequency of key words at the beginning is viewed as more relevant (relative to the key word entered) than one with a low frequency lower down in the document. However, the authors did not find any empirical evidence that substantiated Sullivan's statement.

Thurow (2003:71–73) states that all the search engines consider the words at the top of a
Web page more important than the words on the rest of the Web page. How high up a key word is on a Web page is referred to as 'key word prominence'.

These last two references were central to this research, and the authors had set out to provide evidence to support or contradict these claims by Sullivan and Thurow. No empirical evidence could be found in the literature to substantiate their premises.

### 7 Search engine optimization (SEO)

An increasing number of Web sites are turning to search engines as their primary marketing route (Centaur Communication as cited by Zhang and Dimitroff 2005:667). Driven by this trend, SEO is a lucrative field for entrepreneurs. Many companies offer SEO services to help enhance customers' online experiences by pushing relevant Web sites to the fore (Kanaley as cited by Zhang and Dimitroff 2005:667).

These services range from the submission of Web pages for free optimization, to optimization software for sale, to free Web pages submission. Free Web pages submission cannot ensure that the submitted Web site will eventually earn a good position. Conversely, many Web publishers, especially non-profit institutes or organizations, cannot afford the cost of optimization software or expensive Web site submissions. A growing industry has emerged that offers advice (for a fee in most cases) on maximizing Web pages placement. This advice about which techniques will provide optimal ranking results is hinted at on the Web itself, however none of those offering advice provide details about any empirical research on which their recommendations are based.

While a common theme among these 'advice providers' is 'location, location, location', the specific advice is fairly generic and based on conventional wisdom. Research on this emerging topic has not been reported in research-orientated publications. Some Web sites offer SEO tips based on their experiences, while others merely provide a basic introduction to the topic (Greenberg as cited by Zhang and Dimitroff 2005:668; Sullivan 2002b).

#### 7.1 Strategies and guidelines

There are various ways of ensuring Web site visibility to search engines. One way is the aspect of key word location, and it is crucial in ensuring that Web sites are visible to search engines (Nobles and O'Neil 2000). These authors list the following as some of the strategies that can be used as a guide to creating top-ranking pages:

- Consider your overall Web site. What is the goal of your site? What are you trying to accomplish? Which pages are so important on their own that they should be able to be found separately from your home page?
- Determine what the current rankings are. This can be done by manually searching for important key words associated with the site at each of the engines. Alternatively, special software can be purchased or online services could be used to accomplish this.
- Consider what key words are important to the site. Which key words are central to the Web site (potential visitors might type into the search box of a search engine)?
- Determine where to place those key words for optimum key word prominence (Nobles and O'Neil 2000).

These authors also provide the following guidelines:

- Fine-tune key words to narrow down the competition. By fine-tuning key word
phrases, the Web site designer is targeting traffic to a particular Web page.

- Pair a general key word with a more specific one.
- Consider combining key words that are remotely related.
- If a specific key word is commonly misspelt, include this version as one or more of the key words (e.g. accommodation, acommodation, accomodation, acomodation).

8 Methodology and results

8.1 Pilot study

The research design that was used in this dissertation was an empirical field/natural experimental design. According to Mouton (2003:157), the definition of this kind of research design is 'studies that are usually quantitative in nature and which aim to provide a broad overview of a representative sample of a large population'.

An experiment was conducted to test whether the prominence of a key word has a measurable effect on a Web site's visibility to search engines. A study on Web site benchmarking was done by Boisvert and Caron, during which individual Web sites were visited and inspected in a similar way as in this study (Boisvert and Caron 2006:180).

According to Nielsen (2004), 97% of all Web searches are performed on AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo!. The authors decided to do a pilot study in an attempt to determine whether any patterns of key word location would be noticeable. Subsequently, a random selection of four of the seven search engines was done, and only the first ten search results were inspected, that is on AskJeeves, Google, MSN and Yahoo!

The authors conducted a single-word search on the above four search engines, using the key word 'books'. The top ten Web sites from the result pages of search engines were then visited and evaluated. The text in the 'body area' of the source code within the Web pages was divided into three equal areas, namely top, middle and bottom. The number of characters in the 'body area' was counted and divided by three to achieve this. The key word 'books' was then counted in each of the three areas and a percentage was calculated to determine the occurrence of the key word in each of the three areas. Furthermore, the authors used an industry standard program to calculate the visibility percentage of each of the top ten search results in the respective search engines using the key word 'books'. If the Web site had a first place ranking for the key word 'books', then the visibility percentage was 100%.

Since the data were not normally distributed, the most efficient way to determine whether a relationship between two variables (e.g. key word prominence and visibility percentage) existed was to use the Spearman Rank Correlation Coefficient: 'The Spearman's correlation analysis compares the order, rather than the numeric magnitude, of the variables and is mostly used to examine the strength of the relationship between two variables' (Hirsch and Riegelman 1996).

The Spearman Correlation was used to determine whether there was a significant relationship between the visibility percentage of the Web site and the location of the key words. Table 1 represents the combined statistical results of the pilot study.

Table 1 Statistical results of the pilot study

<table>
<thead>
<tr>
<th>Search engine</th>
<th>Top</th>
<th>Middle</th>
<th>Bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Spearman's correlation analysis produces three possible answers, namely: positive, negative and not-significant relationship. As reflected in Table 1, all the significant answers in the bottom area are negative and all the significant answers in the top area are positive [dashed (-) in the table represent not-significant answers]. This means that if a Web site designer places key words in the top section of the Web page, it will have a positive impact on the Web site's visibility to search engines. Conversely, if a Web site designer places key words in the bottom section of the Web page, it will have a statistically negative impact on the Web site's visibility to search engines.

8.2 Empirical experiment

Finally, the empirical experiment was expanded using the seven search engines mentioned above, namely AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo!. For this empirical experiment, each of the top 20 search results in the respective search engines was used.

Table 2 represents the statistical results of the empirical experiment. As the data were not normally distributed, the Spearman correlation analysis was used.

<table>
<thead>
<tr>
<th>Search engine</th>
<th>Top</th>
<th>Middle</th>
<th>Bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo!</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Google</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MSN</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AskJeeves</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yahoo! and Google</td>
<td>Positive</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yahoo! and MSN</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yahoo! and AskJeeves</td>
<td>Positive</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td>Google and MSN</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Google and AskJeeves</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MSN and AskJeeves</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yahoo!, Google and MSN</td>
<td>Positive</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yahoo!, Google and AskJeeves</td>
<td>Positive</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td>Yahoo!, MSN and AskJeeves</td>
<td>Positive</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td>Google, MSN and AskJeeves</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yahoo!, Google, MSN and AskJeeves</td>
<td>Positive</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The results of the search engines were then grouped, as it was assumed that a Web site designer would like to have his/her Web site rank well with all search engines and not with only one. Again, all the significant results in the top section were positive, while all the significant results in the bottom section were negative.
As can be seen in Table 2, all the significant answers in the bottom area are negative and all the significant answers in the top area are positive. The conclusion is the same as for the pilot study above: key words concentrated higher up on a Web page will positively influence its visibility, while too many lower down will have a negative effect statistically.

The results of the search engines were grouped and, again, all the significant results in the top section were positive, while all the significant results in the bottom section were negative.

9 Conclusions

As a general conclusion, it can be stated that key words should be concentrated at the top of a Web page, and diluted towards the bottom. Web site authors must apply these rules when viewing the actual HTML coding, not the Web site as it is reflected by the browser on the screen.
The experiments indicated that for all the search engines used in this research, with one exception (Yahoo!), there were significant negative relationships between the visibility percentage and the key words listed at the bottom area of the body text of a Web page. A possible reason for this is that Yahoo! uses human editors. Where automatic crawlers can easily calculate key word density and allocate rankings accordingly, human editors are probably more interested in quality content and are unlikely to count key words in specific areas on a Web page.

Designers of e-Commerce based Web sites should pay close attention to the use of key words on Web pages. For every separate HTML page, the relevant key words should be identified and placed inside the top text areas. However, care should be taken to ensure that the key word density does not lead to spam penalties by search engine algorithms. Finally, alternative spelling of complex words should also be included, to match common spelling mistakes made by searchers.

It was the intention of the author to investigate whether there is a link between the correct use of key words and the visibility of a Web site to search engines. The results show that a Web site's ranking to search engines increases when the key words are placed at the top of the Web page. Similarly, the results show that a Web site's ranking in search engines decreases when the key words are placed at the bottom of the Web page. It is finally concluded that this research has highlighted some definite trends in the prominence of key words to be used on Web pages, with special emphasis on commercial sites.

10 References


**Disclaimer**

Articles published in SAJIM are the opinions of the authors and do not necessarily reflect the opinion of the Editor, Board, Publisher, Webmaster or the Rand Afrikaans University. The user hereby waives any claim he/she/they may have or acquire against the publisher, its suppliers, licensees and sub licensees and indemnifies all said persons from any claims, lawsuits, proceedings, costs, special, incidental, consequential or indirect damages, including damages for loss of profits, loss of business or downtime arising out of or relating to the user’s use of the Website.