VISIBILITY TO SEARCH ENGINES: A COMPARISON BETWEEN TEXT-BASED AND GRAPHICS-BASED HYPERLINKS ON E-COMMERCE WEBSITES

Ngindana, M.W., Weideman, M.
Faculty of Business Informatics, Cape Technikon
mongezim@ctech.ac.za, melius@gmail.com

ABSTRACT

A literature search has indicated that most website developers and designers first build a website and later concern themselves about “searchability” and “visibility”. Some companies spend large amounts of money on a website which cannot be indexed by search engines, is rejected by directory editors and is invisible to search engines. The primary objective of this research project is to compare and report on the effect on website visibility of text-based versus graphic-based web page navigation schemes.

The method employed in this project will be to develop two e-Commerce based website with the same functionality, contents and keywords, but different navigation schemes. One will embed all hyperlinks in text-phrases only, while the other will use graphic images of buttons, arrows and other navigational aids for hypertext links. Both websites will be submitted to the same search engines at the same time. A period of two months will be allowed to ensure that spiders will have had enough time to visit and index both sites. An industry standard web ranking program as well as advanced counters will be used to monitor how the two sites feature in the rankings, and if they appear in the indices of search engines. Graphs and other results as well as text-based report produced by the ranking program and the counters will be used to compare results.

Based on a literature survey, it is expected that the text-based website will achieve higher rankings than the graphics based one. The result will provide a clear path to commercial and other website designers in terms of navigational element design. These result will provide a foundation to the construction of a best practice guide for commercial website designer. Although the human website browser finds a certain amount of graphical aids conducive to easier navigation, search engine crawlers find these same graphic aids impossible to index. A balance has to be found between these two extreme to please both the human and the crawler. It is believed that this research project will provide website designers with guidance on achieving that balance.

KEYWORDS
search engines, hyperlinks, text, graphics, visibility, navigation, design
1. INTRODUCTION
As the web expands into a vast sea of information, users increasingly turn to search engines as their fundamental navigation aid. Search engines are forced to select from thousands or millions of possible web pages when answering queries. Ultimately, all search engines in one way or another favour websites that are “search friendly” (Thurow, 2002).

According to Weideman (2003), website navigation schemes are aimed at pleasing either human site visitors or search engine crawlers. It was found that some of the developers design web pages that are created entirely in Flash or with graphic images to preserve the aesthetics of the colours, uncommon typefaces and movement to satisfy users. In the process, these pages end up being search engine unfriendly (Lide et al, 2001).

1.1 Background to the research problem
Several authors (Lu, 1998; Pfattenberger, 2001; Murata, 2003) concur that if a company is going to use search engine marketing as part of its online strategy, creating search-engine friendly design templates can save a company considerable time and expense, and increase visibility.

Thurow (2002) reported that:

“The strategy of placing keyword-rich text in your web pages is useless if the search engines spiders have no way of finding that text. “

This author also identified the following site architecture issues that are problematic to search engine spiders crawling the site:
- poor HTML coding,
- image maps,
- frames,
- Flash movies,
- JavaScript, and
- dynamic or database-driven web pages.

Link popularity is one part of the link analysis systems that search engines such as Google use to rank pages (Beatty, 2003). Eastman et al (2003) identified the following factors as also having an effect on visibility:
- keyword density,
- keyword placement,
- keyword prominence, and
- link popularity.

The use of a text-based link navigation scheme might help developers to design websites that could be more easily indexed by search engine crawlers. Although navigation schemes are put to use by many developers, its effectiveness have not been evaluated. In this study the value of text-based links as opposed to graphic-based links towards the visibility of websites to search engine crawlers is examined.
1.2 Research problem
Most web developers and designers first build a site and only later concern themselves with “searchability” and “visibility” (Sullivan, 2001). Munafo (2002) found that companies spend thousand of dollars on sites which:
- Cannot be indexed by search engines,
- Are rejected by directory editors,
- Are not visible to search engines.

For the purpose of this study, the focus will be on the use of text-links versus graphic-links for e-commerce website navigation scheme, and comparison will be made of their impact on website ranking, visibility, and positioning to search engines.

1.3 Research Question
If a company is going to use search engines as part of its online marketing strategy, creating search engine friendly design templates can save a company considerably time and expense (Thelwall, 2002a).

In the study, the following question relating to the use of text versus graphic website navigation schemes is examined:
- Will the use of a text-based link navigation scheme as opposed to a graphic-based link scheme improve a websites’ visibility to search engines?

The author is testing whether the use of text as navigation scheme has a positive impact on websites’ visibility to search engine crawlers. The study is designed to determine how useful these methods of linking are in improving visibility and ranking of website’s via Internet search engines.

2. LITERATURE STUDY
According to Ford et.al (1997), consumers are steadily moving to electronic commerce (e-Commerce) in the USA in order to pay their bills, purchase goods or use the many other electronic services available. A recent study by Bruemmer (2002) estimated a growth in electronic commerce of 20% every two months. In the USA it has been found that search engines are the source of choice for Internet users to find websites. This could explain why a website with good search engine listings may see a dramatic increase in traffic (Furnis, 2001).

Collin (2002) identified the following as the most common mistakes made by developers of e-commerce websites:
- The use of Microsoft FrontPage and Frontpage Express (may result in unacceptable code),
- novice designers seem to lack knowledge of graphic layout, and
- a lack of graphic optimisation.

This may result in sites that have large graphic headers (70kb – 100kb), and is time consuming to download.
2.1 Search Engines
Thelwall (2002b) claims that, in order to understand how to improve rankings, one needs to have a basic concept of how a search engine works. There are essentially two types of search engines – those that make use of directory listings and those that use indexing. What differentiates a search engine from a directory is that a directory database consists of sites that have been added by human editors. Search engines’ databases are compiled through the use of special software robots, called spiders, to retrieve information from web pages (Mertz, 2001).

According to other literature there are three basic types of search services, namely:
- Crawler-based search engines: they create their listing automatically. Search engines “crawl” or “spider” sites looking for identifying tags that will determine how and when the results will appear (Rose et al. 1999).
- Human-Powered directories: employs cadres of researchers to look at millions of sites and categorize them according to what the researcher deems appropriate. Rao (2000) concludes that strategies for being listed well in search engines differ from those being listed well in directories.
- Hybrid search engines or mixed result: in the web’s early days, search engines either presented crawler-based results or human-powered listings, today, it is extremely common for both types of results to be presented (Li et al. 2002).

Dykehouse et al (2000) reported that if two websites have the same text and link components, the site that end-users click the most will usually rank higher. Sometimes, a popular website will consistently rank higher than sites that use many keywords.

Therefore, building a site that appeals to human directory editors, crawlers and your target audience appears to be important for maximum search engine visibility.

2.1.1 Search Engines Visibility importance
Search engines face a daunting task, forced to select from millions of possible webpages when answering queries (Nobles et al. 2002).

Alexandra (2002a) reported that all search engines in one way or another favour websites that are search-friendly. Whalen (2000) states that, “search engines are rated on five subjective categories”, namely:
- interface: view of engines’ home pages and how effectively it displays content,
- relevancy: how accurate the results are within the scope of a search,
- usability: how effectively one can navigate and find things,
- site owners’ respect quotient: how effectively the engine deals with the people who built their database, and
- link popularity: the quality of sites linked to the website.

2.1.2 Factors that affect Website visibility
According to Beaty (2003), factors affecting visibility include:
- keyword density, i.e. the ratio of keywords to the total number of words on a page,
- keyword placement, i.e. the positioning of keywords in the title, other tags and in alternative areas, such as in image description,
• keyword prominence, i.e. how high a keyword appears on a page. And
• link popularity, i.e. how the site is linked to other sites.

Apart from these (mostly) keyword oriented factors, a number of others exist, a discussion of which falls outside the scope of this paper.

2.2 Navigation Schemes
Website designers should consider their target audience when creating a site’s navigation scheme. Navigation schemes should enable visitors to find what they are searching for as quickly and as easily as possible. A recent study by Eastman & Jansen (2003) reported that a search engine design and user-friendly design have at least two forms of navigation: one that site visitors can follow and one that search engines can follow.

Alexandra (2002b) claims that some websites use multiple navigation schemes. For example, a site might have a series of navigation buttons down the left side of a screen and corresponding text links at the bottom of a screen. Text links can also be placed within the actual body content. Common navigation schemes includes the following:

• hypertext links,
• navigation buttons,
• image maps,
• drop-down/pull-down menus,
• animation/Flash button and
• dynamically generated URLs.

Each one of the navigation schemes has advantages and disadvantages which impacts the search engine visibility of the site as will be discussed below.

On figure 2.2 Amazon.com is using image, button and text as major navigation scheme to navigate the site.

![Figure 2.2. Image maps, Image Buttons and Text-links used as Navigation Scheme](image)

2.2.1 Hypertext Links
Some authors claim that search engine spiders easily interpret text links because they can record the text in and around the link and follow these links from web page to web page (Gardner et al., 2001). In fact, search engine marketers use the term "anchor text" to refer to the HTML text inside a hyperlink.

Usability experts also prefer to use text links and breadcrumbs (Tansley, 2002), because they provide the target audience with valuable information about the trail of visited and unvisited pages behind them. Users automatically know that a blue underlined word or just an underlined word indicates an unvisited link and that a
purple or faded colour indicates a visited link.

Breadcrumbs as shown on figure 2.2.1 are commonly used at the top of a webpage and are hierarchical in nature. Inan (2002) concluded that many directories, including Yahoo!, use breadcrumbs as a navigation scheme. Because breadcrumbs are generally placed at the top of web pages, search engines consider the text placed inside the breadcrumb links important. Therefore, if a site uses breadcrumbs as a navigation scheme, they must use keywords consistently in them (Lide & David, 2001).

Figure 2.2.1 Text-links used as a Navigation Scheme

2.2.1.1 Absolute and Relative Links
According to Whalen (2002) two types of links exist: absolute and relative. Whalen defines an absolute link as a specific location of a web file or document, and a relative link as a link to another page relative to the current one. The difference is that the coder has to specify only the filename of a relative link instead of the full URL.

Gardner et al. (2001) note that it makes no different to the search engine crawlers whether you use an absolute link or a relative link - all search engine crawlers can follow both types of links.

2.2.1.2 Places to use Text Links
Winn et al (2002) concluded that:

“Because all search engines can index the text and follow the links in a hypertext link, for maximum search engine visibility, use them as either the primary or secondary navigation for a website design”.

Beatty (2003) defines the following five general locations for placing text links:
The top of a web page - not always the best place to put text link, because this the first place the crawler reads. As long as the text is short however, it should not interfere with search engine visibility.

The left side of a web page - not the best location to put text links. Search engines often use this text in their snippet in the search results, and if the text is too long, it could cause the web page to appear not to be focused on a specific topic.

The bottom of a web page is the preferred location for text links. These text links can correspond to graphic images (navigation buttons or image maps) at the top of the screen.

The right side of a web page is a good place to put text links because it enables the designer to cross-link important web pages above the fold.

Placing text inks in the middle of a web page highlights the important points in the document, and should also be encouraged.
2.2.1.3 Problems with Hypertext links
A recent study reveals that a text link navigation scheme might seem like the ideal solution because all search engines prefer this type of link (Thurow, 2002). It follows that a web page full of text links tends to download faster than a web page full of graphic images. It seems that the use of text links as main navigation scheme can satisfy both the search engine and targeted user.

It is reported that too many text link can interfere with keyword density, since they can dilute the keyword density and tamper with keyword prominence (Ollins, 2002). According to Tansley (2002):

“... placing breadcrumb links at the top of the page can retain keyword prominence. If you find that you have more words in your text links than you do in the main body content, it is better to use graphic images as part of the navigation scheme.”

Xue et al. (2000) found that text links are often the first body of text that the search engines read. Murrata (2003) claims that too many text links on a single page can also interfere with a page’s legibility. Because this text tends to be the same on many web pages, the first text introduced to the spiders is not unique. If a search engine does not use the meta-tag description to display in the search result, it generally takes the first text found at the top of a web page. A series of text links does not accurately describe the contents of a unique web page. The description can appear as a collection of unrelated words to users.

2.2.2 Graphic Links
Murrata, (2003) states:

“The whole point of writing a web page is to have a target audience read it and perform a desired action. People like simplicity and ease of navigation. Thus, find ways to make your text link more visually distinct, easy to find and legible, such as placing them in a coloured table cell or a coloured side bar”.

A navigation button is a graphic image, generally in a GIF or a JPEG format, that links to a single URL.

2.2.2.1 Navigation Buttons
Navigation buttons as shown in figure 2.2.2.1 give users a visual representation of how to navigate a site. They are visually appealing, and can easily draw attention to important parts of a webpage. Users’ eyes are naturally drawn to a splash of colour or a change in dimension, for example web designers use contrasting colours to highlight the ON and OFF button on a webpage. Xue et al. (2000) recommends that navigation button should always contain alternative text in the HTML code. This will allow visitors to click that text to navigate site if the button image does not load.

Figure 2.2.2.1 Navigation button
2.2.2.2 Problems with Navigation Buttons

All search engine crawlers can follow the link surrounding a navigation button (Furnis, 2001), as long as the navigation button does not contain JavaScript within the anchor tag. JavaScript can pose problems to search engines crawlers, therefore not all search engines follow this type of link.

Furnis also reported that when too many attributes are added to the anchor tag, such as rollover script, the attributes could make links less search engine friendly. Currently, the major search engines prefer link coding without any type of script. According to Ollins (2002), one effective search-engine friendly layout is to use navigation buttons (with or without JavaScript rollover) and corresponding hypertext links at the bottom of a web page.

2.2.2.3 Image Maps

An image map is a single graphic image that enables users to access different web pages by clicking on different areas of the image. It is reported that (Thurow, 2002) many search engines do not follow the links inside an image map, because of the possibility of image spam. The study further concludes that, if image maps are used as part of the navigation scheme, the designer should use text links or navigation buttons elsewhere on the web page.

![Image Map as Navigation Scheme](image.png)

Figure 2.2.2.3 Image map as Navigation scheme

Mertz (2001) finds that a single image downloads much more quickly than multiple graphic images, for example, if a site has 16 navigation buttons that are 2kb each in file size, the total for these graphic images is 32kb. A single image map might only be 8-10kb in size, much smaller than the set of navigation buttons. With the image map there is only one call to the server as opposed to 16 in the case of the navigation buttons.

Mertz (2001) concluded that search engines crawlers have an easier time accessing the main body text on a web page if the navigation scheme is a single graphic image rather than multiple ones.

2.2.2.4 Use of Image Maps

Aguillo (2000) notes:

“When is it not good idea to use an image map? If the only navigation scheme on a website is graphic images, it is best to use navigation buttons because all the search engines can follow the links surrounding navigation buttons.”
Aquillo’s findings are also supported by Xue (2003):

"However, if you are using both graphic images and text links as a navigation scheme, consider the pages’ download times”.

If the image map downloads more quickly than a set of navigation buttons, the image map might be a better choice.

2.2.2.5 Dropdown and Pull-down Menus
The main advantage of using drop-down menus in a navigation scheme is the saving of screen real estate (Holder, 2003). Drop-down menus do not initially take up as much screen space as a series of navigation buttons or text links. Furthermore, by freeing up screen space, website owners are able to place more contents that their target audience wishes to read above the fold.

2.2.2.6 Problems with Menus
According to Mertz (2001), drop-down menus are generally not search engine friendly because they require either JavaScript or a CGI program to work. Because search engines generally do not follow these types of links, you should always have an alternative form of navigation for the crawlers to follow.

Web designers can use a combination of drop-down menus with corresponding text links at the bottom of a web page for search engine visibility (Greenough et al. 2001).

2.2.2.7 Dynamically Generated URL’s
Many Internet crawlers do not index dynamically generated pages (Inan 2002). Dynamically generated web pages are usually created using a technology such as Active Serve Pages (.asp), Cold fusion (.cfm), Hypertext PreProcessor (.php), Java Server Pages (.jsp), or Perl. Often, search engine marketers state that search engines cannot follow the links inside dynamically generated web pages, and they might not prefer the URL used to retrieve web pages.

Dynamically generated URL’s contain some of the following characters: ?, &,%,$, =, and +. Crawlers cannot interpret URL’s containing these characters correctly, hence the information contained on these pages are effectively invisible to the search engine user.

Some dynamically generated URLs can "trap" a search engine crawler and cause it to crash: programmers who fail to close their IF and WHILE statements on webpages can set this "trap".

However, search engines are progressing in their capability to crawl dynamic URLs. Web developers should focus their efforts on delivering web pages to the search engines that do not contain stop characters in their URL’s.

2.3 Website design for Search Engine visibility
Hof et al., (2003) concur that good website design gives a site credibility. An effective website needs to download quickly. If a site does not appear on screen fast enough, competitors are just a click away (Ollins 2002). Furnis (2001) recommends that website design be kept very simple without frames, excessive graphics, database-generated pages, Flash home pages or lengthy JavaScript. Each of these
design strategies can present problems when trying to get a good ranking with search engines.

It has been reported that, to get the best search engine visibility, web designers should follow the Five Basic Rules of Web Design, which state that a website should be:

- easy to read,
- easy to navigate,
- easy to find,
- constant in layout and design and
- quick to download (Eastman et al., 2003).

It is believed that, by following these rules, you will be building a website to satisfy a targeted audience as well as directory editors and search engine crawlers (Rose et al. 1999).

2.3.1 e-Commerce Website architecture
According to Winn et al. (2002), the Internet is a collection of client/server computers and infrastructure that spans the earth. Millions of computers are indirectly connected to one another by routers over the Internet. In many cases (depending on the contract with the ISP), nearly unlimited numbers of files can be transmitted and received across the world for nothing more than the cost of connecting to the Internet.

The Internet acts as a network infrastructure for many types of computing beside the World Wide Web, but it is the key World Wide Web client/server architecture that makes it so germane to B2C e-commerce (Randall, 1999). As a result, a global client-server network of Web-ready computers provides the mechanism for low cost communication between a Web retailer and millions of e-consumers. Murata (2003) reported that, while the infrastructure enables commerce, this same infrastructure also contains technological limitations and obstacles to growth and further development.

2.3.2 Classification and evaluation of Websites
Tansley (2002) sees websites performing three main functions: the processing of a client’s request, the control of service levels and the interaction with remote network caches. A website application can be classified as commercial sites (e-tailing and auctions), financial sites (banking and security trading), information sites (news and sport) and educational sites (digital libraries). Web applications for sites should be able to handle the workload of a site as well as provide a certain level of performance.

2.3.3 Objectives of e-commerce Web Applications
Tansley (2002) states that the objectives of Web applications are information retrieval, dissemination and processing as well as technology exploration. The author claims that web development has seven stages: problem statement, requirement list, conceptual model, derived requirements, detailed analysis, detailed design and finally the development stage, when HTML and scripts are written.
Web applications have different objectives, which will determine the tools to be used by Internet programmers. It is therefore important for Internet programmers to identify these objectives. Rao (2002) identified the following types of objectives:

Business objectives - determine tasks, e.g. enhancing customer relationships, increasing employee productivity, etc.

Information objectives refer to the information resources that the application should provide, e.g. structured or unstructured; text or multimedia.

Functional objectives include, for example, the capabilities the application must have e.g. keyword searching, forms-based data collection, hypertext linking and traversal downloading capabilities as well as functionalities for scripts i.e. CGI scripts.

3. METHODOLOGY

Two virtually identical URL's namely e-books1.co.za and e-books2.co.za will be designed and linked to Amazon.com. The sites will be registered with the same search engine (Google) and both hosted on same server. The site will be monitored for a period of three months and data from the counter will be recorded. Statistics will be done based on the data and feedback from the user. Conclusion will be made based on the information analysed.

3.1 Method
The method employed in this project will be to develop two e-Commerce based websites with the same functionality, content and keywords, but different navigation schemes as shown in Figure 3.1 and 3.2 in appendix-A. One will embed all hyperlinks in text-phrases only, while the other will use graphic images of buttons and images. Both websites will be submitted to the same search engine at the same time, and hosted on the same server.

3.2 Data
A sample of 100 IT students will be handed questionnaires - refer to appendix B - to evaluate them in terms of user friendliness. A period of three month will be allowed for search engine spiders to be able to visit and index the two websites. To gather data on search engines, an industry standard web ranking program as well as counters will be used to monitor how the site feature in the rankings and if they appear in the indices of search engines.

3.3 Data Analysis
The results and data from end-user evaluation questionnaires will be tabulated and analysed using graphs and statistical methods. Data from the standard program, text-based reporting and the counters will be used to compare results.

The two websites above have the same look, except e-books1 is using text-based link figure 3.1 and e-books2 is using graphic-based link figure 3.2. The content of these two website and the functionality is the same, which is both sites are selling and buying books online.
4. CONCLUSION

From the literature reviewed, it is evident that website designers should exercise extreme care in the choice and application of a navigation scheme. The main problem identified is that the requirements for a human friendly scheme are in direct opposition to those of a crawler friendly scheme.

It appears as if most of the solutions to the problems generated by some navigational schemes proposed in the literature could create new problems of their own. Duplication of existing information, creating extra clutter and adding to download delays are some of these problems, which need to be researched further. Potential future research in this area could look at this duplication element, the irritation factor of Flash components and download delays.

Knowledge about and insight into e-commerce features is desirable to enable designers to enhance the visibility of webpages to search engines. Whether planning for e-commerce or following an existing strategy, it is important to know which features should be embedded and which should be omitted. Only after this has been determined, can decision be taken on which search services and what elements to add to the coding process to make a company’s website search engine friendly and visible.

Numerous opportunities for further research exist in this area.
Appendix A: E-Books Websites

Figure 3.1 E-Books1 - a pure text-based website with no graphics

Figure 3.2 E-Books2 - a graphic-based website with graphic-navigation scheme
Appendix B: Questionnaire

To complete this questionnaire you must first visit the two websites namely: E-Books1.co.za and E-Books2.co.za, the objectives are for you to compare the two sites. Please complete the questionnaire by placing an X in every relevant box.

Section A – Demographics Information

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>20 or younger</td>
<td>20 to 30</td>
</tr>
<tr>
<td>Field</td>
<td>Inf. Techn</td>
<td>Cost Mang.</td>
</tr>
<tr>
<td>Internet use</td>
<td>Frequently</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Website use</td>
<td>Frequently</td>
<td>Sometimes</td>
</tr>
</tbody>
</table>

Section B – E-Books1 evaluation

Rate the site (e-books1) using the scale of 1 = Excellent and 5 = Bad

<table>
<thead>
<tr>
<th>Read</th>
<th>Navigate</th>
<th>Find</th>
<th>Layout</th>
<th>Download</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Rate the site (e-books2) using the scale of 1 = Excellent and 5 = Bad

<table>
<thead>
<tr>
<th>Read</th>
<th>Navigate</th>
<th>Find</th>
<th>Layout</th>
<th>Download</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
5. REFERENCES


