PROCEEDINGS

7th ANNUAL CONFERENCE
ON WORLD WIDE WEB APPLICATIONS

29-31 August 2005

Cape Town
South Africa

URL: http:www.zaw3.co.za

TO WHOM IT MAY CONCERN

Herewith to testify that the papers accepted for the 7th Annual Conference on World Wide Web Applications, 29-31 August 2005, Cape Town have been peer-reviewed by independent and external peer-reviewers.

Prof P.A. van Brakel
Conference Chair
7th Annual Conference on WWW Applications
c/o Faculty of Business Informatics
Cape Peninsula University of Technology
Cape Town

+27 21 469-1015 (landline)
+27 82 966-0789 (mobile)

Search engine visibility: a pilot study towards the design of a model for e-commerce websites

R. Chambers
Faculty of Business Informatics
Cape Peninsula University of Technology
Cape Town
South Africa
rickard@polka.co.za

M. Weideman
Faculty of Business Informatics
Cape Peninsula University of Technology
Cape Town
South Africa
meliusw@yahoo.com

Abstract

In this research project, the authors attempt to compile a model that consists of guidelines towards achieving an electronically visible and well optimised website. This model will provide a strategy for achieving website visibility which could be used by e-Commerce companies to achieve higher page ranking within most search services. This model is expected to also improve the website usability, benefiting the client and searcher.

There are many methods for finding a website on the Internet, one being the use of a search service which usually is a directory or a search engine. Each one of these services has their own search strategy for categorising websites, which in the end will determine where a site would be listed. Due to the importance of these search strategies, one should emphasize the importance of improving visibility techniques in the development of a website, to satisfy both directory and search engine strategies.

For the purpose of explaining the effective use of optimisation techniques, an e-commerce based website was selected to be used as an example. Elements gathered through academic literature were then applied on the website without modifying the layout of the site. Areas affected were: meta tags, frames, graphics descriptions, JavaScript and a site map. It should be noted that the testing of the applied changes would only occur at a later stage due to the nature of the research project.

The model that is based on principles identified in academic literature, if applied, is expected to increase the visibility of most websites and in the process satisfy some of the requirements of the directories and search engines to index webpages.

Keywords: internet, website, search engine, visibility

1. Introduction

In a global information economy, there are no boundaries. Making use of the Internet effectively, can play an important role in enhancing global competitiveness and attract foreign investment.

Authors of websites of commercial concern, whose income depends on traffic, sometimes fail to get / keep their pages indexed and ranked high in the search engines, due to lack of skill, knowledge and resources. One author stated that:

> "Unless further research is undertaken to analyse and document the problems being faced by Small Businesses when they connect to the web and present this in a formal easily accessible format, by both the academic community and SME community, many Small Businesses' future efforts to develop effective websites will prove to be unsuccessful" (Boyes and Irani, 2004:193).

For most of SMMEs, the only solution towards a more visible website and therefore improved listing in search service results are to turn to outsourcing their visibility needs. The problem with outsourcing is that there are always the problems of finding reliable, cost-effective, independent advice that will solve their specific problem (Baard, 2004).

It is evident that there is a real need for assisting SMMEs, to help them gain the advantage when providing a visible website.

A literature survey has identified a number of authors (Rowley, 2001; Strinivasan et al, 1996; Thurow, 2003; Nobles and O'Neil, 2000; Weideman, 2004) in the field of website visibility. These authors collectively claim that:

- visibility is an ongoing process,
- content must be accurate and compelling,
- frames should be avoided where possible,
- meta tags should be included even though they are not often used,
- Flash should be avoided as it increases download time of web content,
- graphics should be minimized to increase download speed,
- a site map should be included,
- JavaScript should not be used, as it is illegible to crawlers, and
- websites should be registered with the most popular search engines.

2.   Literature Survey

2.1 SMMEs

As large enterprises have restructured and downsized, small, medium and micro enterprises (SMMEs) have come to play an increasingly important role in South Africa's economy (Bester, 2003).

According to the Department of Trade and Industry (2001), the small business sector makes a valuable contribution to the economic development in South Africa. The harsh reality though, is that between 70% and 80% of these small businesses face failure within the first three years of existence (Barron, 2000:1).

Possible reasons for these high failures have been identified in several research projects. Research done by Baard (2004) and Fillis, et al. (2004:180) respectively indicated that failures of SMMEs are partially due to internal factors, as a single manager operates the majority of small firms. A manager as sole decision maker sometimes lacks the necessary skills to manage vital tasks, such as adequate planning and financial control. But failures in

SMMEs are not always to be blamed on bad management. The SMME sector also exists in a hostile environment containing several constraints.
One of which, and probably the biggest, is the operation within a global environment intensified by competition (Baard, 2004).

According to Isaacs (2004), many small to medium-sized companies can improve their productivity dramatically with small improvements to their IT infrastructure and therefore could provide a possible competitive advantage. One of these improvements involves the adoption of the Internet.

2.2 Adoption of the Internet

During the last decade the Web has experienced a continuous growth rate to become an important information resource. It took the Internet just 7 years to reach a 25% market share as opposed to the 13 years for cell phone technology. This growth makes the Internet the fastest growing technology the world has ever encountered. The growth of the most popular technologies is indicated in figure 1 (Singh, 2002).

Figure 1: Time taken for technology to reach 25% market share (Singh, 2002)
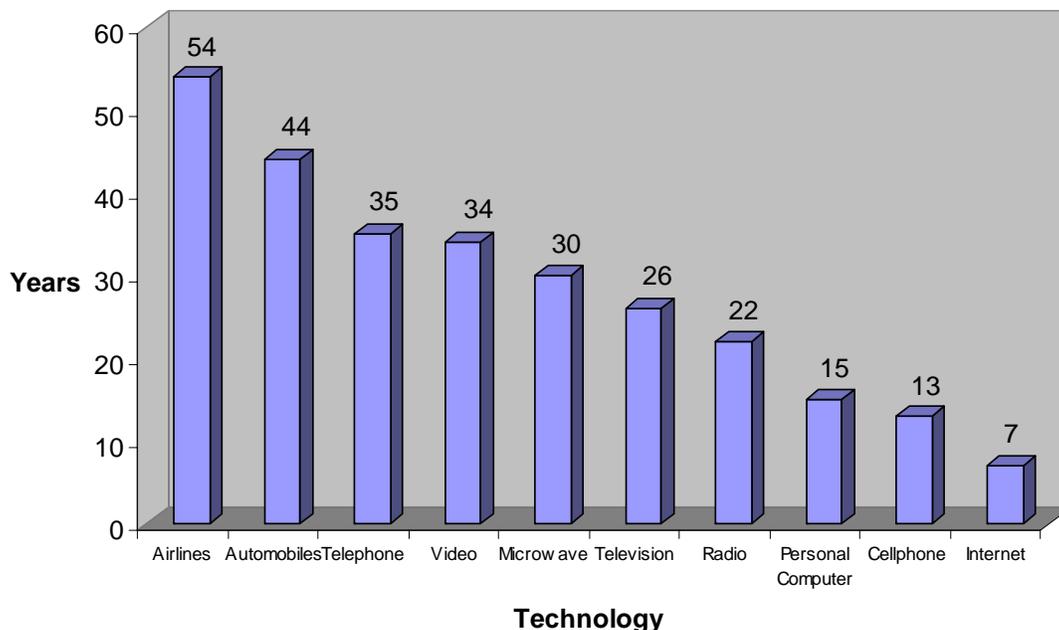
Miller (1997:15) in Hoffman (2002) also states that since the development of the personal computer in 1981, the Internet has undoubtedly been the single most important development.

Today, relatively cheap communications technology, software and the growth of Internet service providers offer Internet access to both individuals and organisations at a relatively small cost. Internet access has been made available in most public libraries and in community centres, providing access to information previously only accessible by large organisations (Bester, 2003).

Because of the lack of knowledge, time and funds experienced by SMMEs, website owners will often simply launch their site, do nothing to ensure visibility and wait for surfer

traffic that could never come. According to Guenther (2004a:47-48), it is an area where project teams spend the least amount of time and thought.

It could be compared to buying an expensive user-friendly phone system and then having an unlisted phone number, or placing an attention-grabbing billboard in the middle of a forest, next to a dirt road (Subia Creative, 2002).

A potential buyer starts a search for information in the hope of finding a solution to current needs. Those websites displayed on the first page of the returned results, will enjoy a greater market share of search engine referred leads. One author stated:

> "simple, mainly non-technical tests need to be carried out on any newly designed website, including checking how easy it will be for potential customers to find it, by typing relevant key words into major search engines" (Thelwall, 2000:150).

2.3   Type of Search Services

There are many methods for finding a website on the Internet, but one of the leading methods used, is through the use of a search service (Oppenheim et al, 2000; Green, 2000). These search services can generally be categorized into two categories, each with its own strategy for categorizing websites.

To understand the features required to design a website with higher visibility, one must first understand how these two search services operate.

2.3.1   Directories

Green (2000) defines a web directory as a pre-defined list of websites, compiled selectively by human editors through categorization, according to subject or topic.

A web directory provides a service that allows a user to navigate through several listings and an option to search the entire directory. According to Green (2000), the major web directories also make use of search engine indices to provide secondary results, whenever their human compiled indices fail to produce matching results to the user's query.

Yahoo is one example of a directory, which consists of human indexers examining documents and identifies its principle concepts with a controlled vocabulary. Web owners need to submit the address of the site in order to have it reviewed and registered, so that it can be found in a search. Once indexed by web directories, it will remain listed within that directory, unless it is removed due to reasons such as excessive use of spam (Green, 2000).

2.3.2   Search Engines

A Search engine is a search service that uses retrieval software comprising out of spiders/crawlers that examine websites and then index them into a database of website listings according to their relevancy (Green, 2000).

Search engines use their own indexing software and strategies to continuously travel the web searching for the most up-to-date content possible. The indexing software (also referred to as spiders or crawlers) is responsible for downloading webpages for the purpose of including them in the search engine results. Although all search engines use different algorithms to rank a website, they operate according to similar principles. All

search engines primarily strive to retrieve and display relevant results (webpages), which contain words, or terms that match the user's search query (Green, 2000; Guenther, 2004a).

Green (2000) also states that although search engines with its automatic indexing nature can analyse far more websites than directories such as Yahoo!, it could become less useful, the larger their index becomes.

2.4   Visibility

"There are two distinct elements of the Web: the visible and the invisible" (Green, 2000:127).

In practice, information searchers would not only like to see the item or content-rich page that they are searching for in the top ten hits, but also shown on the first screen. Businesses would also like their webpages to appear in this set, where it is most likely to capture the interest of the searcher (Rowley, 2001:203-204). Ideally, every web developer should invest in a search engine strategy that is based on the effective use of techniques and programming tools, for example meta-tags or ALT-tags to increase the chances of the website being listed in high positions in search queries. One author stated:

"A sound search engine strategy provides mechanisms for allowing the frequent evaluation of the site visibility" (Constantinides, 2002:205).

In most cases, except where there is no need for a website to be visible, the developer's responsibility should also include constantly improving the visibility of the website, by getting the pages indexed and ranked high in the search engines and also to keep them there. This way, the developer ensures constant website visibility which, in turn, should generate constant traffic to the website.

Several methods exist in making a website more visible for most search engines to receive a higher relevance score. The exact methods used by search services are a closely guarded secret due to several reasons such as increased competition and problems with spam (see section 2.5.8). Some general ranking principles are sometimes discussed in search services'help sections and other methods are obvious from research done on search results (Notess, 1999:2).

Many sources claim that the first and most effective way to make a site more visible, is to have it listed with the most popular search engines and web directories (see Table 1), to allow their crawlers to visit and index these pages (Rowley, 2001:205-207; Strinivasan et al. 1996:79-81; Thurow, 2003; Nobles and O'Neil, 2000; Weideman, 2004; Guenther 2004a). Although this is possibly the most effective step, one should also adopt Web development standards for HTML use, metadata, and content structure prior and during the development of a website.

Some of these design elements are general and could be expected from any site that is well designed (Guenther, 2004b:54-56). When these standards are not implemented accordingly, it places the burden of getting listed high in an index on the shoulders of the web developing team who is responsible for these standards.

Table 1: Popular Search Services (Dickson and Marshall, 2004:34)

| TOP 10 SEARCH CHANNELS IN EUROPE - JAN 04	UNIQUE AUDIENCE	ACTIVE REACH	PAGES PER PERSON	TIME PER PERSON
Google Search	55,641,382	47.30%	52	00:15:24
MSN Search	27,151,382	23.10%	12	00:04:08
Yahoo Search	12,676,097	10.80%	21	00:07:30
Google Image Search	10,275,673	8.70%	36	00:09:13
AOL Search	5,846,613	5.00%	20	00:09:05
Virgillo Ricerca	4,350,538	3.70%	24	00:07:57
T-Online Suche	3,898,809	3.30%	8	00:04:15
Volla Search	3,458,755	2.90%	17	00:08:03
Lycos Europ Search	3,117,113	2.70%	13	00:04:54
Microsoft Search	2,683,728	2.28%	3	00:01:51

2.5   Factors Affecting Visibility

2.5.1 Frames

Frames could be seen as a useful feature to implement on a webpage, which could keep the user from having to continuously scroll down to look for what they want. Although frames could be used in some cases, it is not as popular as it was when it was first offered. Frames present several problems to both the user and the search engine crawler. One of these is the problem users experience when trying to print a website. Adams (2003) states that the easiest way to tell if a page is using frames is to scroll down towards the bottom of the page. If one portion of the page moves, but another portion doesn't, the page contains frames.

According to Adams (2003), when printing information from a webpage containing frames the printer usually prints the information in the frame last clicked on. For example, in figure 3, when a user clicked on frame 1 and then clicks on the print button, the printer would print the information in frame 1. If the user wants to print the information in frame 3 then the user should first click on frame 3 and then request a printout.

Figure 3: Layout of website containing frames
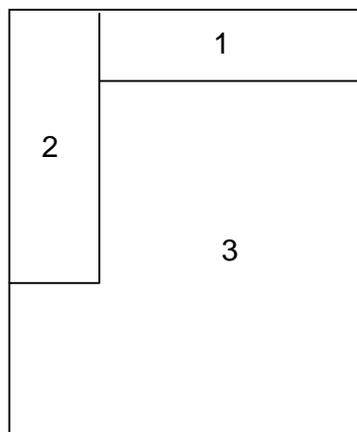
One author also stated that crawlers often ignore sites containing frames beyond the home page and that significantly reduces the chance of users visiting it (Thelwall, 2000:157). It is also known that some websites could index each frame as a distinct page, which would cause the site to be displayed partially. This means that if the content matching a query is in a pane, visitors clicking on those links will see only the pane, not the full page as it was originally designed.

2.5.2 HTML

Structure of data plays a major role on how data is retrieved in a search. Blair (1990:122) states:

> "Document retrieval is based on how the documents on the system are represented, and the effectiveness of retrieval will rely, more than anything else, on the quality of these document representations."

HTML tags mostly appear in pairs. For example when using the tag <Title> one must close it again with the tag </Title>. The typical webpage consists of a header and a body section. The header section is usually used to give information about the content of a webpage which is not displayed as part of the webpage. The body section contains several meta tags which determine how the user will see the webpage. A basic structure is listed below.

```
<HTML>
<HEAD>
    <TITLE>Title of Website</TITLE>
</HEAD>
<BODY>
    ----Code which determine how webpage appear to user-----
</BODY>
</HTML>
```

Webpages written in HTML fall neither close to free text nor to well-structured data as seen in figure 4.

Figure 4: Example of HTML code

```
<HTML><HEAD><TITLE>Adventure Cycling Association</TITLE>
<META http-equiv=content-type content=text/html;charset=ISO-8859-1>
<META content=  "Adventure Cycling is America's bicycle travel inspiration and
                resource; creating bike routes, offering bicycle tours, and
                publishing the Adventure Cyclist magazine. Self-contained bike
                trips are our specialty." name=description>
<META content=  "bicycle tours, bike magazines, bike tours, bicycle maintenance,
                bicycle touring, cycling tours, bike trips, adventure cycling,
                bike maps, bike routes, cycling, bicycling, cross country, recumbent b
                transamerica, tandem bicycles, touring bikes, cycling club, bicycle tr
                bike travel, cycling association, bicycle clothing, touring, bicycle t
                recreational bicycling" name=keywords><LINK href="ac/assets/favicon.ic
<SCRIPT language=JavaScript
src="Adventure Cycling Association_files/ac_open.js"></SCRIPT>

<META content="MSHTML 6.00.2800.1276" name=GENERATOR></HEAD>
<BODY bgColor=#ffffff leftMargin=0 topMargin=0 marginheight="0" marginwidth="0">
<DIV align=center>
<TABLE cellSpacing=0 cellPadding=0 width=767>
  <TBODY>
  <TR>
    <TD align=middle colSpan=5><A |
      href="http://www.adventurecycling.org/index.cfm"><IMG
      title="Adventure Cycling Association Home" height=51
      alt="Adventure Cycling Association"
      src="Adventure Cycling Association_files/ac_logo.gif" width=243
      border=0></A><BR><IMG height=1
      src="Adventure Cycling Association_files/graydot.gif" width=758
    border=0></TD></TR>
```

HTML coding provides limited structural information, typically used to control the layout of the webpage. The code might seem useless by looking at it, but it is in fact a particularly valuable source of metadata. Most metadata is usually invisible to the user and thus creates opportunities for several spam techniques. The value of the metadata controlling the layout of the webpage in HTML is that it is more difficult to use spamming techniques without affecting the experience of the user (Henzinger et al., 2002).

Although webpages might seem to be unstructured, they do provide some structure exhibited through the use of HTML coding. One author believes that:

> "HTML coding provides unintentional structure because it is not typically the intent of the webpage author to describe the document's semantics. Rather, the author uses HTML to control the document's layout, the way the document appears to readers" (Henzinger et al., 2002:9).

By analysing the data provided within the HTML code, the search engine could determine the weight of a phrase in several ways which is discussed in the paragraphs to follow.

2.5.3 Keywords

A study done by Duncan and Fichter (2004) highlighted the importance of keywords by establishing how users search and navigate websites:

- Users do not read through an entire site, but rather scan for words that matched their particular need.
- Users over time have learned to ignore links above the main content area, expecting these images to be banner ads.
- Users are not familiar with some abbreviations.
- Text links work better that graphic links.

The use of effective keywords is extremely important in describing the content of a site, as search engines extract information from a webpage's initial content to effectively categorise it (Guenther, 2004a).

One author describes keyword prominence by making use of an inverted pyramid writing method. This method explains a top-down approach where content is displayed from most relevant to less relevant, thus benefiting the author of a site and the user visiting it (Guenther, 2004a).

Another author also stated that one should also try avoid repeating keywords more than two or three times and never list the same keyword twice in a row (Craven, 2001:203).

2.5.4 Meta Tags

Meta tags are HTML tags that supply information about the content of a webpage, such as what HTML specifications a webpage follows, or a description of a webpage's content. A meta tag, however, does not affect how a webpage is displayed on a browser.

According to Henzinger et al. (2002:7), meta tags are currently the primary way to include metadata within HTML. This is also one retrieval method that is still included by most search engine algorithms when it comes to indexing web content (Guenther, 2004a).

Although meta tags used to be possibly one of the best methods to increase a webpage's weight and relevancy, its use does not guarantee a top place in the search results. This is mainly due to the spamming techniques (see also 2.5.8) created to manipulate search engines into indexing sites not relevant to a search request (Guenther, 2004a).

Research done by Craven (2003) to determine the relevancy of meta tags clearly showed that the title tag carries the highest weight, while the H1 tag (heading 1) and H2 tag (heading 2) ended up second and third (see Header Tags, section 2.5.10).

According to Henzinger et al. (2002:9), one should also constantly check for pages containing syntax errors as it could be rated lower than pages containing no errors. It is therefore essential to make sure that when designing or updating a page to check for spelling mistakes in the content, as well as in the meta tags. By doing this the author assures that the page is of high quality, thus benefiting the user and setting high standards for the search algorithm.

2.5.5 Title Tag

The title tag provides the description that appears at the top of the browser window, in order to let an Internet user know where they are. One author states that the title tag should be a richly written, keyword-dense sentence that accurately describes the content of that particular webpage (Galon, 1999). The company name alone in the title would tell the visitor very little about the content of the particular page and does not tell the crawler what type of content the page has. The title should be focused on describing the content of the particular webpage, in a keyword-rich manner. This technique should be repeated with each and every page of a website. Since none of the major engines are case sensitive, it is considered in order to list keywords in lower case only.

Another author also states that creating an effective title for a website

> "is not simply a question of what data elements are to be included, in what order and with what punctuation, font and capitalization, but also of how the values of those data elements are to be derived" (Craven, 2002).

2.5.6 Description Tag

Meta description tag provides a description of a website and is sometimes used on the search engine results-page. According to Konia (2002:18), one of the most important factors in determining page relevancy is the keyword prominence of the content in the description tag.

Although addition of a description tag does not improve retrievability of Webpages on infoseek and AltaVista (Craven, 2003) and several other search services it should still be included for those single ones, which still do support it.

Of all the pages on a website, the home page tends to be less likely to retain its description in the meta tag due to the frequent redesign of home pages. Usually, the use of some page creation programs causes the descriptions to be easily overlooked due to its invisibility when previewing the page (Craven, 2001).

According to Craven (2003), the description tag can be used for an abstract of what the site is about. The description should be no more that 200 characters and it should be concise. The author also advises not to repeat the title in the description.

To compile a description tag that best describes the whole nature of a particular webpage, one could use words drawn from the title or meta tagged keywords, text from the top bit of text on the page and also text and passages emphasized by different fonts (Craven, 2003).

2.5.7   Header Tag

As with Microsoft Word, HTML also has built in styles for headings to differentiate between importance levels of text which is usually used to break up text into paragraphs. The different options allow the designer to develop large and bold text in a HTML document marking the beginning of a new paragraph or section (Henzinger et al., 2002:9).

In HTML there are six levels of headings. H1 is the most important; H2 is slightly less important, and so on down to H6, the least important. Some search engines recognise the use of header tags as a safe method to weight keywords, due to its connection with a heading of a paragraph.

The different types of heading tags are highlighted by way of examples below:

The code:

```
<H1>Heading 1 </H1>
<H2>Heading 2 </H2>
<H3>Heading 3 </H3>
<H4>Heading 4 </H4>
<H5>Heading 5 </H5>
<H6>Heading 6 </H6>
```

will be displayed on a webpage as in Figure 5.

Figure 5: Representation of the different types of headings

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6

2.5.8 Spamming

The term spam as used in website visibility should not be confused with for example e-mail spam - users receiving large numbers of unsolicited mail. Spam in the current context is described below. Since the use of search engines is seen as a very popular way of finding information on the web (Oppenheim et al, 2000; Green, 2000), authors of (mostly) commercial websites quite often turn to manipulative ways to get a higher placing in the top results of search engines. This process of deliberately manipulating search engine ranking algorithms is known as search engine spam (Henzinger et al., 2002:2; Notess 1999:2).

According to Henzinger al. (2002:3) spamming approaches can be categorised into 3 categories; text spam, link spam and cloaking.

- **Text Spam**
  An example would be, the author of a website repeating one or more keywords several times, in such a way that it does not disturb the user. This is sometimes accomplished by presenting text in the smallest size or in the same colour as the background (Henzinger et al., 2002:4; Notes, 1999).
- **Link Spam**
  An example would be an author putting a link farm (page containing several links to other pages) at the bottom of every webpage to manipulate systems that make use exclusively of incoming links to determine a webpage's importance (Henzinger et al., 2002:4).

  Another form of link farms that are more sophisticated is known as doorway pages. These pages consist entirely out of links sometimes pointing to thousands of sites and are usually not intended to be viewed by users conducting a search (Henzinger et al., 2002:4).
- **Cloaking Spam**
  The cloaking technique is very popular on websites containing intense multimedia content. In such a webpage the developer often creates a separate text page to assist the search engine ranking in the site. However, cloaking is not always used for this reason; spammers also make use of cloaking as a spam technique and therefore are viewed as spam by search engines. (Henzinger et al., 2002:4; Robinson, 2005).

As soon as new search engine techniques develop, new spam techniques are developed in parallel. To stay on top, search engines have to continuously develop and improve techniques for detecting and fighting webpage spam. Due to this never ending battle between search engines and spamming techniques, search engines fail to publish their anti-spam techniques in the hope of making it more difficult for webpage spammers to succeed (Henzinger et al., 2002).

Some countermeasures developed against spam have been identified by Henzinger et al. (2002) as:
- search engines ignoring webpages with text in the same colour as the background, and
- search engines detecting certain distinct patterns of link-farms and doorway pages.

These spamming techniques caused search services to constantly turn to new ways of ranking webpages.

One author states: "The bottom line is - implementing ethical search engine marketing practices will deliver long term qualified traffic. Spam techniques may give you traffic today, but tomorrow your rankings will drop or your site will be black listed" (Robinson, 2005).

2.5.9 Anchor Text

Anchor text is the text usually displayed underlined and in blue by the web browser with the intent to describing a location of a specific hyperlink (Henzinger et al., 2002:5).

Notess (1999:85) gives the following example of the use of an anchor text:

> "a Webpage that both mentions the publisher Bowker and offers a link to Bowker's Website from the word "Bowker" has "Bowker" as the anchor text."

One advantage of using anchor text is that the more pages pointing to a specific webpage from the same anchor text, the higher the weighting of the target page on a query that matches the anchor text (Notess, 1999).

According to Henzinger et al. (2002:5), some search engines benefit from including anchor-text analysis due to the resemblance between the anchor description and the hyperlink. This resemblance make it difficult for spamming techniques, as it is could affect the perception of the user.

2.5.10 Graphics

Automated spiders can presently only read text formats, such as HTML files (example in Figure 13) and therefore are unable to read graphical elements. Some engines do consider text used in these ALT tags (text describing images or graphics) as keywords. Therefore ALT tags should be applied where possible to describe the graphics on the webpage (Hubbard, 2004). ALT tags must accurately describe the particular graphic.

Some problems that occur due to lack of ALT tags include:
- When a user turns graphics off, to increase loading speed, a site could become impossible to navigate due to lack of text describing the graphics.
- The use of ALT tags provides the ability to cater for those with a visual disability. Voice-output screen readers (benefiting the blind) will not "*read*" a non-text element (Oppenheim et al, 2000:204).

2.5.11 JavaScript

According to Goetsch (2003), search engine crawlers cannot interpret JavaScript and therefore cannot interpret the content it refers to. However, making use of the correct design techniques when making use of JavaScript, can still make a webpage "visible" to search engine crawlers.

An interviewee strongly disapproved of the use of JavaScript, Flash and session ID's in an interview conducted by Dickson and Marshall (2004:28-36). The argument was that a web crawler views a website as seen through a text browser. The author even goes further, by saying Google might look for URL's in JavaScript, but states that the average crawler would not be able to do so (Dickson and Marshall, 2004:28-36).

3.   Methodology

This pilot study was implemented by mimicking a real-life, e-commerce based website, basing the new design on the literature survey described in this study. This was done without disturbing the look and feel of the original site, as this was an express condition of the owners. A site was provided with the help of Cowie, author of www.sa-cycling.com, for the duration of this research project.

Figure 11 represents a screenshot of the appearance of the chosen website before any changes were made, while Figure 12 shows the webpage after the changes where implemented. The reader should not be able to detect any major differences between these two webpages, although their design approach differs in many aspects.

3.1   Headings

As discussed earlier (Sections 2.6.4 and 2.6.8), search engines rank keywords listed in header tags higher than those listed as normal text. The first focus was placed on the logo. The logo was design as a complete image which consisted out of the South African flag and the text "sa-cycling" as seen in Figure 11. The logo was changed to only an image of the South African flag, and the text "SA-Cycling" was then placed next to the logo in a header tag. The headlines and paragraph headers throughout the webpage was also replaced with header tags due to the higher weight given to header tags.

3.2   Title

Due to the weight given to the description of the title tag (as mentioned in Section 2.5.5), it was decided to change the title from "South African Cycling - http://SA-Cycling.com/" to "South Africa Cycling – news, fun rides, tours, clubs and much more", which now contained more relevant keywords accurately describing the contents of the site.

3.3   Meta keyword tags

Although meta keyword tags consist out of several categories, it was decided to only apply changes to the meta keyword tag and meta description tag, as these are the well known ones. A large variety of keywords, listed in figure 6, was derived from 15 other cycling websites and added to SA-Cycling's keyword section.

Figure 6: Compilation of keywords derived from several websites

```
<meta name="keywords" content= "road bikes, roadbikes, cycle, cycles, road
                                cycling, roadcycling, ride, SA cycling,
                                SA-Cycling, sacycling, team, cycle club,
                                cycling club, cycling, South Africa cycling,
                                cycle touring, cycle tour, WP cycling, best
                                cycling tours, fun ride, funrides, bicycle,
                                bicycling, rapport, giro del capo, giro,
                                cape argus, david cowie, bike touring, RIP,
                                hill, cycle news, cycling news, pedal power,
                                PPA, WPPA, track cycling, rules of cycling,
                                Tour de France, professional cycling, pro cycling,
                                procycling, track, fixed-gear, ride">
```

In the description tag, the frequently used words such as "or", "to", "not", etc. were minimised to increase the keyword prominence of the tag. As seen below, the text has been re-written so that it still reads well to the visitor but also positions the keyword phrase closer towards the beginning of a sentence or paragraph.

The description tag was changed as listed in figure 7 and 8.

ORIGINAL VERSION

Figure 7: Original description tag

```
<meta name="description" content="The South African Cycling website is provided
                                  to cover all topics of Cycling in South Africa:
                                  Touring, Important events, Racing, Pedal Power
                                  organisations, Local Provincial Bodies, National
                                  Federation, Mountain Biking and Club information">
```

NEW VERSION

Figure 8: Modified version of description tag

```
<meta name="description" content="South African Cycling website provide cover
                                  on all topics of Cycling in South Africa:
                                  Touring, Important events, Racing, Pedal Power
                                  Organisations, Local Provincial Bodies, National
                                  Federation, Mountain Biking, road deaths and
                                  club information">
```

3.4   Graphics

ALT tags were added to all graphics to accommodate blind people and people preferring to surf the internet with graphics turned off. To better explain the changes made to the graphics the author included the modifications (see figure 9) made to the logo graphic of the website:

Figure 9: Original Graphics tag

```
<img src="Pics/logo.gif" width="66" height="50" align="top"
                         name="SA-Cycling" alt="SA-Cycling Logo"><font size="+7">
```

The crawler can now interpret the name of the graphic (logo.gif), and read the ALT text that describes the image. It can also read the width and height of the graphic, but it can not interpret the graphic itself.

3.5   JavaScript

The success resulting from the removal of JavaScript is highly dependant on a web designer's personal point of view. JavaScript has just as many advantages as disadvantages. For the purpose of this research it was decided to remove the JavaScript from the sample webpage and redesign the menu to contain anchor links listed on a table (grid).   When comparing Figures 11 and 12, it is clear that the changes in the menu structure are barely visible to the average human browser.

3.6   Frames

The website provided contained several frames which could have caused several problems relating to low visibility and printing problems. It was decided to remove the frames completely and redesign the content as one html page (e.g. "SA-Cycling -

homepage.html" see Figure 12). This page contains all the code for the home page and separate pages for "news.html, contacts.html" etc. These changes are expected to make SA-Cycling much more visible to most crawlers and also provide an error free printing environment (Nobles and O'Neil, 2000).

3.7   Site Map

A site map (a webpage listing of all sections of a website) was created to increase usability as well as visibility (see Figure 10). According to Nobles and O'Neil (2000:94), many of the engines are allowing web developers to submit only the main page of a site. Therefore, one should ensure that the website's pages are found by the crawlers. With the use of JavaScript, several links could disappear in the code, but with the use of a site maps one could provide the links lost in JavaScript code. The site map not only provides the crawler with an easy navigation path throughout the website, but also assists in increasing the usability of the website to the user.

Another author claims that "Site maps are especially relevant since many users do not actually 'read' content, but instead skim the information, looking for items of relevance to their purpose" (Trunble, 2003:27).

A button was added to the menu bar which provided easy access to the site map for users and search engine spiders.

Figure 10: Site map

4.   Expected results

Throughout the implementation of the visibility features onto the webpage, the authors retained the layout as close to the original as possible. The only major perceived difference between the old site (see Figure 11) and the new improved site (see Figure 12) is the menu bar in the left margin. In designing the new menu, which consists out of

multiple anchor links in rows linking to appropriate pages, the authors kept the layout very close to the original. Each of the menu options can now easily be read by a crawler, allowing the menu options to be listed as keywords. More importantly, crawlers could now also follow the embedded hyperlinks to other pages on the same site.

The second noticeable change is the heading, which now starts with capital letters and is displayed in bold by using only the "header tag". The heading which now contains larger and bold text is expected to receive a higher ranking by search engine algorithms.

The third noticeable change is the title of the webpage. It has been changed to a more keyword-dense sentence that accurately describes the content of that particular webpage, rather that just the site name and address. It is expected that the new title should also put some emphasis on some keywords associated with the website.

The fourth and final noticeable change concerns the frames. The removal of the frames should eliminate the risk that some search engine crawlers would not be able to find keyword rich webpages on the site. It also eliminates the chance of a crawler indexing the outside of the frame as a distinct page, and in the process miss the actual page containing the body text. The printing problem is now also solved.

The ALT tags (text describing the graphics) are now visible when graphics are turned off. This improvement is expected to allow end-users who turned the graphics off to find their way around the site, by reading the graphic descriptions. At the same time the descriptive text should provide valuable keywords for the crawlers to interpret. See Figure 13 for an example of the value of ALT tags.

The newly added site map now provides anchor texts, providing several links to the rest of the site. This allows most search engine crawlers to read every page on the site. Not only should it increase the visibility to the crawler but it should also increase the usability of the webpage to human visitors. Figure 10 shows the design of the implemented site map.

Due to the ongoing nature of this research project future research could include the Implementation of the newly designed SA-Cycling website, submitting it to search enines and monitoring its ranking. The website could then also be analysed with WebPosition Gold's Traffic Analyzer which tracks visitors using an invisible counter on the website.

5.   Conclusion

It should be noted that there is no one quick way to dramatically increase the visibility of a website. It is a complex process which involves the preferences of the website's owners, technical issues and ethical aspects. Making a website visible for most search services is a time-consuming process, and it should be ongoing.

Enforced development standards are often achieved through the creation of a web development style guide, or a formalised set of policies and procedures to ensure consistency across the site.

Should these development standards be implemented,
- higher website visibility,
- increased crawler friendliness without sacrificing human friendliness,
- increased website's usability to disabled users by implementing text that describe the graphics, and

- increased usability made possible by the site map could be experienced.

Although the lack of JavaScript, Flash, frames and sounds might seem to decrease the human appeal of a website, it could be a step required for financial survival. Several changes as well as the removal of JavaScript where implemented without major changes to the webpage layout as seen in Figures 8 and 11.

6.   References

Adams, I. 2003. Overcoming Webpage Printing Problems. *Online Magazine*, Issue September / October 2003, 36-41.

Baard, V.C. 2004. Interactive information consulting system for South African small businesses – Part 1. *South African Journal of Information Management*. 6(2). [Online]. Available WWW: http://www.sajim.co.za (Accessed 5 December 2004).

Barron, C. 2000. *Brilliant ideas but spectacular flops*. Sunday Times 9 April. 9: (1).

Bester, F. 2003. Challenges and risks for South Africa in the information economy. *South African Journal of Information Management*. 5(2). [Online]. Available WWW: http://www.sajim.co.za (Accessed 10 January 2005).

Blair, D.C. 1990. *Language and Representation in Information Retrieval*. Amsterdam: Elsevier Science Publishers.

Boyes, J.A & Irani, Z. 2004. An analysis of the barriers and problems to web infrastructure development experienced by small businesses. *Information Technology and Management*. 11(2):189-207.

Constantinides, E. 2002. From Physical Marketing to Web Marketing: The Web-Marketing Mix. In: *Proceedings of the 35th Hawaii International Conference on System Science* – 2002. Big Island, Hawaii, 7-10.

Craven, T.C. 2001. Changes in Metatag Descriptions Over Time. *First Monday*. 6(10), October 2001.

Craven, T.C. 2002. What is the title of a Webpage? A study of Webography practice. *Information Research,* 7(3). [Online]. Available WWW: http://InformationR.net/ir/7-3/paper130.html (Accessed 30 January 2005).

Craven, T.C. 2003. HTML Tags as Extraction Cues for Webpage Description Construction. *Information Science Journal*. 6():1-12.

Department of Trade and Industry. 2001. *State of small business development in South Africa: Annual review 2001*. Pretoria: Ntsika Enterprise Promotion Agency.

Dickson, M & Marshall, G. 2004. Search engine secrets. *.Net.* Issue May. 28-35.

Duncan, V. and Fichter, D.M. 2004. What words and where? Applying usability testing techniques to name a new live reference service. *Medical Library Association,* 92(2):218-225.

Fillis, I, Johannson, U. & Wagner, B. 2004. Factors impacting on e-business adoption and development in the smaller firm. *International Journal of Entrepreneurial Behaviour & Research,* 10(3):178-191.

Galon, D. 1999. *The savvy way to successful website promotion: Secrets of successful Websites, Attracting on-line traffic. The most up to date guide to top positioning on search engines.* Canada:Trafford Publishing.

Goetsch, D. Search engine optimization. Search engine robots: how they work, part one. [Online]. Available WWW: http://www.rankforsales.com/crawlers-and-spiders-and-robots.html (Accessed 13 May 2004).

Green, D. 2000. The evolution of Web searching. *Online Information Review*, 24(2):124-137.

Guenther, K. 2004a. Getting your Website Recognized. *Online Magazine*, Issue May/June. 47-49.

Guenther, K. 2004b. Know the fundamentals and good design would follow. *Online Magazine*, Issue January/February. 54-56.

Henzinger, M.R., Motwani, R., Silverstein, C. 2002. Challenges in web search engines. [Online]. Available WWW: http://www.acm.org/sigs/sigir/forum/F2002/henzinger.pdf (Accessed 12 April 2005).

Hoffman, E.C. 2002 Information and communications technology, virtual offices and telework. *South Africa Journal of Information Management*, 4(2). [Online]. Available WWW: http://www.sajim.co.za (Accessed 17 january 2005).

Hubbard, J. 2004. *Indexing the Internet*. Essay for Drexel University College of Information Science and Technology.

Isaacs, Z. 2004. Good IT strategy helps SMMEs get it right. CITInews. [Online]. Available WWW: http://www.CITInews.com (Accessed 16 April 2004).

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

Nobles, R & O'Neil S. 2000. *Maximize website Traffic: Build website Traffic Fast and Free by Optimizing Search Engine Placement*. USA: Adams Media Corporation.

Notess, G.R. 1999. On the Net: Rising Relevance in Search Engines. *Online Magazine.* Issue May-June 1999. 84-86.

Oppenheim, C., Morris, A. & McKnight, C. 2000. The Evaluation of WWW search engines. *Journal of Documentation,* 56(2):190-211.

Robinson, P. 2005. Stop Search Engine Spam. ClickThrough Marketing. [Online]. Available WWW: http://www.clickthrough-marketing.com/articles/stop-search-engine-spam.php (Accessed 17 April 2005).

Rowley, J. 2001. Remodelling marketing communications in an Internet Environment. *Internet Research: Electronic Networking Applications and Policy,* 11(3):203-212.

Singh, A.M. 2002. The Internet – strategies for optimal utilization in South Africa. *South African Journal of Information Management*, 4(1). [Online]. Available WWW: http://www.sajim.co.za (Accessed 15 February 2005).

Strinivasan, P. Ruiz, M. Lam, W. 1996. An investigation of indexing on the WWW. In: *Proceedings of the ASIS '96*, (33):79-83.

Subia Creative. 2002. Justifying an Investment in Search Engine Marketing. [Online]. Available WWW: http://www.subiainteractive.com/art-pref-leads.html (Accessed 20 May 2004).

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

Thurow, S. 2003. *Search Engine Visibility*. New Riders Publishing. Indianapolis.

Trunble, I. 2003. Is your website a usable tool, or a designer's dalliance? *inCite.* August. [Online]. Available WWW: http://www.alia.org.au/publishing/incite/2003/08/usable.tool.html (Accessed 24 June 2004).

Weideman, M. 2004. Ethical issues on content distribution to digital consumers via paid placements as opposed to website visibility in search engine results. In: *Proceedings of ETHICOMP 2004*, April 2004. Syros: University of the Aegean: (2):904-915.

Acknowledgements

7.   Appendix A

Figure 11: Site before changes were implemented

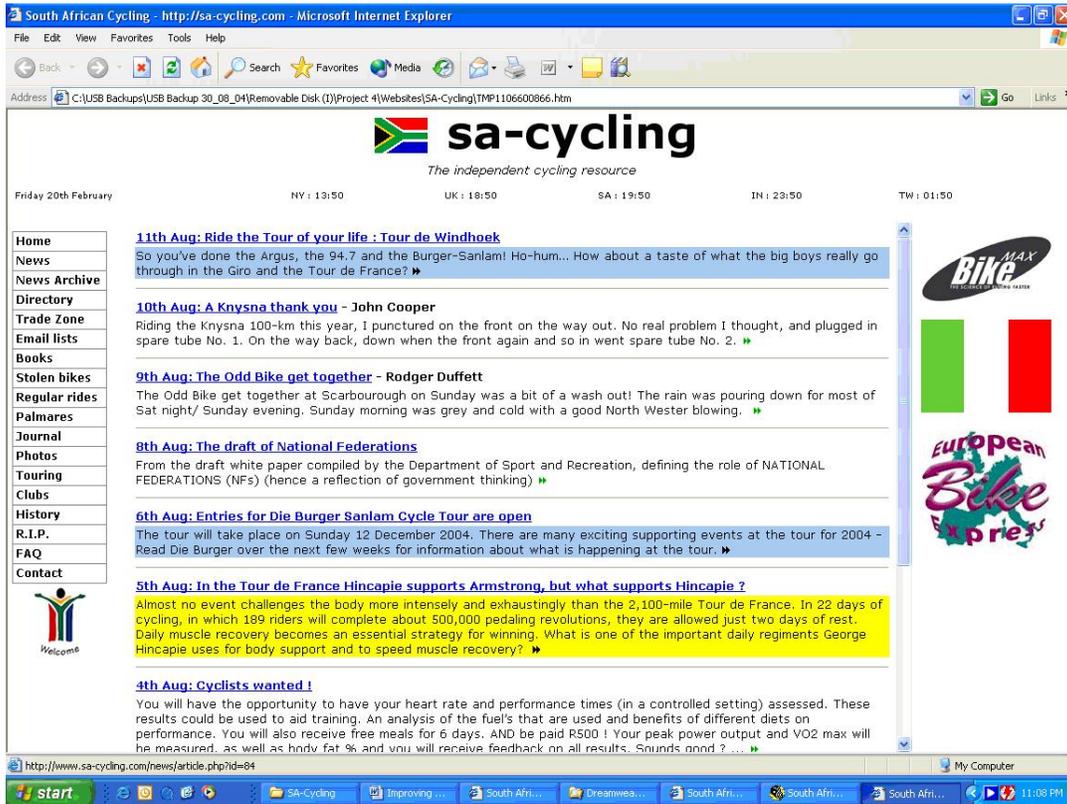**Figure 12: Site after changes were implemented**

South Africa Cycling - news, fun rides, and much more - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

file:///C:/SA-Cycling/SA-Cycling.html

🏴 SA-Cycling

The independent cycling resource

Monday 17th May   NY : 14:30   UK : 19:30   SA : 20:30   IN : 00:30   TW : 02:30

Home
News
News Archive
Directory
Trade Zone
Email Lists
Books
Stolen Bikes
Regular Rides
Palmares
Journal
Photos
Touring
Clubs
History
Links
R.I.P
FAQ
Site Map
Contacts

11th Aug: Ride the Tour of your life : Tour de Windhoek

So you've done the Argus, the 94.7 and the Burger-Sanlam! Ho-hum... How about a taste of what the big boys really go through in the Giro and the Tour de France? ☐

10th Aug: A Knysna thank you - John Cooper

Riding the Knysna 100-km this year, I punctured on the front on the way out. No real problem I thought, and plugged in spare tube No. 1. On the way back, down when the front again and so in went spare tube No. 2. ☐

9th Aug: The Odd Bike get together - Rodger Duffett

The Odd Bike get together at Scarbourough on Sunday was a bit of a wash out! The rain was pouring down for most of Sat night/ Sunday evening. Sunday morning was grey and cold with a good North Wester blowing. ☐

8th Aug: The draft of National Federations

From the draft white paper compiled by the Department of Sport and Recreation, defining the role of NATIONAL FEDERATIONS (NFs) (hence a reflection of government thinking) ☐

6th Aug: Entries for Die Burger Sanlam Cycle Tour are open

The tour will take place on Sunday 12 December 2004. There are many exciting supporting events at the tour for 2004 - Read Die Burger over the next few weeks for information about what is happening at the tour. ☐

5th Aug: In the Tour de France Hincapie supports Armstrong, but what supports Hincapie ?

Almost no event challenges the body more intensely and exhaustingly than the 2,100-mile Tour de France. In 22 days of cycling, in which 189 riders will complete about 500,000 pedaling revolutions, they are allowed just two days of rest. Daily muscle recovery becomes an essential strategy for winning. What is one of the important daily regiments George Hincapie uses for body support and to speed muscle recovery? ☐

4th Aug: Cyclists wanted !

You will have the opportunity to have your heart rate and performance times (in a controlled setting) assessed. These results could be used to aid training. An analysis of the fuel's that are used and benefits of different diets on performance. You will also receive free meals for 6 days. AND be paid R500 ! Your peak power output and VO2 max will be measured, as well as body fat % and you will receive feedback on all results. Sounds good ? ... ☐

Welcome

Bike MAX
THE SCIENCE OF RIDING FASTER

European Bike Express

Done

start   Websit...   SA-Cycl...   South A...   WebPo...   ACDSe...   South A...   Homep...   09:49 AM

Figure 13: Webpage loaded without graphics