

**VISIBILITY OF E-COMMERCE WEBSITES TO
SEARCH ENGINES: A COMPARISON BETWEEN
TEXT-BASED AND GRAPHIC-BASED HYPERLINKS**

by

Mongezi Ngindana

DISSERTATION

Submitted in partial fulfilment
of the requirements for the degree

MAGISTER TECHNOLOGIAE

in

INFORMATION TECHNOLOGY

in the

FACULTY OF BUSINESS INFORMATICS

at the

CAPE PENINSULA UNIVERSITY OF TECHNOLOGY

Supervisor: Prof M Weideman

January 2006

DECLARATION

I declare that 'Visibility of e-commerce websites to search engines: A comparison between text-based and graphic-based hyperlinks' is my work, that it has not been submitted before for any degree or assessment in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by means of complete references.

Furthermore, this dissertation contains no references to other work without acknowledgement.

Student Name: *M Ngindana*

Signature:

Date:.....

ACKNOWLEDGEMENT

The author would like to thank the following:

1. 'Qamata Wezinyanya, Ama-ndungwana, Ama-gcina', Almighty Jah Rastafari's, The Creator of Heavens and Earth, for giving me the strength, perseverance, and ability to get to where I am.
2. My supervisor Prof Weideman, who has provided excellent guidance towards completion of this dissertation.
3. The Cape Peninsula University of Technology for study leave.
4. The Cape Peninsula University of Technology librarians, for assisting with the searching of information, and the research resources that they provided in aiding the completion of this dissertation.
5. The National Research Foundation for financial assistance towards this research.

"Never be a coward, and yet never be too bold. Never put all your trust in man. Always leave room for surprises and disappointments. Never feel sorry for yourself, if you fall, stand up and start afresh."

(Lawrence Tutu)

ABSTRACT

Research has shown that most website developers first build a website and only later focus on the 'searchability' and 'visibility' of the website. Companies spend large amounts of money on the development of a website which sadly cannot be indexed by search engines, is rejected by directory editors and which is furthermore invisible to crawlers. The primary objective of this dissertation is to compare and report on the impact of text-based versus graphic-based hyperlinks on website visibility.

The method employed in the research was to develop two e-Commerce based websites with the same functionality, contents and keywords, however utilising different navigation schemes. The one website had all hyperlinks coded in text-phrases, while the other embedded the hyperlinks in graphics. Both websites were submitted to the same search engines at the same time. A period of eight months was allowed to ensure that the websites drew sufficient 'hits' to enable a comparative analysis to be conducted. Two industry standard website ranking programs were used to monitor how the two websites feature in the search engine rankings. Graphs as well as text-based reports produced by the ranking programs and the t-test were used to compare and analyse the results.

Evidence based on the reviewed literature indicated that there are conflicting reports on the impact of text as opposed to graphic hyperlinks on website visibility. However, there is unsupported evidence that text hyperlinks achieved higher rankings than graphics-based hyperlinks. Although the 'human website browsers' find a certain amount of graphical aids conducive to easier navigation, 'search engine crawlers' find many of these same graphic aids impossible to index. The study supported that the graphic-based website ranked higher than the text-based website, which calls for a balance to be found between these two extremes. This balance would satisfy both 'human website browsers' and 'search engine crawlers'.

It is posited by this author that this dissertation provides website designers with the abilities to achieve such a balance.

KEYWORDS:

search engines, hyperlinks, text, graphics, visibility, navigation, e-commerce, design.

RESEARCH OUTPUTS

During the formulation of this dissertation, the author produced the following research outputs, all related to the research topic:

RESEARCH OUTPUTS				
Output Type	Authors	Title	Conference/ Journal	Status
Journal Article	Weideman, M. & Ngindana, M.	An empirical study contrasting the role of two navigational aids in website visibility	South African Journal of Information Management (SAJIM)	To be submitted in February 2006 to www.sajim.co.za
Conference (Full Paper)	Weideman, M. & Ngindana, M.	Visibility to search engines: A comparison between text-based and graphic-based hyperlinks on e-commerce websites	6 th Annual Conference on World Wide Web Applications	Published in: Proceedings of WWW2004, www.zaw3.co.za , 1-3 September 2004
Conference (Short Paper)	Ngindana, M. & Weideman, M.	Website navigation architectures and their effect on website visibility – A literature survey	South African Institute of Computer Scientists and Information Technologists (SAICSIT)	Published in: Proceedings of SAICSIT, 4-6 October, 2004

Supervisor: *Prof. M. Weideman*

Signature:.....

Date:.....

TABLE OF CONTENTS

CHAPTER ONE- INTRODUCTION

1.1 INTRODUCTION	1
1.2 BACKGROUND TO THE RESEARCH PROBLEM	2
1.3 STATEMENT OF THE RESEARCH PROBLEM	3
1.4 SIGNIFICANCE OF THE STUDY	3
1.5 RESEARCH HYPOTHESES	5
1.6 AIMS OF STUDY	5
1.7 DELINEATION OF STUDY	6
1.8 OVERVIEW OF CHAPTERS	7
1.9 CONCLUSION	7

CHAPTER TWO - LITERATURE REVIEW

2.1 INTRODUCTION	8
2.2 INTERNAL LINKING STRUCTURE ELEMENTS	10
2.2.1 Graphic components	10
2.2.2 Text components	11
2.2.3 Popularity components	12
2.2.4 Alternate text components	13
2.3 NAVIGATION SCHEMES	14
2.3.1 Hypertext links	15
2.3.1.1 Absolute and relative links	17
2.3.1.2 Places to use text links	17
2.3.1.3 Search engines and hypertext links	18
2.3.2 Graphic links	19
2.3.2.1 Navigation buttons	19
2.3.2.2 Image maps	20
2.3.2.3 Use of image maps	21
2.3.2.4 Search engines and navigation buttons	21
2.3.3 Dropdown and pulldown menus	22
2.3.3.1 Search engines and dropdown menus	22
2.3.4 Dynamically generated URL's	22

2.3.4.1 Dynamic URL's	23
2.3.4.2 Importance of Dynamically generated URL's	23
2.3.5 Cascading Style Sheets (CSS)	25
2.3.6 Frames, Flash, DHTML, JavaScript rollovers	25
2.3.6.1 Frames and search engine visibility	26
2.3.7 Flash components and search engine visibility	27
2.4 SEARCH ENGINES	27
2.4.1 Differences between crawlers and web directories	29
2.4.2 Directory search engine	31
2.4.3 Crawler/Spider search engine	32
2.4.4 Hybrid search engines or mixed results	33
2.4.5 How search engines, Directories, and Hybrids work	34
2.5 SEARCH ENGINES VISIBILITY IMPORTANCE	35
2.5.1 Search engine friendly e-commerce websites	36
2.5.2 Search engine optimization strategies	41
2.6 WEBSITE DESIGN FOR SEARCH ENGINE VISIBILITY	43
2.6.1 Factors that affect website visibility	44
2.7 CONCLUSION	45

CHAPTER THREE - METHODOLOGY

3.1 DISCRPTION AND MOTIVATION OF THE RESEARCH DESIGN	48
3.2 RESEARCH DESIGN AND METHODOLOGY	49
3.2.1 Procedure	50
3.2.2 Population and sample types	51
3.2.3 Sampling methods	51
3.2.4 Data collection methods	52
3.2.4.1 Type of data	52
3.2.4.2 Source of data	52
3.2.4.3 Level of measurements	54
3.3 CONCLUSION	55

CHAPTER FOUR - RESULTS AND ANALYSIS

4.1	INTRODUCTION	57
4.2	DATA COLLECTED	57
4.2.1	Data collected from AWStats	57
4.2.2	Data collected from Internet Visibility Service™	61
4.3	RESULTS ANALYSIS AND INTERPRETATION	63
4.4	DESCRIPTIVE STATISTICS	66
4.5	CONCLUSION	70

CHAPTER FIVE - CONCLUSION AND RECOMMENDATIONS

5.1	INTRODUCTION	71
5.2	PROPOSED SOLUTION	72
5.2.1	Hypertext	72
5.2.2	Navigation Buttons	73
5.2.3	Image maps navigation	73
5.2.4	Flash and Animation navigation scheme	74
5.2.5	JavaScript or DHTML Dropdown or Pulldown Menus	74
5.2.6	Framed navigation	74
5.3	CONCLUSION	75
5.4	FURTHER RESEARCH	76

REFERENCES

LIST OF REFERENCES	77
--------------------	----

APPENDICES

Appendix A:	Website layout: Graphic-based website	87
Appendix B:	Website layout: Text-based website	87
Appendix C:	Program code: Graphic-based website	88
Appendix D:	Program code: Text-based website	92
Appendix E:	Visibility report for Graphic-based website from 'Real World Technology'	95

Appendix F:	Visibility report for Text-based website From 'Real World Technology'	98
Appendix G:	Statistics report for Graphic-based website from 'AWStats'	101
Appendix H:	Statistics report for Text-based website from 'AWStats'	103
Appendix I:	Descriptive statistics for mean, mode and standard deviation	105

GLOSSARY

GLOSSARY	111
----------	-----

TABLES

Table 1.1:	Internet hosts in Africa for 1998	4
Table 2.1:	Crawler vs. directory based search engines	29
Table 2.2:	Six basic functions of SEO campaign	42
Table 3.1:	A classification framework of design types	48
Table 4.1:	Monthly statistics for graphic-based website	58
Table 4.2:	Monthly statistics for text-based website	59
Table 4.3:	Consolidated monthly statistics for graphic and text-based websites	59
Table 4.4:	Search Engine visits statistics for graphic-based website	60
Table 4.5:	Search Engine visits statistics for text-based website	60
Table 4.6:	Consolidated Search Engine statistics for graphic and text- based websites	60
Table 4.7:	Comparison for the most active month (January)	61
Table 4.8:	Comparison of the monthly average	61
Table 4.9:	Internet Visibility Service™ ranking for graphic-based website	62
Table 4.10:	Internet Visibility Service™ ranking for text-based website	62
Table 4.11:	Combined data for text versus graphics websites	63
Table 4.12:	Consolidated comparison of text vs. graphic websites	67
Table 4.13:	Descriptive statistics for mean, median and std. deviation	69
Table 4.14:	Independent sample t-test	70

Table G.1: Website statistics report for graphic-based website	101
Table G.2: Quick summary of graphic-based website	102
Table H.1: Website statistics for text-based website	103
Table H.2: Quick summary of text-based website	104

GRAPHS

Graph 4.1: Consolidated monthly statistics for graphic vs text websites	64
Graph 4.2: Comparison of search engine statistics	64
Graph 4.3: Internet Visibility Service™ analysis of graphic versus text-based websites	65
Graph 4.4: Overall visibility to search engines	66
Graph 4.5: Consolidated comparison of number of requests	68
Graph 4.6: Consolidated comparison of number of page requests	68
Graph 4.7: Consolidated comparison of number of bytes transferred	69
Graph E.1: Graphic-based website statistics	96
Graph F.1: Text-based website statistics	99
Graph G.1: Monthly report of graphic-based website	102
Graph H.1: Monthly report of text-based website	105

FIGURES

Figure 2.1: Images, buttons and text-links used as navigation scheme	14
Figure 2.2: Text-links used as a Navigation Scheme	15
Figure 2.3: Navigation button	19
Figure 2.4: Image map as navigation scheme	20
Figure 2.5: Search engine friendly website	39
Figure 2.6: Header with navigation links	39
Figure 2.7: Left hand side navigation	40
Figure 2.8: Elements of ranking strategies	42

CHAPTER ONE: INTRODUCTION

1.1 INTRODUCTION

The problem of finding material on the World Wide Web has been extensively debated in information science journals, computer literature, and the popular media (Turner & Brackbill, 1998: 258, Weideman, 2005:11). Internet Visibility Service™ (2005) states that,

‘Two-thirds of first-world populations use the Internet. This means your “offline” business could be using the Internet to drive new business to it. Getting this to work is no accident, though. You have to know exactly what you are doing.’

Search engines and directories are the most important method Internet users employ to discover websites (Sullivan, 2001). According to this author, most search engines look at links as referrals or recommendations. From this the analogy can be drawn that the website links determine how a website will be ranked. Xue, Harker & Heubs (2000: 37) reported that links from websites with similar content to a developer’s own website will be considered as quality links. Thurow (2003b) reported that the strategy of placing keyword-rich text on webpages is useless if the ‘search engine spiders’ have no way of finding such text. As a result, the way webpages are linked to each other, and the way websites are linked to other websites could impact adversely upon search engines’ visibility.

Thurow furthermore identified the following website navigation schemes that are problematic to ‘spiders crawling’ a website:

- Poor HTML coding.
- Image maps.
- Frames.
- JavaScript, dynamic or database-driven webpages.
- Flash components.

It is important for SMMEs to make their e-commerce websites visible to search engines so that they will be visited by a large number of users. For the purpose of this research, the term search engines will refer to 'crawler' or 'spider based' search engines.

1.2 BACKGROUND TO THE RESEARCH PROBLEM

It is reported that, if search engine spiders can find and navigate pages quickly and easily, the website has a much better chance of appearing at the top of search results (Kerr, 2001: 22). Position Research (2002), states that an optimally crafted navigation or links structure would increase page ranking. Furthermore, an optimal navigation architecture can be constructed to satisfy the needs of both human visitor and search engine at the same time.

Research indicates that the SMME sector contributed 53.9% to formal private sector employment and 34.8% to the total domestic gross product in 2001, while small business form 98% of the total business population in South Africa (South Africa: Department of Trade and Industry 2001). According to Baard & van den Berg (2004), Information Technology (IT) provides benefits to SMMEs, including gains in efficiency and business performance, increased managerial competence, and the provision of information to increase flexibility and responsiveness to external environments.

Research has shown that most web developers and designers first build a website and only later focus on the 'searchability' and 'visibility' of the site. According to Munafo (2002), this can be a costly mistake. This author is of the opinion that companies spend thousands of dollars on websites which:

- Are rejected by directory editors.
- Cannot be indexed by search engines.
- Are not visible to the search engines.

A study by Nobles & Alexander (2002) indicated that many webpages contain content that is not 'search engine friendly'. Search engines have great difficulty in reading and ranking certain website content such as animations, images, Java applets, movie files, and frames (Alexander, 2002b).

There appears to be a link between a SMME website, Internet search engines, and navigation schemes employed on websites. This research will endeavour to identify, compare and report on what impact the use of text-link navigation schemes or graphic-link navigation schemes has on search engine success. Furthermore, the study is designed to determine how useful these methods of linking are in improving visibility and ranking of websites in Internet search engines. This author regards this situation serious enough to warrant close inspection, experimentation and reporting.

1.3 STATEMENT OF THE RESEARCH PROBLEM

The research problem of this study is the lack of empirical evidence on the effect different hyperlink structures have on website visibility. Furthermore, there are also conflicting claims in the literature on how search engine crawlers interpret these links. This research examines the following question related to the two most commonly used navigation schemes namely, 'hypertext' and 'graphics links':

- Do webpages that use a text navigation scheme have a higher ranking and visibility to search engines than those that use graphics-based navigation schemes?

1.4 SIGNIFICANCE OF THE STUDY

Knowledge about and insight into e-commerce website navigational features are desirable to enable designers to enhance the visibility of webpages to search engines. Website developers should understand the respective benefits and disadvantages of using hypertext links and graphical navigation icons or buttons.

South Africa is currently ranked 25th in the world in terms of Internet connectivity. Table 1.1 depicts the spread of Internet hosts in Africa, which shows a more positive picture in this context.

Table 1.1 Internet hosts in Africa for 1998 (Source: Clarks, 2000)	
Country	No. of hosts
South Africa	95%
Egypt	2%
Namibia	1%
Zimbabwe	
Botswana	
19 other African countries	2%

Hof & Sager (2003: 12) found that most search engine marketers label both 'search engines' and 'directories' as 'search engines', even though they have unique individual characteristics. It is of importance for website owners to understand the difference between the two terms, because the strategies deployed for listing in search engines are different from the strategies deployed for listing in directories (Thurow, 2003b). The single most important marketing strategy that a company can deploy is to list with the top search engines and indices. It is recommended that SMMEs develop their e-commerce websites according to very specific guidelines, which will culminate in more website hits, more customers, and a potentially higher return on investment.

The significance of this research is the focus on SMME website navigation architectures for e-commerce based websites, specifically using text as opposed to graphics, with the aim of improving visibility and ranking of websites to search engines. This dissertation explains how search engines find and rank webpages, with emphasis being placed on what webmasters, website owners and web marketers can do to improve their search engine ranking by utilising improved links. This can serve as a basis for further research in this field of study.

1.5 RESEARCH HYPOTHESES

Should a company use search engines as part of its online marketing strategy, creating search-engine friendly design templates can save such a company considerable time and expense (Thurow, 2003b). This study examines the relationship between the text-based and the graphic-based navigation architecture and visibility to search engines. The following hypotheses will form the basis of the research:

H₀: There is no difference, in ranking and visibility of SMME websites to search engines, between pages that use text-based navigation schemes and pages that use graphics-based navigation schemes.

H₁: There is a difference in ranking and visibility of SMME websites to search engines between pages that use text-based navigation schemes and pages that use graphics-based navigation schemes.

Results as to how these navigation schemes have an impact on website visibility, were based on logs and statistical data produced by two industry-standard website evaluation programs namely: Report Magic for Analog and Real World Technology. The two variables for the primary focus of the study are Visibility and Ranking, while text-based and graphic-based website navigation schemes are the grouping variables.

Investigative questions were formulated as follows:

- i. What impact does a text-based navigation architecture have on website visibility to search engines?
- ii. What impact does a graphic-based navigation architecture have on website visibility to search engines?

1.6 AIMS OF STUDY

Research has indicated the existence of a positive and significant relationship between the capabilities of IT and the effective adoption thereof and business performance (Baard & van den Berg, 2004). The aim of this research is to evaluate the effectiveness of using different

navigation schemes to improve visibility and ranking of websites. This transposes into the way pages are linked to each other, and the way a website is linked to other websites. To achieve a visible website without an effective navigation architecture and other visibility factors, website marketers are forced to pay for search engine visibility through 'paid inclusion' or 'pay-per-click' programs.

The primary objectives of this study are to investigate and compare the difference in e-commerce website visibility to search engines between the navigation architectures employed by a website. This fact limits the scope of this study to:

- SMME e-commerce websites only.
- Text and graphic website navigation architectures only.
- Crawler/spider based search engines only.

To achieve the primary objective of the study, the secondary objectives of this study are:

- To evaluate the theory and investigate the most commonly used navigation architectures and their advantages/impact on search engines.
- To establish whether or not there is a significant difference in website visibility to search engines between the two website navigation schemes.

1.7 DELINEATION OF STUDY

The following factors that could impact website visibility to search engines are excluded from this study:

- Dynamic webpages and Flash components.
- The use of meta-tags and meta-data to increase website visibility.
- Keyword options.
- Element inclusion and exclusion for visibility of the website.

1.8 OVERVIEW OF CHAPTERS

The remaining part of the study is divided into chapters as follows:

- Chapter 2 (the literature review) covers differences between human-based, spider/crawler based and hybrid search engines. Furthermore, navigation schemes are inspected in terms of advantages and disadvantages, also including their impact on search engines. In addition, website design and factors that affect website visibility are identified.
- Chapter 3 considers the methodology used in the study, followed by a discussion and comparative analysis of the different methodologies and their suitability. Motivations for the methods to be used in this study are discussed.
- Chapter 4 specifies the type, source, nature and form of data collected. Furthermore, data collection and analysis methods, results and data interpretation are elaborated upon.
- In chapter 5, the research is concluded by discussion of findings, conclusions and recommendations made to solve the research problem.

1.9 CONCLUSION

In this chapter, the reader is introduced to the field of study by providing a holistic overview of the research. Background to the research problem is provided and the significance of the proposed research emphasised. The two hypotheses are stated, further supported by the investigative questions for the research. The aims of the study are elaborated upon and the delineation of the research listed.

CHAPTER TWO: LITERATURE REVIEW

2.1 INTRODUCTION

An Internet presence provides SMMEs with opportunities to enhance business offerings in a cost effective and practical manner (Fontanella 2000). It can help companies increase profitability, reach new markets, improve customer service, distribute products faster, and communicate more effectively with supply chain partners (Kleindl, 2001; Watson, Pierre, Leyland & George, 2000). Because of these facts, few businesses can afford to ignore the competitive potential of engaging in business-to-business (B-to-B) electronic commerce. It is therefore apparent that SMMEs of all sizes should consider having a website that is visible to search engines and users.

The development and implementation of B-to-B websites have become important issues for SMMEs. The ability of SMMEs to successfully exploit the web is of utmost importance, since over 80% of global economic growth comes from SMMEs alone (Boyes & Irani, 2004: 192). Furthermore, small businesses are often the most ill equipped organisations to effectively exploit this medium for commercial purposes.

In this chapter the various types of search engines and navigational elements that are applied within an internal linking structure strategy are investigated. Furthermore, the different methods of juxtaposing the various navigational elements in order to achieve the most beneficial linking structure for targeted website visitors and search engine spiders are analysed.

The navigation schemes to be discussed include:

- Simple hypertext links.
- Graphical navigation icons or buttons.
- JavaScript or DHTML dropdown or pulldown menus.
- Flash, GIF's and Flash navigation.

- Image map navigation.
- Framed navigation.

In addition, website design in terms of search engine visibility will be discussed.

Thurrow (2003a: 85) lists some of major factors which contribute to a website's visibility to search engines namely, targeted keywords, meta-tags, quality content, navigation scheme, and others. Reasons for the 'invisibility' of websites to search engines have been reported to include hidden text or links to search engines, Flash page introductions (splash pages), redirecting users and doorway pages (Bruemmer, 2002). Alexander (2002b) claims that the use of metadata enhances website visibility to search engines. Alexander (2002b) furthermore claims that, 'if a website does not appear on the first page of the search engine, the website is invisible'.

Rose, Khoo & Straub (1999: 3) identified the following six categories of technological limitations posing the greatest threats to the development of e-Commerce:

- Download delays.
- Limitations in the interface.
- Search problems.
- Inadequate measurement of web application success.
- Security (real and perceived) weaknesses.
- Lack of Internet standards.

Collin (2002: 15) identified the following as the most common mistakes made by developers of e-commerce websites:

- Using Microsoft FrontPage and FrontPage Express, which may result in unacceptable code and 'cheesy' templates.
- Novice designers seem to lack knowledge of graphic layout.
- No graphic optimisation.

This may result in websites that contain large graphic headers (70 – 100k), which could become time consuming to download.

2.2 INTERNAL LINKING STRUCTURE ELEMENTS

An internal linking structure, referred to as a Search Engine Optimisation (SEO) strategy and commonly known as a website navigation scheme, is a fundamental requirement for end-user and search engine spiders (Mitra, 1999). Logical and intuitive navigation will assist a website visitor to find or navigate around a website effortlessly, and furthermore assist to increase response rates (be it sales, phone calls, or simply readers) (Foot, Schneider, Dougherty, Xenos & Carsen, 2003). Paulussen (2004), concludes that a well-constructed and coded linking structure will help search engine spiders navigate a website with ease. This could ultimately assist visitors to locate website content via search engines easily.

2.2.1 Graphic components

In a business environment, real-time information about customer shipments is often required to retain a customer's business. The provision of interactive website content may present even more of an opportunity for SMMEs to better serve their customers than the inclusion of better informational content. According to Rea-Smith, Alexander, Ellinger (2002: 7), interactive content can offer customers approximately the same level of service through the Internet as opposed to the level of service they can get directly from a service representative. However, research findings suggest that relatively few SMMEs are utilizing their websites to differentiate their business by incorporating interactive content offering customers total product visibility for supply chain shipments (Rea-Smith *et al*, 2002: 7).

Download delays impact upon the development and use of Internet applications such as multimedia for B-to-B commerce (Lide & David, 2001: 30). For example, technology exists to allow the viewing of a TV program over the Internet. However, the amount of 'wait time' required before such a program is downloaded is prohibitive and therefore it is not often done this way (Lide & David 2001: 31).

Download time is primarily a function of:

- The size of the data files being transmitted.
- The technological configuration of nodes.
- The network infrastructure.
- The bandwidth of the connection between nodes and infrastructure (Pfattenberger, 2001).

2.2.2 Text Components

Even though search engine spiders are powerful data-gathering programs, Hypertext Markup Language (HTML) coding or scripting can prevent a spider from crawling webpages. According to Thurow (2003a: 89), examples of website navigation schemes that can be problematic include the following:

- Poor HTML coding on all navigation schemes: Browsers can not display webpages with insufficient HTML coding.
- Image maps: Many search engines do not follow links inside image maps.
- Frames: Some search engines e.g. Google, Inktomi, and Lycos do follow links on a frames-based website. However, the manner in which pages display in the search results is not ideal.
- JavaScript: The major search engines do not follow the links, including mouseovers/rollovers, arrays, and navigation menus, embedded inside JavaScript.
- Dynamic or database-driven webpages: Pages that are generated through scripts or databases, or that have a ?, &, \$, =, +, or % in the URL, pose problems for search engine spiders. Uniform Resource Locaters (URLs) with CGI-BIN in them can also be problematic.
- Flash: Currently, only Google and FAST Search can follow the links embedded in Flash documents.

It is therefore necessary for web developers to include a text navigation scheme to enable spiders to record the text on webpages.

This culminates in two forms of navigation on a website being possible, namely one that pleases human targeted audiences visually and one that the search engine spiders can follow.

2.2.3 Popularity Components

The popularity components of a search engine algorithm according to Anon (2000: 25) consist of two subcomponents:

- Link popularity.
- Click-through or click popularity.

The quality of the website linked to a website has more value than the number of websites linked to the website. Thilmany (2003: 11) concludes that:

‘... to develop effective link popularity to a website, the website should be listed in the most frequently visited directories - it can boost your search position if a directory that is associated with a search engine lists your website’.

Obtaining links from other websites is not enough to maintain optimal popularity. Major search engines and directories are measuring how many times end users are clicking the link to webpages, and how long they are staying on a website. Furthermore, measuring how often end users return to a website has value (Kerr, 2001: 24). The number of clicks to a webpage and duration of a visit are all measurements that constitute a website’s ‘click through popularity’.

Search engines and directories measures both ‘link popularity’ (quality and quantity of links) and ‘click-through popularity’ to determine the overall popularity components of a website.

2.2.4 Alternate text components

Alternate text is the text that is associated with a graphic image in HTML code. Alternate text 'instructs' the browser to display the text if the image itself is not downloaded. An example of the HTML code for alternative text follows:

```
<img src='image/home.gif' height='25' width='60' alt='BooksLand'>
```

- *img* represents the HTML coding which instructs the browser to insert a graphic image.
- *src* represents the location of the file, an attribute that indicates the filename, source or URL of the graphic image.
- *width and height* represent the graphic image dimensions.
- *alt* represents the alternative text attribute.

Whalen (2003) reported that search engines build lists of words and phrases from webpages. It follows that to do well on search engines, designers should place these words on the webpage in strategic HTML tags. The most important part of the text component of a search engine algorithm is keyword selection (Thelwall, 2000a: 151). Different search engines do not place emphasis on the same HTML tags. For example, Inktomi places some emphasis on meta-tags, while Google ignores meta-tags. To do well on all the search engines, it is recommended to place keywords in all the HTML tags (Furnis, 2001). Thurow (2003a: 76) concluded that keywords need to be placed in the following areas:

- Title tags.
- Visible body text.
- Meta-tags.
- Graphic images (alternative text).

This is supported by Crowell (2004), who is of the opinion that the title tag and the visible body text are the two most important places to insert keywords, because search engines index and place significant 'weight' on the text. Many search engine marketers believe that placing keywords in domain names and filenames impact upon search engine positioning

(Fagan, 2001: 169). Fagan further states that some search engines marketers believe that this strategy is beneficial, while others believe that the benefits derived from this approach are minimal.

2.3 NAVIGATION SCHEMES

The Internet is currently one of the most significant infrastructures in the business world. Its capacity for representing information is 'unsurpassed by any other current system in the world' (Smith, 2002: 6). For SMMEs, the Internet offers an excellent means to compete globally with other businesses of differing sizes. Furthermore, businesses in any country have the ability to market themselves in the global marketplace.

Alexander (2002a) finds that many websites use multiple navigation schemes, (see Figure 2.1). For example, a website may have a series of navigation buttons down the left side of the screen and have corresponding text links at the bottom of the screen, with holder text links within the actual body content. Common navigation schemes include the following:

- Hypertext links.
- Navigation buttons.
- Image maps.
- Drop-down/pull-down menus.
- Animation/Flash buttons.
- Dynamically generated URLs, such as those generated by a website search calling up a database.



Figure 2.1: Images, buttons and text-links used as navigation scheme (Source: Amazon, 2004)

A study by Eastman & Jansen (2003: 385) proved that a search engine friendly and user-friendly design have at least two forms of navigation, namely one that website visitors can follow, and one that search engines can follow. It was suggested that website designers should consider their target audience prior to the creation of a website's navigation scheme. Navigation schemes should enable visitors to find what they are searching for as quickly and as easily as possible.

Each one of the navigation schemes selected has some advantages and disadvantages, which impact on the website's search engine visibility. This is supported by Weideman & Ngindana (2004: 295), in a recent study on navigation schemes. The authors concluded that the main problem identified is that the requirements for a 'human friendly scheme' are in direct opposition to those of a 'crawler friendly scheme'.

2.3.1 Hypertext Links

Hypertext links are the simplest navigation elements available for building an internal linking structure. Thurow (2002) defines a hyperlink as 'a computer-based text retrieval system that enables a user to access particular locations in a website or other electronic documents by clicking on links within a specific webpage or documents'. A hypertext link is represented by a word or set of words placed inside an anchor tag, as depicted in Figure 2.2. Hypertext is the text placed between the `<a>` and `` tags. The following HTML code represents a simple hypertext link: `E-Books` where

- i. `<a` starts the anchor tag, and
- ii. `href` represents the attribute of the anchor tag.

This refers to the target location of a web document. In this example, the location of the web document is a webpage named 'ebooks.html'. Thurow (2003b) recommends that the location of the webpage in a hypertext link be enclosed in quotation marks such that *e-books* is the anchor text, and `` closes the anchor tag and hypertext link.



Figure 2.2: Text-links used as a navigation scheme (**Source:** Amazon, 2004)

It has been found that all search engine spiders read text links because they have the ability to record the text in and around the link, and follow these links from webpage to webpage (Gardner & Pinfield 2001: 93). According to Thurow (2002), search engine spiders prefer the simplicity of the hypertext link to any other form of linking structure elements. In fact, search engine marketers use a specific term that refers to the HTML text inside a hyperlink anchor text. A recent report on search engine visibility (Kerr, 2001: 22) found that many search engines consider anchor text relevant, because webmasters generally link to pages that contain information in which their target audience is interested, and as a result, anchor text is deemed to be important. Usability experts often recommend 'breadcrumb' links.

Breadcrumbs are text link schematics on every webpage that guide the user on how he/she navigated to the current webpage. Whalen (2003) states that:

'Breadcrumbs are commonly used at the top of a webpage and are hierarchical in nature, they provide the target audience with valuable information about visited and unvisited pages. Users automatically know that a blue underlined word or just an underlined word indicates an unvisited link and that a purple or a faded colour indicates a visited link.'

Inan (2002) concludes that many directories, including Yahoo!, use breadcrumbs as a navigation scheme. Due to the fact that breadcrumbs are generally placed at the top of webpages, search engines consider the text placed inside the breadcrumb link as important. As a result, if a website uses breadcrumbs as a navigation scheme, they are compelled to use 'keywords'.

However, there are some disadvantages to hypertext links which include (Inan, 2002):

- i. Hypertext links are visually less attractive than any other form of navigation elements.
- ii. By default, hypertext links create blue underline text and if used too much will make a page difficult to read.
- iii. If overused, they can interfere with a webpage's keyword density.

2.3.1.1 Absolute and Relative Links

Two types of hyperlinks exist, namely 'absolute' and 'relative' hyperlinks (Thurow, 2003b). An absolute link defines a specific location of the webpage or document. The location information includes the protocol to use to access the document, the server from which to obtain it, the directory in which it is located, and the name of the document. The following is an example of an absolute link:

```
<a href='http://www.BooksLand.com/BooksUses.HTML'>BooksUses
```

With relative links, the search engine spider or browser already knows where the current document is located on a server. As a result, if it is a link to another document in the same directory, it is not necessary to write out the full URL. The following serves as an example of a relative link:

```
<a href='BooksUses.html'>Books Uses</a>
```

Gardner & Pinfield (2001) note that it makes no difference to the search engine spiders whether an absolute link or a relative link is used. All search engine spiders can follow both types of links.

2.3.1.2 Places to use Text Links

Winn & Beck (2002: 18) recommended that, for maximum search engine visibility, text links as either the primary or secondary navigation for a website design should be used. Thurow (2003b) finds the following as 'best' and 'worst' places to put a text link:

- i. The top of a webpage. Because this is the first place the search engine “sees”, as long as the text is short, it does not interfere with search engine visibility.
- ii. The left side of a webpage. This is not always a good place to put text links. If the text is too long, search engines often use this text in their excerpt in the search results, which could cause the webpage to appear not to be focused on a specific topic.
- iii. Bottom of a webpage. This is a good place to put a text link, as these text links can correspond to graphic images, navigation buttons or an image map at the top of the screen.
- iv. The right side of a webpage. This is also a good place to put a text link since cross-linking important webpages becomes possible.
- v. Placing text links in the middle of a webpage. Highlights the important aspects of the document.

2.3.1.3 Search Engines and Hypertext links

A recent study conducted by Thurow (2003b) found that text-link navigation schemes might seem like the ideal solution because many search engines favour this type of link. Furthermore, a webpage full of text links tend to download faster than a webpage full of graphics images. Using text links as the primary navigation scheme can satisfy both the search engine and the targeted user. This research intends to determine whether graphic or text links are interfering with search engines’ visibility and page ranking. Too many text links can interfere with keyword density and with keyword prominence (Ollins, 2002).

Tansley (2002: 58) claims that placing breadcrumb links at the top of the page could retain keyword prominence. If there are more words in text links than in the main body content, it is better to use graphic images as part of the navigation scheme (Thurow, 2003b). A further disadvantage of using text links as the primary navigation scheme is that they are often the first words of text that the search engines read. Text tends to be the same on many webpages as they are inherent to the navigation scheme of every

webpage. The first text introduced to the spiders is unique if a search engine uses the meta-tag description to display in the search result. The crawler generally uses the first text found at the top of a webpage. (Murata, 2003: 39)

A series of text links does not accurately describe the contents of a unique webpage, and could therefore appear as a set of unrelated words to users (Xue *et al.* 2000: 31). Murata (2003: 40) finds that too many text links on a single page can also interfere with page legibility. Due to the fact that users prefer simplicity and ease of navigation, web designers should find innovative ways to make text links more visually distinct, easy to find and legible, such as placing them in a coloured table cell or a coloured side bar (Murata, 2003: 39).

2.3.2 Graphic Links

A study conducted by Murata (2003: 40) reported that the objective of writing a webpage is to draw a target audience who would respond positively to it.

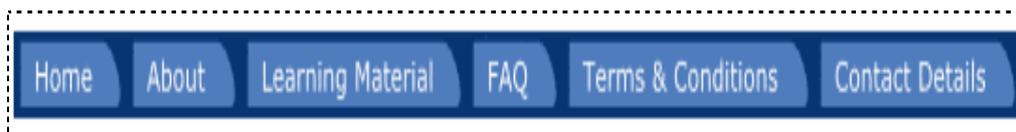


Figure 2.3: Navigation button (Source: Amazon, 2004)

As a rule, navigation buttons take the form of a graphic image (as seen in Figure 2.3), generally in a GIF or a JPEG format, that links to a single URL.

2.3.2.1 Navigation Buttons

Navigation buttons provide users with a visual representation of how to navigate a website. They should be visually appealing to easily draw attention to the important parts of the webpage. Users 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' buttons on a website. Xue *et al.* (2000: 33) recommend that navigation buttons

should always contain alternative text in the HTML code, in order for website visitors to navigate the website.

2.3.2.2 Image Maps

An image map is represented by a single graphic image that enables users to access different webpages by 'clicking' on different areas of the image (see Figure 2.4). It is reported that many search engines do not follow the links contained within the ambit of an image map, because of the possibility of 'image spam' (Thurow, 2003a: 104). Furthermore, if image maps are used as part of the navigation scheme, text links or navigation buttons should be used elsewhere on the webpage.



Figure 2.4: Image map as navigation scheme (Source: Amazon, 2004)

Mertz (2001: 23) is of the opinion that a single image downloads faster than multiple graphic images. For example, if a website has 16 navigation buttons of 2kb each in file size, the total size for these images is 32kb. A single image map may only be 8-1kb in size, which is much smaller than the set of separate navigation buttons. With an image map, there is only one call to the server as opposed to 16, which also speeds up the

downloading process regardless of file size. Furthermore, search engines spiders find it easier to access the main body text on a webpage, if the navigation scheme is represented by a single graphic image as opposed to multiple graphic images (Mertz, 2001:31).

2.3.2.3 Use of Image Maps

According to Agillo (2000: 243):

‘When is it not a 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’.

Xue *et al.* (2000:35), supports the finding of Aguillo, pertaining to graphic images and text links as a navigation scheme. The analogy can be drawn that should the image map download more quickly than a set of navigation buttons, the image map may be a better choice.

2.3.2.4 Search Engines and navigation buttons

All the search engines can follow the links surrounding a navigation button (Furnis, 2001), as long as the navigation button does not contain JavaScript within the anchor tag. JavaScript *per se* can pose several problems for search engines spiders, and as a result, not all search engines can follow this type of link. It has been reported by Thurow 2003a: 109, that the following type of link is considered to be search engine friendly:

```
<a href='BooksMedicine.html'><img src='images/Books.gif' width='60'  
height='20' alt='Books Medicine'></a>
```

The following type of link represents a navigation button with a rollover effect, and is considered search engine unfriendly:

```
<A href='BooksMedicine.html'  
OnMouseOut='MM_swapImgRestore()'
```

```
OnMouseOver='MM_swapImage('image/Bookson.gif','Images/BooksMed
.gif',1)';><img src='Image/Booksmed.gif' width='60' height='20' alt='Books
Medicine'></a>
```

Furnis (2001) also reported that when too many attributes are added to the anchor tag, such as rollover script, the attributes could culminate in making the links less search engine friendly. Currently, the major search engines prefer straightforward link coding without any type of script. According to Ollins (2002), an effective search-engine friendly layout example is to use navigation buttons (with or without JavaScript rollover) and corresponding hypertext links at the bottom of the webpage.

2.3.3 Dropdown and Pulldown Menus

The primary advantage of using drop-down menus in a navigation scheme is a saving of screen real estate (Holder, 2003: 135). Drop-down menus do not 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 content on the site that their target audience can read above the fold.

2.3.3.1 Search Engines and dropdown menus

Drop-down menus are generally not search engine friendly because they require either JavaScript or a CGI program to work (Crowell, 2004). Due to the fact that search engines generally do not follow these types of links, Thurow (2003b) is of the opinion that a webpage should always have an alternative form of navigation for the search engines to follow.

Web designers can use a combination of drop-down menus with corresponding text links at the bottom of a webpage for search engine visibility (Greenough, Fakun, & Kay, 2001: 389).

2.3.4 Dynamically Generated URL's

Often, search engine marketers state that search engines cannot follow the links inside dynamically generated webpages, due to the fact that they can not interpret the URL used to retrieve the webpages (Sullivan 2001).

Search engines have difficulty in reading the following characters or symbols as part of a URL: ?, &,%,\$, =, and +. This can be attributed to the following reasons:

- i. Search engines frown on content delivered to them repeatedly.
- ii. Search engines want their search results to be accurate, so they
- iii. update their indices every four to eight weeks. Should they include dynamically generated URLs in their search results, the content may change between the time they recorded the URL and the time the URL appears in the search results, culminating in the search result not being accurate.
- iv. Certain dynamically generated URLs can trap a search engine spider and cause it to crash. Programmers who fail to close their 'if' and 'while' statements on webpages can cause a search engine spider to enter an infinite loop (Greenough *et al.*, 2001: 40).

Search engines are progressing in their capability to spider dynamic URLs. Web developers should focus their efforts on delivering webpages to search engines that do not contain stop characters e.g. ?, &, % etc. in their URLs.

2.3.4.1 Dynamic URL's

Dynamic server-based technologies such as Hypertext PreProcessor (.php), Java Server Pages (.jsp), Active Server Pages (.asp), ColdFusion (.cfm) and Perl are used to develop websites containing a large number of pages (Hout, 1999). Such technologies provide programmers with the tools to build sites so that adding products or pages does not require extensive HTML work. It is recommended by Turner & Brackbill (1998: 260) that all high volume sites should use one of these technologies in

order to maximize efficiency and maintain profitability. The power and flexibility provided by such technologies offer outstanding benefits.

There are however, issues that need to be addressed regarding the way search engines 'crawl' a dynamically driven Website. Craven reported that these issues do not have to do with the pages that are generated, but with the URLs these technologies generate (Craven, 2001: 210).

2.3.4.2 Importance of dynamically generated URL's

In order to better elucidate the understanding of what a search engine 'sees', a sample URL will be analysed (a normal URL):

<http://www.freshpair.com/underwear.html>.

A complex URL example:

[http://www.freshpair.com/catalog.php?formid=4&id=8&brand=&brasize=\\$I on=women](http://www.freshpair.com/catalog.php?formid=4&id=8&brand=&brasize=$I on=women).

According to Smith (2001), it is the question marks, equal signs and ampersands within the URL that could cause problems for crawlers. Such entities are commonly referred to as 'stop characters' in search engine terms. Stop characters signal to search engines to stop crawling past a certain point, limiting the number of pages which can be crawled on the website (Thelwall, 2000b: 155). The objective of the search engine is to place pages into its index that are unique, and not those containing repetitive information (Smith, 2001: 250).

Search engines generally combat this issue by 'pruning off' the URLs after a specific number of variable strings (i.e.. ?, =, &) (Thurow, 2002). A typical website address which reads as follows:

[http://www.rustybrick.com/portfolio_client_list.php?industry=6,](http://www.rustybrick.com/portfolio_client_list.php?industry=6)

may be pruned down to:

http://www.rustybrick.com/portfolio_client_list.php

This 'pruning' approach in order to limit the number of content could be repeated (Craven, 2001: 214). Search engines according to Turner &

Brackbill (1998: 265), want to keep the number of pages that a website contains to a minimum in order to:

- i. Eliminate duplicate search results with the same content.
- ii. Make the crawling of the pages efficient.

This in particular is problematic where there are a number of methods of finding the same product with an unlimited number of pages. The solution suggested by Byrne (2004: 28) is to program a Mod_Rewrite on the URLs

to remove the stop characters from the URLs. Byrne modified the following URL:

<http://www.freshpair.com/catalog.php?formid=5&query=bra&ion=women>

to read:

http://www.freshpair.com/catalog_section_women_id_8.html

This was achieved by replacing all stop characters with underscores and more friendly URL characters and names.

2.3.5 Cascading Style Sheets (CSS)

Cascading Style Sheets (CSS) is a feature of HTML developed by the World Wide Web Consortium (W3C), to assist developers in designing graphic-based websites to attract users.

With style sheets, website designers and end users can create style templates that specify how different text elements should appear on a webpage without impacting upon its structure. In a study by Winn & Beck (2002: 18) it was reported that using CSS in website design, enhances the end user experience as well as search engine visibility. Furthermore, style sheets can assist to reduce 'code bloat' and the overall file size of a webpage. For example, CSS enables a web designer to display fonts and typeface exactly as was intended. Another way CSS benefits webpages vests in the headings. Heading tags, especially the <h1> tag, can be more attractive when using CSS (Thurow, 2002).

2.3.6 Frames, Flash, DHTML, JavaScript rollovers

Flash, DHTML, JavaScript rollovers, style sheets, Frames and animations can add style to a website and help users to understand the navigation scheme better and make the website easier to read, thus enhancing the user experience. However, the graphics created by these programs can interfere with a search engine marketing campaign. Search engine crawlers always look for text on a webpage, and some website designs present the search engines with no text to index, as is the case with many webpages containing Flash designs (Furnis, 2001).

- i. Technical choices or methods for building a website can also interfere with how well a search engine can spider a website. Collin (2002: 18) claims that the design of Flash navigation buttons is not problematic for search engines, but that the actual Flash technology used to produce the navigation button is. JavaScript as defined by Thurow (2003b), is a programming language that enables website designers to add flair and interactivity to their websites. Some of the most common uses for JavaScript include: Rollovers/mouseovers, Pop-up windows, Navigation menus.

The JavaScript rollovers has become so widely used that it is now considered a standard component of website design packages (Graham & Inan, 1995).

These authors further report that when JavaScript is used in a website's navigation scheme, the script can greatly decrease the 'crawlability' of the links. Currently, most search engines do not follow, or are limited to crawling the links embedded inside the JavaScript code (including rollovers and menus). However, some JavaScript code segments are more search engine friendly than others. As a general rule, a script which is more 'crawlable' includes Mouseover/Rollover JavaScript placed between the <head> and </head> tag (Thurow, 2003b).

2.3.6.1 Frames and search engine visibility

Whether or not websites should have frames has been debated at length. Usability experts do not recommend the use of frames because it limits the ability to properly bookmark a page, and furthermore the URL appearing in the browser does not match the content (Eastman & Jansen, 2003: 390). Furthermore, these authors state that, due to the fact that search engine spiders index text links, the main problem with initial frameset code is that it does not present search spiders with keyword-rich text to index and set links to follow.

If a website uses frames, it is recommended to include keyword-rich content and links to the most important pages within the website. This would facilitate the ability of the search engine spiders to record the most important text.

2.3.7 Flash components and search engine visibility

Flash provides opportunities of creating websites with visual flair. Web designers use Flash as it provides a way to deliver vector images as opposed to bitmap images over the web. Vector graphics are scalable when visitors resize a browser window. Webpages designed in Flash remain static no matter how large or small the browser window becomes (Davidson, 2001: 38).

Flash movies represent 'streaming', which transposes into the fact that when part of a vector image downloads, that part of the image displays on the browser screen while the rest of the movie downloads. As a result, Flash designs have benefits for both web designers and website visitors. In terms of search engine visibility however, Flash components are not an ideal choice. Although a few of the major search engines can crawl the links embedded within a Flash navigation scheme, the main problem with Flash content is that it contains very little text for search engines to index.

2.4 SEARCH ENGINES

The growth of the World Wide Web (also known as WWW, or simply Web) has been one of the most remarkable developments in the history of

documentation (Oppenheim, Morris & McKnight, 2000: 190). These authors furthermore add that, within a short period, search engines have proliferated, often claiming to provide better retrieval of material from the WWW than using other methods. For many reasons, the Internet and WWW is becoming a main arena for the business community worldwide. The Internet is both an enabling tool for business and the new business environment that is transforming the economy. According to Smith (2002: 6), the Internet is the most global, borderless, cost-effective and open business application and communication infrastructure. Interactions and relationships between SMMEs and their customers have changed due to the Internet, as the Internet now allows customers many more choices.

According to Oppenheim *et al.* (2000: 191), search engines can be divided into four categories namely, robots, directories, meta-search engines and software tools. Furthermore, some search engines combine characteristics of more than one of these categories.

Internet search engines have been developed to aid in the finding of materials (Turner & Brackbill, 1998: 258). Kaplan and Carl (1997: 17) reported that the Internet acts as a network infrastructure for many types of computing beside the WWW. The key to WWW client/server architecture that makes it so germane to Business-to-Customer (B-to-C) e-commerce is its utilization of webpages with text, colours, graphics, sounds, video and links to other pages. Those seeking information from the Internet often start from a search engine, using either its organised directory structure or its text query facility (Thelwall, Binns, Harries, Page-Kennedy, Price, & Wilkinson, 2001).

According to Van Greunen and Wesson (2004: 75), search systems can be divided into two subjects. The first allows users to enter a query in a search interface and the second displays records containing metadata to represent each document. The term 'search engine' is often used commonly to describe crawler-based search engines and human-powered directories (Sullivan, 2001).

These two types of search engines however gather their listing in radically different ways.

Table 2.1: Crawler vs. directory based search engines (Source: Dykehouse & Sigler, 2000: 268).				
	HAR-VESTING	SUBMISSION	SE RESULTS	EFFECT
Crawler (Google)	Automated programs	Follow links	Index Catalogue	Has effect
Directory (Yahoo!)	Human Editors	Short description	Match description	Has no effect

According to Dykehouse & Sigler (2000: 268), search services can generally be categorized into two types of sources, namely directories and search engines. The search services use two main sources to obtain the listing as shown in Table 2.1. According to Dykehouse & Sigler (2000: 269), if two websites have the same text and link components, the website that users click the most will usually rank higher. Often, a popular website will rank consistently higher than sites that use multiple keywords. As a result, building a website that appeals to both directory editors and the target audience is very important for maximum search engine visibility.

Although search engines search an enormous volume of information at considerable speed, they have been the subject of widespread criticism, including slow response times and the retrieval of duplicate records. To retrieve relevant items a vast amount of irrelevant items are included. This study will identify differences between spider-based search engines (Google, FastSearch, Inktomi, Teoma) and human-based search directories (Yahoo!, OpenDirectory, Ananzi, Business.com), and compare the response in terms of indexing a text-based website as opposed to graphic-based website navigation scheme.

2.4.1 Differences between Crawlers and Web Directories

‘Crawlers’ and ‘web directories’ refer to two different search services available to the Web community, although they are often confused

(Barnes & Vidgen, 2002: 118). Crawlers-based search engines have indices that are built up by robots or crawlers. These robots reside on a host computer and retrieve information from sites on the web using standard protocol (Oppenheim *et al.* 2000: 191), whereas web directories build up their indices through human editors. Many crawlers and directories contain both a computer-generated index and a human generated index, and are referred to as 'hybrids'. Spider-based search engines 'crawl' sites looking for identifying tags that will determine how and when the results will appear. A human-based search engine employs cadres of researchers who look at millions of sites and categorize them according to what the researcher deems appropriate. Strategies being listed well in search engines differ from those being listed well in directories (Thurow, 2002).

Google, Inktomi, AltaVista, AlltheWeb and the like represent various forms of search engines. These search engines write programs known as robots, crawlers and/or spiders that have the following functions according to Li, Hsiung, Po, Candan, & Agrawal, 2002: 155:

- i. To locate webpages.
- ii. To read the contents of the webpages.
- iii. To report its findings back to the search engine's indices or databases.

Many search engines update their index either on a bi-monthly or monthly basis. When web searchers use a search engine to locate websites that are relevant to the keyword search, they are searching the search engine's index. A search engine with a larger and more up-to-date index provides a better representation of the information available in the Web (Browne & Jeremy, 2001).

Yahoo!, Open Directory Project (dmoz.org) and Gipsy are all forms of web directories. These directories rely on human editors to review sites that are submitted to the directory. According to Gardner & Pinfield (2001: 97), directories, unlike search engines, use a hierarchical tree structure to

organize their database. Another common distinction is that a directory tends to list websites (root directory of a website or homepage) whereas a search engine will list webpages (individual pages of a website) (Adam & Dean, 1999). Due to the manual process of adding sites to a directory, directories often have to supplement their search results with a search engine partner to increase the relevancy of the produced search results.

According to Browne & Jeremy (2001: 66), all search engines have their own confidential algorithms that determine which webpages are to be shown first. The algorithms assign weights to certain components or variables that it finds within a page. Many search engines consider the text within the title of the page to be deemed very important (Foot *et al.*, 2003).

The title of a page is considered important to search engines and is given more weight because:

- i. It is displayed on the top of the menu bar in the browser.
- ii. It is displayed in the search engine results page.
- iii. It is displayed in browser bookmarks when the page is added to favourites or bookmarks.

These weights provide the search engine's algorithms with a method to prioritise one webpage over another (Mitra, 1999).

2.4.2 Directory Search Engine

Directory search engines are also known as 'subject collections' or as 'subject gateways', and represent a collection of links to relevant URLs created and maintained by a subject expert (Oppenheim *et al.*, 2000: 192). Furthermore they rely mainly on people to identify and group resources. The first type of search service is called a directory, which uses human editors to manually place websites or webpages into specific categories. A directory is commonly called a 'human based' or a 'human-powered' search engine (e.g. Yahoo!). A human powered directory, such as the Open Directory, depends on humans for its listings.

The user submits a short description for the entire website to the directory, or editors write a short description for sites they review.

A search engine looks for matches only in the description submitted. Changing webpages has no effect on website listings (Kaplan & Carl 1997: 18). A search engine such as Yahoo! typically lists the name of a website with a description of about ten words - known to contain up to 20 words. The sites are organized according to categories and as in the instance of Yellow Pages directory, are designed for browsing a topic. When a search is performed in a directory, only the website titles and short descriptions are searched, which does not always reflect the entire content or diversity of the website.

A further limitation is that the directories usually list only a home page. Unfortunately, there is little that can be done to improve the ranking within a directory, however all is not lost as most of these directories use a second search engine to process results. Should a search fail to match any directory listings, an 'indexing' search engine will automatically perform the search (Li, 2003).

2.4.3 Crawler/Spider Search Engine

The second type of search service is called a 'search engine', and uses special software robots, called 'spiders' or 'crawlers', to retrieve information from webpages. The crawler starts from the homepage of a website, extracts all of its links, and then download the pages pointed to by these links (Li, 2003). It is found that although both Yahoo! and Google enable users to search, the information they provide in their search results differs (Thurow, 2003a).

Crawler-based search engines, such as Google, create their listings automatically. The indexing done by search engines such as Google, is executed in an entirely different way. In essence, all text on a page will be analysed based on words and phrases. This culminates in a profile or index being built on the website (Thurow, 2003a: 89).

The advantage of this approach is that it becomes possible for every page on the website to be listed, unlike a directory listing, which will only list the homepage. Kaplan & Carl (1997: 16) state that,

‘If you change your webpages, crawler-based search engines eventually find these changes, and that can affect how you are listed. Page title, body copy and other elements all play a role’.

Google is a search engine that makes extensive use of link popularity as a primary way to rank websites. This can be especially helpful in finding good sites in response to general searches such as ‘cars’ and ‘travel’. The system works so well that Google has gained widespread praise for its high relevancy (Hof & Sager, 2003: 13). Google also has a large index of the web and provides some results to Yahoo! and Netscape and several other search engines.

Search engines may implement policy changes that affect their search capabilities, such as the policy of excluding script-based webpages from searches (McGuigan, 2003: 69). McGuigan further states that the spiders of most search engines, i.e. the programs that search the content of pages on the web, generally will not locate pages containing a ‘?’ in the URL. This is the result of the policy established by the programmers to ignore scripted pages.

2.4.4 Hybrid Search Engines or Mixed Result

When the web was in its infancy, search engines either presented crawler-based results or human-powered listings (Li *et al.*, 2002: 159). Currently it is common for both types of results to be presented. Usually, a hybrid search engine will favour one type of listing over another. These search engines utilise databases maintained by other search engines, by accepting a single query from the user and sending it to multiple search engines in parallel (Oppenheim *et al.*, 2000: 192). The MSN search for example, is more likely to present human powered listings from LookSmart, however it also produces crawler-based results (as provided by Inktomi), especially for more obscure queries.

The functionality of these hybrid engines depends largely on the performance of the standalone search engines, as they create and manage their own databases. They are useful because users do not have to visit multiple search engines and re-enter their search terms for further searches.

2.4.5 How Search Engines, Directories and Hybrids work

Search engines automatically search the Internet, following links from documents and collecting information according to the HTML structure of the document (such as URL, document title, keywords in the text) about the resources they come across (Oppenheim *et al.*, 2000: 191). A search engine that is slow or which does not tend to provide relevant results on the first page is likely to lose customers (Li, 2003). Thelwall (2002) claims that in order to understand how to improve rankings, one needs to have a basic concept of how a search engine works.

It is also found that search engines perform three basic tasks (Thelwall, 2002):

- i. Search spiders find and fetch webpages, a process called crawling or spidering, and build lists of words and phrases found on each webpage.
- ii. Search engines keep an index (or database) of the words and phrases found on each webpage they are able to crawl. The part of the search engine that places the webpages into the database is called the 'indexer'.
- iii. Search engines then enable end users to search for keywords and keyword phrases found in their indices. Search engines endeavour to match the words typed in the search query with the webpages that are most likely to have the information for which end users are searching. This component is called the 'query processor'.

The question can be asked: How do search engines begin to find webpages? According to Thurow (2002), the usual starting points are lists of heavily used servers from major Internet Service Providers (ISPs), such

as America Online, and the most frequently visited websites, such as Yahoo!, the Open Directory, LookSmart, and other major directories (Dykehouse & Sigler, 2000: 269). Search engines spiders will begin crawling these popular sites, indexing the words on every single page of a website and following every link found within a website.

This serves as a primary reason why it is important for SMME websites to be listed in the major directories. Many search services are hybrids of a search engine and a directory. A hybrid search service usually gets most of its listings from one source, resulting in hybrid search services being classified according to the main source used. If a hybrid search service obtains its primary result from a directory and its secondary results from a search engine, the search service is generally classified as a directory, like MSN Search. MSN's primary results come from the LookSmart database, while secondary (fall-through) results come from Inktomi, which is a search engine.

Comparative research into the effectiveness of search engines produced conflicting conclusions as to which services are better at delivering superior data. One of the problems associated with the evaluation of search engines is the fact that they are constantly changing and developing their search mechanisms and user interfaces. Combining these facts with the constantly changing contents of the web, the analogy can be drawn that no specific evaluation of a search engine is likely to remain valid for any length of time.

2.5 SEARCH ENGINES VISIBILITY IMPORTANCE

As the web expands, users increasingly turn to search engines as their fundamental navigation aid (Nobles & Alexander, 2002). Search engines are forced to select from billions of possible webpages when building up its index. Ultimately, all search engines in one way or another favour websites that are search-friendly. Weideman (2001: 197) agreed that searching on the Internet using standard search engines, is not an easy task.

The number of current Internet users were estimated to be around 200 million in 2002, with an anticipated one billion users by end 2005 (Clarke, 2000). Clarke further states that the technology-mediated communication aspect of the Internet is enabling fast global contact with growing numbers of these individuals and groups. The SMME sector makes a valuable contribution to the economic development in South Africa (Baard & van den Berg, 2004: 1-32).

Whalen (2003) states that search engines are rated on five subjective categories, namely:

- i. Interface: View of the home pages and how effectively it displays on various browsers, screen widths, and platforms.
- ii. Relevancy: How accurate the results are within the scope of a search.
- iii. Usability: How effectively one can navigate and find information in the index (the most important aspect).
- iv. Website owners' respect quotient: How effectively the engine deals with the designers who built their database.
- v. Link popularity: The quality of the website linking to the website under review, and

2.5.1 Search Engine friendly e-Commerce websites

There has been a big drive for SMMEs to produce websites and be part of the Internet (van Steenderen, 2001: 3). Website content ranges from basic sites providing customers with general company information, to more complex sites where interactive applications offer customers virtual product catalogues, opportunities to provide feedback, and an array of services including the ability to pay for and fulfil orders online (Rea-Smith *et al*, 2002:15)). Successful websites add value to the user experience due to their ability to present fresh, useful, relevant, and comprehensive information. Virtual product catalogues on websites are replacing the laborious and expensive requirement of printing and updating the physical catalogues that are customarily used in B-to-B sales. Website content can also make information seeking more convenient for customers.

Many firms handle common information requests that would normally require access to a service representative by posting customers' most Frequently Asked Questions (FAQs) and the associated answers on their websites (Watson *et al.*, 2000).

Boggs (2000) found navigation and the title tag to be critical in the development of friendly e-commerce websites. The fundamentals of search engine friendly design apply to all types of sites, including e-commerce catalogue sites (Crowell, 2004). This author discusses the home page, category, sub-category and then the product page.

According to Van Steenderen (2001: 5), the overall view and the mission of the site are established first, which includes:

- i. Who the target audience is: Development needs to start with a target audience. Some companies do more damage by getting onto the Internet because they 'should' without thinking through who will be using the site and why. The identity of the target audience will impact significantly on the type of site promotions that will be carried out. If it is a site offering a service or selling a product that requires a more general audience, then advertising will cover a wider base – like using a banner on Yahoo! or television advertisements. If the audience is more specialized, then only related sites would be used for advertising, such as advertisements placed in specific trade magazines.

- ii. What products will be developed and delivered: Websites are generally centred on a target product. They typically offer a combination of goods and services. Goods may be any content, information, or commodity that may be stored by the site and delivered electronically. Elderbrock, Ezor, Dalton, and Weissberg (1997: 72) define services as, 'meaning the methods by which these goods are delivered'. These authors further identify three ways that the WWW is related to a company's standard products:

- a. Conventional – existing products, where the WWW is the delivery channel (e.g. software distributed on-line; electronic brochures).
 - b. Extended – enhanced or adapted products due to the WWW being used as the medium (e.g. a tracking service or on-line customer service centre).
 - c. Original – new goods with the WWW as its foundation (e.g. new electronic publication).

- iii. What the business objectives are: In the midst of competitive pressure to create a website presence and technological pressure to make use of all the latest advances, it is surprisingly easy to lose sight of a simple but critical question namely: 'Why does the company want a website?' (Elderbrock *et al.*, 1997:72). These authors divide the business objectives of websites into direct and indirect revenue.
 - a. Direct revenue includes advertising, sales, membership, subscriptions and technology licensing. If the site draws enough visitors, advertising is perhaps the easiest way of generating revenue, with membership being the most difficult (due to the Internet's tradition of free information).
 - b. Indirect revenue includes lead generation and prospecting, sales and marketing support, brand awareness, product supplementation and cost reduction.

Deciding on the business objective leads to the focus of the site. The major focus of the site has a direct impact on the mechanisms provided for interaction (Van Steenderen, 2001: 8). The typical e-commerce homepage has a company logo, some company specific information at the top, side bar navigation at the left, more prominent navigation in the middle of the page and featured products and sales on the right side of the page (Browne & Jeremy, 2001).

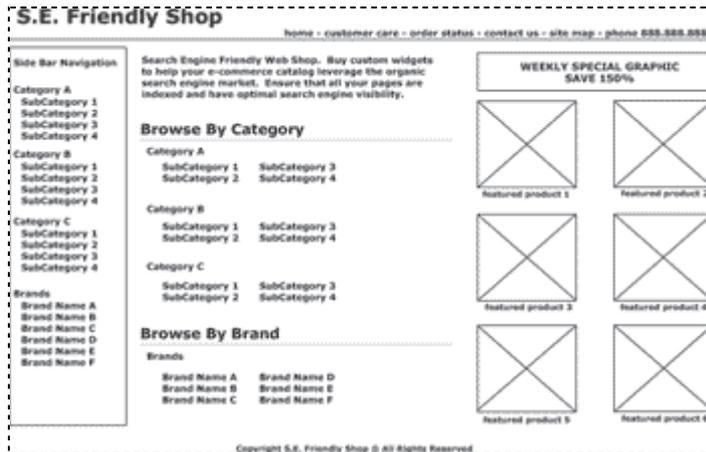


Figure 2.5: Search engine friendly website (Source: Boggs, 2000)

The homepage of a website will typically be depicted as set out in Figure 2.5. An analysis of each section of the homepage from a search engine optimization standpoint produces the following:

- i. Header: The header (see (Figure 2.6), which is typically an include file on each page, contains the company logo and top bar navigation. It is recommended that the logo should represent a link to the homepage, and contain an 'alt tag' and possibly a title attribute in the link tag (Boggs, 2000).

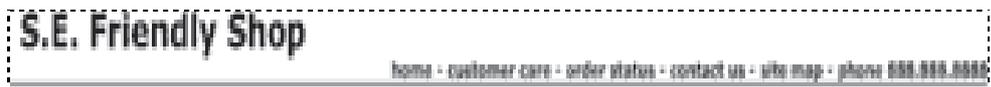


Figure 2.6: Header with navigation links (Source: Boggs, 2000)

- ii. Left side navigation (see Figure 2.7): The advantage of the left hand side navigation is its ability to solve two common e-commerce issues namely 'usability' and 'visibility'.

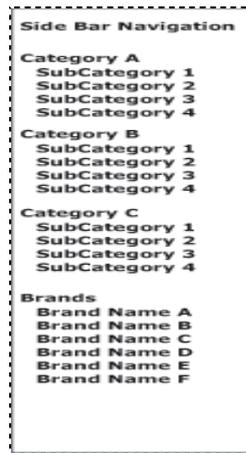


Figure 2.7: Left hand side navigation (**Source:** Boggs, 2000)

- a. Usability: A left hand side navigation bar will provide a level of clarity to the website. Website visitors will know how to easily navigate a website, as long as the navigation is kept consistent throughout.
- b. Visibility: The use of a mini website map on the left side of each page makes it easy for search engines to crawl and then index webpages.
- c. Homepage content area: For search engine friendly design, content is a very important factor. The content should read well and contain targeted keywords. Google normally uses the content on the page for the description in the results page (Boggs, 2000).

According to Boggs (2000), search engine optimisation covers the primary elements for building a search engine friendly e-commerce homepage. The header, left hand side navigation and middle area components of the homepage are all important for building an e-commerce website which has a good internal linking structure.

The WWW provides a vehicle for implementing a public identity in new innovative ways because it is live, interactive, and global in reach (Elderbrock *et al.*, 1997: 146). These authors provide some guidelines to creating a WWW identity:

- i. Be flexible – the WWW identity does not have to be the same as the corporate identity.
- ii. Be consistent – the identity should be consistent throughout the site.
- iii. Be clear – the mission and the impression one wants users to have must be clear throughout the development, the launch and consequent advertising.

Once the overall mission of the site has been decided upon, the focus is to get a target audience to the site (not just visitors) and, once they have visited, to keep them coming back. It is important to decide whether the target audience is going to be international or local and to keep the demographics of the current users in mind. To generalize, the users of the Internet tend to be from high quality niche markets (Van Steenderen, 2001). All the factors above pertaining to website visibility design forces designers to use graphics to attract users, however the same graphics could be harmful to search engine visibility.

2.5.2 Search engines optimisation strategies

Search engine optimisation (SEO) is the process of designing, writing, coding (in HTML), programming and scripting an entire website so that there is a good chance that the webpages will appear at the top of search engine queries for selected keywords (Pfattenberger, 2001). Optimisation serves as a mechanism of helping potential customers find a website. Figure 2.8 shows some elements of a successful search engine optimisation campaign (Position Research, 2002).



Figure 2.8: Elements of ranking strategies (**Source:** Position Research, 2002)

A SEO campaign may be broken down into six (see Table 2.2) basic functions (Position Research, 2002).

1 & 2	Strategic Analysis	Analysis of optimal theme-word & link structure of competing sites
3	Link Optimisation	Add relevant external inbound links and outbound links to and from other websites
4	Page Optimisation	Optimisation using theme-word for high ranking
5	Result reporting	Report on the ranking results and campaign status
6	Analyse & Revise	Analyse results and revise tactics to maintain or improve positioning

According to Thilmany (2003: 13), the presence of the following components ensures the best long-term search engine visibility:

- i. Text component.
- ii. Link component.
- iii. Popularity component.

Furthermore, according to Thurow (2002), all the major search engines (Google, FAST Search, MSN Search, and other Inktomi-based engines) use these components as part of their algorithms. Most webpages have different combinations of these components. Websites rank well in the search engine index when they have:

- i. All the components on their webpages.
- ii. Optimal levels of all the components.

2.6 WEBSITE DESIGN FOR SEARCH ENGINE VISIBILITY

When a search transaction is executed in a search engine, the matches should be ranked, so that the most relevant ones are displayed first. This way the first few sites that the user sees will contain the most pertinent information. A well designed website should therefore appear in the first 10 to 20 hits, specifically within the site's area of content (Furnis, 2001: 8). The problem with ranking though is that search engines have many different ways of applying ranking criteria. It is recommended that meta-tags must be complete with viable and applicable keywords. Because there are so many search engines, it is sometimes worthwhile considering using one of the many tools available on the Web to submit the site automatically to a batch of search engines.

Over the past few years, the Internet (in particular the World Wide Web) has become one of the primary areas for advertising and communication research (Chang & Zhao, 2004). Good website design has been found to give a website credibility. However, if a website does not appear on the screen fast enough, competitors are just 'a click away' (Ollins, 2002). Progressive SMMEs are utilizing web-based applications and technology to enhance operational control and real-time transactional data to improve planning and customer service (Rea-Smith, *et al*, 2002: 10). Perhaps most importantly, web-based information systems are helping some SMMEs handle transactions more efficiently in addition to giving their customers' shipments more visibility with features like real-time tracking and instantaneous proof of delivery.

Furnis (2001) recommends that website design be kept simple without frames, excessive graphics, database-generated pages, Flash home pages or lengthy JavaScript components. Each of these design strategies can present problems when trying to attain a good ranking with search engines. Van Greunen & Wesson (2004: 75) report that it is important to consider who the website is aimed at and what its purpose is.

A study by Alexander (2002a) shows that Americans experience 'search rage' when they do not find what they are looking for within 12 minutes.

This rage could cause them to move off to a different site with similar content and better navigation scheme, which could culminate in loss of income and customers for a badly developed site. Alexander (2002a) suggests that developers keep this in mind and recommends that they use simple, clear navigation with the inclusion of a website map that has links to all the pages in the website, which would be of value for both users and search engines.

Inan (2002: 35) is of the opinion that a website that looks good to a human, packed with useful information about products and/or services, but which is not visible on the search engines, can expect minimum traffic. There are many factors that contribute to the success of a website, especially if the website sells something over the Internet. Lu (1998: 168) agrees that a fast, clean, organized website with good content, crisp graphics and functional effect, will lure customers in and keep them longer than a busy, overstated and slow-loading website. Search engine databases are important entities, as they have the potential to generate large volume of user traffic. Navigation schemes assist publishers in their quest to ensure that their materials are found when appropriate searches are executed. Little has been researched on the effectiveness of different navigation schemes on website visibility to search engines (Sullivan, 2001).

According to Thurow (2003a: 45), to get the best search engine visibility, web designers should follow the Five Basic Rules of Web Design, which dictates that a website should be: Easy to read; easy to navigate; easy to find; constant in layout and design; quick to download. It is believed that by following these rules, a website will satisfy the target audience. Directory editors and search engines expect the same characteristics (Thurow, 2003b).

2.6.1 Factors that affect website visibility

According to Beatty (2003), factors affecting visibility include:

- i. Keyword density, i.e. the ratio of keywords to the total number of words on a page.
- ii. Keyword placement, i.e. the positioning of keywords in the title, other tags and in alternative areas, such as in image description.
- iii. Keyword prominence, i.e. how high a keyword appears on a page.
- iv. Link popularity, i.e. how is the website linked to other sites.

Search directories by contrast, require that the website operator selects the most appropriate directory categories for a website and write descriptions which accurately describe the content of the website (Lide & David, 2001: 30). To be accepted into a major directory, it is imperative that a website has content that show credibility in the specific field (Agillo, 2000: 25). A 'tip' page, 'how to' page or 'link' page can demonstrate the credibility of a product or service. Factors that affect directory placement are selecting the right category and writing a good description.

2.7 CONCLUSION

The Internet's connectivity network enables SMMEs to establish closer links with not only their customers, but also with search engines, directories, suppliers, partners, etc. by using websites as their primary source of information. It is only natural, with the advent of Internet technology throughout the world, that SMMEs should begin to use these networking facilities for both internal and extend communication. Electronic commerce is the latest economic profiteer for SMMEs who have embraced this opportunity.

The literature reviewed illustrated the following about website design:

- i. Reduce graphic content and focus on providing relevant information for visitors. Furthermore, the search engines cannot see graphic contents. If a page is loaded with graphics, search engines cannot interpret it as a result; a paragraph or two of text above the images that describes what the page is about should be included.
- ii. Build confidence by interacting with website visitors and anticipating and answering their questions.

- iii. Focus on information delivery and website architecture (a simply designed site) and exceptional writing.

The following are recommended to design a website for search engine visibility:

- i. Keep website design very simple.
- ii. Do not use:
 - a. Frames,
 - b. excessive graphics,
 - c. database-generated pages,
 - d. Flash home pages, or
 - e. lengthy JavaScript.

Each of these design strategies can present problems for search engine crawlers.

A simple, clear navigation structure that includes a site map link to all the pages in the website is valuable to both users and search engines.

Developers should avoid a technology that is incompatible with search engines. Many sites are being built with no preconceived planning for search engine spiders, and often contain dynamic content that cannot be indexed by search engines. It is recommended that websites should not be created using frame sets or dynamic URL's containing %, \$, ?, &. The search engines will often exclude these webpages. Using too many graphic images increases load time and does not benefit search engines, as they cannot 'crawl' images. Some of the programming technologies can be problematic to spiders, including:

- i. Complex JavaScripts.
- ii. Dropdown menus.
- iii. Image maps.
- iv. Flash.
- v. Framesets.
- vi. Java applets.
- vii. Dynamically generated webpages.

Optimally crafted navigation or link structures will increase the page rank of target pages. This increased page rank provides better opportunities for higher ranking within search engine result pages. It is important to note that the way human visitors navigate through a website can be different from that of a spider. A visitor can navigate through links that are constructed inside scripts - search engines can only navigate through HTML links. Optimal navigation architectures can be constructed to satisfy the need of both visitor and search engine at the same time.

CHAPTER THREE: METHODOLOGY

3.1 DISCRIPTION AND MOTIVATION OF RESEARCH DESIGN

Welman & Kruger (1999: 46) define research design as the plan according to which research participants (subjects) are obtained, and information is collected from them. Welman & Kruger (1999: 68) distinguish between four different types of research designs:

- i. Experimental research: Involves measuring the dependent variable before intervention, pre-measurement, and measuring it after the intervention, (post-measurement, single-group design).
- ii. Quasi-experimental research: Researcher cannot randomly assign subjects to different groups.
- iii. Non-experimental research: Neither random assignment, nor any planned intervention occurs in non-experimental research.
- iv. Qualitative research: can be used in the description of groups, (small) communities and organisations.

Mouton (2001: 35) developed a classification framework of design types that is depicted in Table 3.1.

Dimension	Type
Dimension 1: Ranging from empirical to non-empirical (conceptual)	Empirical, Non-empirical
Dimension 2: Primary or new data collected versus analysing existing data	Primary, Secondary, Hybrid
Dimension 3: Type of data - from numeric to textual	Numeric, Textual, Combination
Dimension 4: Degree of control or structure in design	High control, Medium control, Low control

A comprehensive research design also suggests the methods in terms of which statistical analysis is done on the data obtained (Welman & Kruger 1999; 65).

3.2 RESEARCH DESIGN AND METHODOLOGY

Babbie & Mouton (1998: 138) defines exploratory research as research into an area that has not been studied and in which a researcher wants to develop initial ideas and a more focussed research question. An exploratory research project was initiated with the intention to develop initial ideas and a more focussed research hypotheses based on academic literature. During this first stage of the project, a literature review of advantages and disadvantages of navigation schemes was conducted. The objective was the development and clarification of ideas and the formulation of research questions and hypotheses. The method used was to study secondary data sources. Two common navigation schemes were selected, and used for the second stage of the research. This second stage followed an experimental design to address the research hypothesis. Two websites (participants) were designed and submitted to crawler based search engines. Primary data was collected and quantitative statistical methods were used to analyse and interpret the results.

- i. Subjects of the study: Two e-commerce websites were developed, one with a navigation scheme based on text links (see Appendix B), and the other with a navigation scheme using graphic links in the form of buttons and image-maps (see Appendix A). The two URL's were submitted to a number of crawler-based search engines (research participants, 'subjects') namely Search Engine.co.za, Ananzi, Google, Yahoo!, AltaVista and DMOZ.
- ii. Units of analysis: The units of analysis of the research are the visits to e-commerce websites using two navigation schemes.
- iii. Sampling methods: Two types of commonly used sampling methods exist, namely probability and non-probability sampling (Welman & Kruger, 1999: 60).

From the exploratory research findings, two commonly used navigation schemes were selected from the population of website navigation schemes. Experimental research was conducted to analyse the impact of each on website visibility to search engines. All visits to the websites between October 2004 and June 2005 have been recorded.

A purposive sampling method was used to select the navigation schemes. Purposive sampling allows researchers to rely on their experience, ingenuity and/or previous research findings to deliberately obtain units of analysis (Welman & Kruger, 1999: 63). Purposive sampling is a non-probability sampling method.

3.2.1 Procedure

The process followed to implement the research methodology was as follows:

- i. Review available literature on website navigation architecture, pertaining to advantages and disadvantages of text-based and graphic-based systems.
- ii. Design two websites (see Appendix A and B), each with different navigation schemes. The two websites are probabilistically equivalent to each other (e.g. selling books over the Internet), and have the same functionality, keywords and content. One contains text links (experimental group) and the other graphic links (the control group).
- iii. Host and register both websites with search engines simultaneously.
- iv. Monitor and record crawler visits to both sites.
- v. Compare the websites using an industry standard program that provides statistical reports on website ranking, positioning and indexing.
- vi. Statistically summarize and compare the data using a t-test.
- vii. Perform hypothesis testing to conduct a scientific procedure for assessing the validity of a claim about the population.
- viii. The following scientific procedure was followed:

- a. Write the claim mathematically and identify H_0 and H_1 .
- b. Find the critical value(s) and identify the rejection region(s).
- c. Find the standardized test statistics.
- d. Decide whether to reject or fail to reject the null hypothesis and interpret the decision in the context of the original claim.

The data collected was tabulated and described using tables and graphs. A comparative analysis was performed to determine whether the use of links substantially improved accessibility, ranking and visibility of the e-commerce websites to the search engines.

3.2.2 Population and sample types

According to Welman & Kruger (1999: 218), the population is the study object, which can be individuals, groups, organisations, human products and events or other conditions to which they are exposed. In the instance of this project, the population of this study is website navigation schemes. Population numbers may sometimes be very large and therefore samples must be used to make inferences about the corresponding population properties. In such cases, inferential statistics may be used (Welman & Kruger 1999: 219). The size of the population usually makes it impractical and uneconomical to involve all the members of the population in a research project (Mouton, 2001: 83). In such instances, the researcher relied on the data obtain for a sample of the population.

3.2.3 Sampling Methods

In a research design, the following need to be specified (Babbie & Mouton 1998: 138):

- i. The number of groups that should be used in order to decide which statistical technique to use.
- ii. Whether or not these groups are to be drawn randomly from the population involved.
- iii. What exactly should be done with the groups in the case of experimental research?

3.2.4 Data Collection methods

When collecting data for quantitative studies, it is impossible or impractical to consult all the people (i.e. population) in a specific category as indicated by the research project (Struwig & Stead, 2001: 109).

3.2.4.1 Type of data

Larson & Farber (2003: 8), identified sets of data to consist of two types, namely:

- i. Qualitative data: Consist of attributes, labels, or non-numerical entries.
- ii. Quantitative data: Consist of numerical measurements or counts.

The type of data decision focuses on whether the research data is to be numeric or non-numeric (Babbie & Mouton 1998: 138). Data can also be gleaned in the form of non-verbal (questionnaires) or verbal data (interviews). In this study quantitative data was collected, textual data was gathered during the literature review, and numeric data was collected during the empirical study.

3.2.4.2 Source of data

There are two possible sources of data in business related research (Babbie & Mouton 1998: 85), namely:

- i. Primary data - the researcher personally collects primary data.
- ii. Secondary data - data which has been collected by someone other than the researcher.

In this study, a combination of secondary data for exploratory design (first stage) and primary data for experimental design (second stage) was used. Primary data was collected using two industry standard programs.

Data sources can be classified into various categories (Mouton, 2001: 64), including:

- i. Observation: Systematic observation under controlled experimental or laboratory conditions.
- ii. Self-reporting: Personal and group face-to-face interviewing, telephone interviews, mail and electronic surveys.
- iii. Archival/documentary sources: Historical documents, diaries, letters, speeches, literary texts, narratives, official memoranda, business plans, annual reports, medical records, etc.
- iv. Physical sources: Blood samples, cell tissues, chemical compounds, materials etc.

Using Mouton's definition of observation, this approach of data collection was used. It was systematic and controlled since the two industry standard programs AWstats, and Internet Visibility Services™, (Real World Technology) recorded all visits (requests and page requests) to the two websites over a period of eight months. These programs were used to provide data of website performance in terms of ranking, visibility, and position. The two websites were monitored and visited on a regular basis over an extended period.

The AWStat program provided the following statistics about the target website:

- i. When: Summary, days of month, days of week, hours.
- ii. Who: Robots/spiders visits, countries, full list, hosts, last visits.
- iii. Navigation: Visit duration, file type, viewed, full list, operating system and browser.
- iv. Refers: Origin, referring SE, referring sites, search keyword and key phrase.
- v. Number of visits.
- vi. Pages visited.
- vii. Hits.
- viii. Bandwidth.

The Internet Visibility Services™ analysis results were used to further analyse websites. This analysis primarily shows how visible a website is to

search engines in terms of: topic clarity, subject representation, header representation, and overall visibility.

All these issues were evaluated on a scale of 'Excellent' to 'Fail', represented by a range of 100% to 0% respectively.

3.2.4.3 Level of measurements

Another data characteristic is the data's level of measurement. The level of measurement determines which statistical calculations are meaningful, and has four levels from highest to lowest (Larson & Farber, 2003: 9) described as follow:

- i. Nominal level of measurement: Is qualitative only. The data at this level is categorised using names, labels, or qualities. No mathematical computations can be done at this level.
- ii. Ordinal level of measurement: Is qualitative or quantitative. Data at this level can be arranged in order, but the differences between data entries are not meaningful.
- iii. Interval level of measurement: Is quantitative. The data can be ordered and meaningful differences between data entries can be calculated. At the interval level, a zero entry simply represents a position on a scale; the entry is not an inherent zero.
- iv. Ratio level of measurement: Is similar to the interval level, with the added property that a zero entry is an inherent zero. A ratio of two data values can be formed so one data value can be expressed as a multiple of another.

For this research, the interval level of measurement was used, because of the quantitative nature of the data. Furthermore, meaningful calculations can be performed on the data entries.

3.3 CONCLUSION

The first phase of this study follows an exploratory design. At this stage, available literature on navigation schemes was reviewed and a sample of

two commonly used website navigation schemes was selected. The second stage was an experimental research design. The research design was carried out in six steps:

- i. Identify the variables of interest, and the population of the study.
- ii. Develop a detailed plan for the collection of data.
- iii. Collection of the data.
- iv. Describe the data using descriptive statistics techniques.
- v. Interpret the data and make decisions about the population using inferential statistics.
- vi. Identify any possible errors.

The author developed two websites, one with an embedded text navigation scheme and the other with graphics in the form of 'Buttons' and 'Images' (see Appendix A and B). The FrontPage webpage authoring program was used, and both websites were hosted with a professional, commercial company specializing only in website hosting.

The two websites were submitted to a number of crawler based search engines. Ananzi did not index the two websites, but Google, AltaVista, SearchEngine, TheIndex and AlltheWeb did. Two industry standard programs were used to monitor the websites for a period of eight months, and the following data was recorded and tabulated: 'Number of requests', 'number of pages requested', and 'number of bytes transferred'. This information was used to compare the visibility and ranking of the two websites. Industry standard programs as well as search engines use this information to determine the visibility and ranking of a website.

Descriptive statistics were used to interpret and analyse the results. The data from the tables was graphed in order to compare the difference between the variables. Furthermore, a t-test comparison of mean, mode and standard deviation of 'number of hits' and 'pages transferred' for the two websites was performed to determine whether there is a difference between the two variables of the research hypothesis.

To achieve the above objective, the following procedure was followed:

- i. The assumption was made that there is no difference in the mean ranking of the two samples, that is $\mu_1 - \mu_2 = 0$, where μ_1 is the sample data of the graphic-based website and μ_2 is the sample data of the text-based website.
- ii. A two-sample t-test was performed to test the hypotheses that there is no difference in ranking and visibility of SMME websites to search engines between pages that use text-based navigation schemes and pages that use graphics-based navigation schemes.
- iii. The resulting two-sample test statistics $\bar{x}_1 - \bar{x}_2$ were used, where \bar{x}_1 is the mean ranking of graphic-based website and \bar{x}_2 is the mean ranking of text-based website.
- iv. A conclusion was reached (based on the t-test) from the website statistics provided by the two industry standard programs used to collect the data and the decision to accept or reject the null hypothesis.

CHAPTER FOUR: RESULTS AND ANALYSIS

4.1 INTRODUCTION

All fieldwork culminates in the analysis and interpretation of some set of data, be it quantitative survey data, experimental recordings, historical and literary text, qualitative transcripts or discursive data (Mouton, 2001: 39). Once a decision has been made on an appropriate research design and suitable means of measuring the relevant variables, the next step is to choose an appropriate statistical procedure to analyze the data obtained.

Analysis involves 'breaking up' the data into manageable themes, patterns, trends and relationships (Babbie & Mouton 1998: 38). The aim of an analysis is to understand the various constitutive elements of collected data. In this study, it was achieved through an inspection of the differences in the means of variables for two independent groups, and an evaluation to determine whether there are any patterns or trends that can be identified to establish themes in the data. The results would then provide feedback to the tenability or otherwise of the originally formulated research hypothesis.

4.2 DATA COLLECTED

4.2.1 Data collected from AWStats

The data collected from the AWStats program was tabulated as shown in Table 4.1 to Table 4.11. Table 4.1 to Table 4.3 show the statistics in terms of server visits, and Table 4.4 to Table 4.6 show statistics in terms of user visits. This information is required to analyse and compare the two websites so as to draw a conclusion from the search engine visibility reports. Table 4.1 summarizes visits to ebooks1, the graphic website. The monthly report identifies activity for each month in the report time frame. Each page clicked can result in several server requests as the images for each page are loaded.

Month	Percent age of requests	Number of requests	Percent age of page requests	Number of page requests	Percentage of bytes	Number of kbytes transferred
Oct-04	25.78%	124	26.35%	34	32.02%	179.121
Nov-04	3.53%	17	5.42%	7	7.52%	42.075
Dec-04	0%	0	0%	0	0%	0
Jan-05	35.13%	169	27.13%	35	31.74%	177.563
Feb-05	3.95%	19	7.75%	10	6.36%	35.615
Mar-05	2.49%	12	3.87%	5	0%	0
Apr-05	15.38%	74	8.52%	11	4.99%	27.934
May-05	0.00%	0	0.00%	0	0%	0
Jun-05	13.72%	66	20.93%	27	17.35%	97.051
Total		481		129		559.359

Table 4.1 shows that the most active month was January 2005. In this month, 35 pages were requested, 177,563 bytes were served and 169 requests handled. Data contained in Table 4.1 was provided by AWStats (www.ebooks1.co.za) and consist of three categories:

- i. Percentage and number of requests. Number of request indicates the number of times the website has been visited in that month.
- ii. Number of pages requested. This indicates the number of linked pages on the website that the user clicked to navigate.
- iii. Percentage and number of bytes transferred. These indicate the number of bytes, referring to the size of the files transferred in megabytes.

This information is used to determine the visibility of the website as well as ranking by most standard programs and search engines. Table 4.2 shows the same data that has been collected for the text-based website (www.ebooks2.co.za) from the AWStats program.

Month	Percentage of requests	Number of requests	Percentage of page requests	Number of page requests	Percentage of bytes	Number of kbytes transferred
Oct-04	18.86%	73	15.28%	24	22.94%	61.772
Nov-04	3.10%	12	5.73%	9	7.63%	20.551
Dec-04	0%	0	0%	0	0%	0
Jan-05	37.20%	144	27.38%	13	38.07%	102.511
Feb-05	4.13%	16	8.28%	13	3.31%	8.924
Mar-05	2.06%	8	3.18%	5	1.07%	2.888
Apr-05	24.28%	94	22.29%	35	16.99%	45.764
May-05	0.77%	3	1.91%	3	0.85%	2.298
Jun-05	9.56%	66	15.92%	25	9.10%	24.520
Totals		416		127		269.228

The data reflects that the most active month was January 2005. In this month 13 pages were requested, 102,511 bytes were served, and 144 requests handled.

Table 4.3 combines the data from Table 4.1 and Table 4.2 in order to summarise it.

Features	ebooks1 (Graphics)	ebooks2 (Text)
No. of Requests	481	416
No. of Pages	129	127
Kbytes transferred	559.359	269.228

Tables 4.4 and 4.5 reflect the number of visits, pages visited, number of hits, as well as bandwidth collected for the graphic-based website (ebooks1) and for the text-based website (ebooks2) respectively. The data in the two tables maps to the data in Table 4.1 for ebooks1 and for ebooks2 respectively.

Table 4.4: Search Engine visits statistics for graphic-based website Source: (AWStats, 2005)				
Month	No. Visits	No. Pages	Hits	Bandwidth
Jan-05	8	50	169	173.4 kb
Feb-05	7	10	19	34.78
Mar-05	3	5	12	0
Apr-05	4	15	74	27.28
May-05	0	0	0	0
Jun-05	15	23	62	88.47
Total	37	103	336	323.93KB

Table 4.5: Search Engine visits statistics for text-based website (Source: AWStats, 2005)				
Month	No. Visits	No. Pages	Hits	Bandwidth
Jan-05	13	49	144	100.11 kb
Feb-05	7	13	16	8.71
Mar-05	5	5	8	2.82
Apr-05	4	39	94	44.69
May-05	5	3	3	2.24
Jun-05	15	25	37	23.95
Total	49	134	302	182.52 KB

Table 4.6 reflects the consolidated statistics of Table 4.4 and 4.5.

Table 4.6: Consolidated Search Engine statistics for graphic and text based-websites		
Feature	ebooks1 (Graphic)	ebooks2 (Text)
No. Visits	37	49
No. Pages	103	134
Hits	336	302
Bandwidth	323.93	182.52

Table 4.7 compares data from the most active month for both websites as provided by the AWstats reports.

Table 4.7: Comparison for the most active month (January)		
Features	ebooks1 (Graphic)	ebooks2 (Text)
Pages Sent	35	43
Kilobytes served	177.563	102.511
Requests handled	169	144

Table 4.8 compares the monthly average data for both websites as collected from the AWstats reports.

Table 4.8: Comparison of the monthly average		
Features	ebooks1 (Graphic)	ebooks2 (Text)
Pages Sent	14.33	17.44
Kilobytes served	62.151	29.914
Requests handled	53.44	43

4.2.2 Data collected from Internet Visibility Services™

Internet Visibility Services™ is a program that primarily shows how visible a website is to search engines. Internet Visibility Services™ claims to provide all the tools and services needed to transform a website into a successful online business. The program comprises the entire scope of scientific website creation, optimisation, sales and marketing services.

The data provided by Internet Visibility Services™ was recorded as shown in Table 4.9 to Table 4.11. The tables provide data on visibility of websites to search engines in terms of topic clarity, subject representation, header representation, and overall visibility.

Table 4.9 contains data for the graphic-based (ebooks1) website as collected from the Internet Visibility Services™.

Table 4.9: Internet Visibility Service™ ranking for graphic-based website (Source: Internet Visibility Services™, 2005)				
Rating	Topic Clarity	Subject Representation	Header Representation	Overall visibility
Excellent				
Acceptable	57%		62%	
Borderline				
Fail		0%		16%

The website's page headers are missing some components. The visibility to search engines is very poor, returning 16%.

Table 4.10 contains data for the text-based website (ebooks2), as collected from Internet Visibility Services™.

Table 4.10: Internet Visibility Service™ ranking for text-based website (Source: Internet Visibility Service™'s, 2005)				
Rating	Topic Clarity	Subject Representation	Header Representation	Overall visibility
Excellent				
Acceptable	50%		52%	
Borderline				
Fail		0%		13%

The data from the graphics-based website also showed that the website's page headers are missing some vital components, but is in general acceptable. The visibility to the search engines is poor at 13%.

Table 4.11 contains the combined data of Table 4.9 and Table 4.10, so as to compare the two websites.

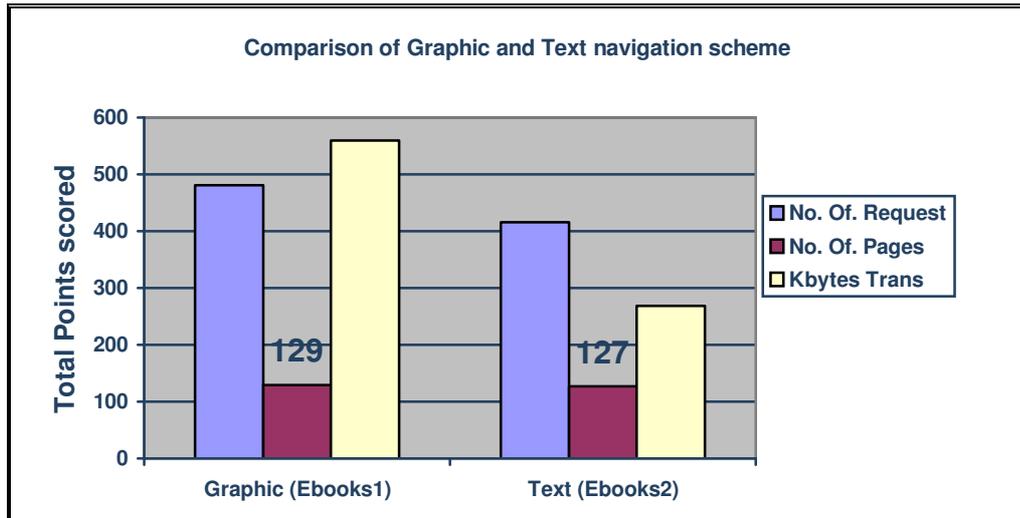
Table 4.11: Combined data for text versus graphics websites		
Features	ebooks1 (Graphics)	ebooks2 (Text)
Topic clarity	57	50
Subject Representation	0	0
Header Representation	62	52
Overall Visibility	16	13

The data shown in Table 4.11 reflects higher values for the graphic-based website than the text-based website in all cases except the centre row. The topic clarity of the graphic-based is seven points higher than that of the text-based website. Header representation has a 10 points difference, while the overall visibility varies by only three points. In all three cases, ebooks1 seems to be more visible than ebooks2.

4.3 RESULTS ANALYSIS AND INTERPRETATION

The consolidated results from Table 4.3, Table 4.6 and Table 4.11 are represented graphically (see Graph 4.1 to Graph 4.3). A comparative analysis was performed between text and graphic navigation schemes to determine whether the use of links substantially improve accessibility, ranking and visibility to search engines.

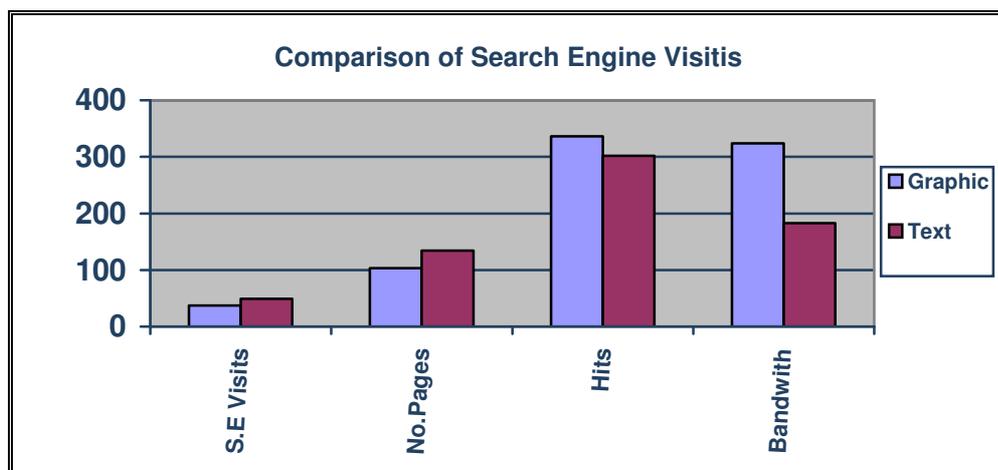
The consolidated monthly statistics data from Table 4.3 compares a graphic-based and text-based website in terms of number of requests (Hits), number of pages, kilobytes transferred, and total points scored. This data is reflected in Graph 4.1 for the comparison of the relationship between the variables of the study.



Graph 4.1: Consolidated monthly statistics for graphic vs. text websites

Graph 4.1 shows that the graphic-based website has a higher value for all four attributes. The graphs reflect that graphic-based website has higher values on every data point. This could mean that crawlers had favoured the graphics-based websites.

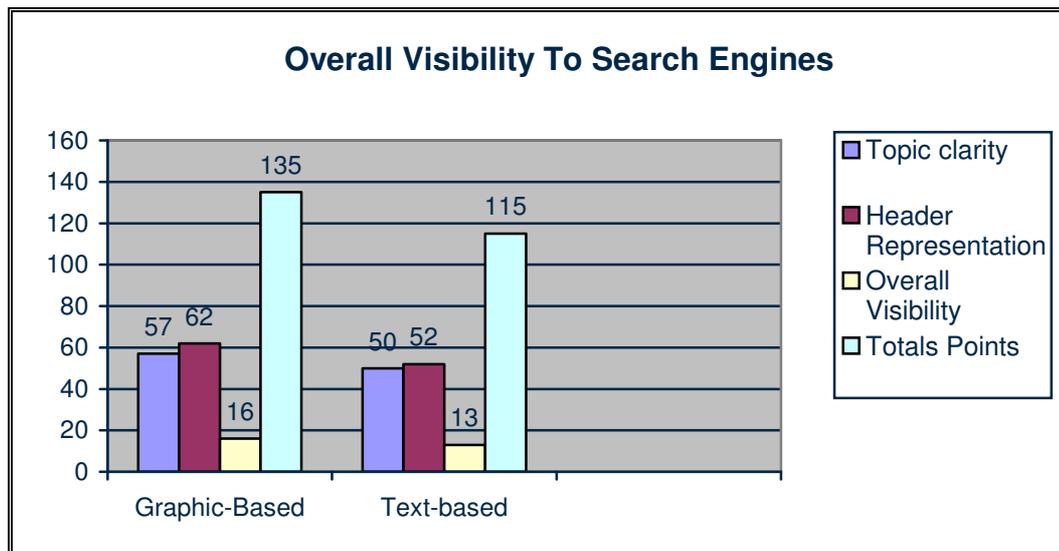
The consolidated search engine visit data from Table 4.6 was represented in Graph 4.2, comparing the number of search engine visits, number of pages, hits, bandwidth, and total points scored for text and graphics.



Graph 4.2: Comparison of search engines statistics

The text-based website had a higher number of search engine visits and number of pages requested. However, the number of hits, the bandwidth and the total points scored was higher for graphics.

The consolidated visibility data from Table 4.11 was graphed as shown in Graph 4.3, whereby the visibility of graphic-based website to text-based website was compared.



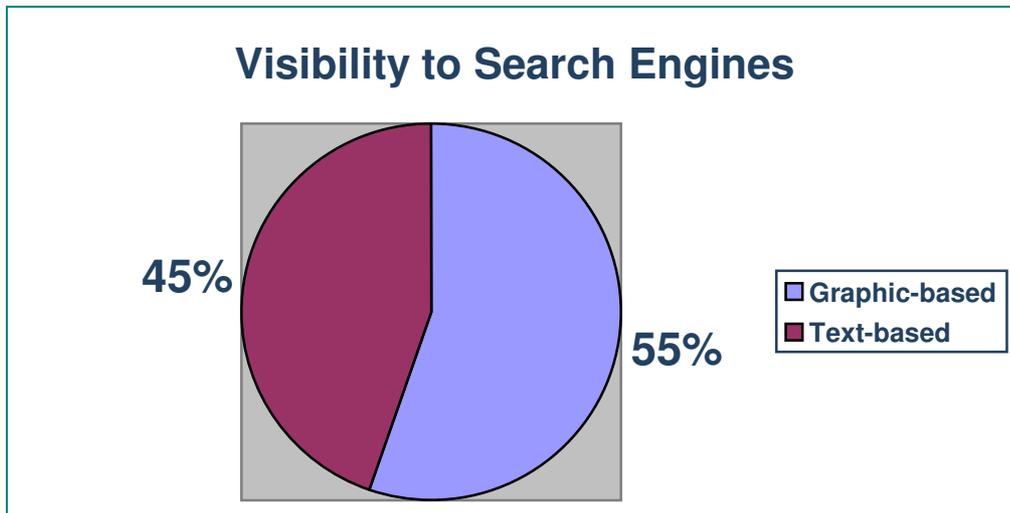
Graph 4.3: Internet Visibility Service™ analysis of graphic versus text-based websites

From Graph 4.3, it is evident that graphic-based websites rank higher than text-based websites. Graphic-based websites appears to be higher in all of the factors considered for ranking of a website, namely:

- i. Topic clarity: Graphic-based is 57 compared to 50 of text-based.
- ii. Header representation: Graphic-based is 62 compared to 52 of text-based.
- iii. Overall visibility: Graphic is 16% visible to search engines compared to 13% of text-based.

These statistics are similar to the statistics provided by AWStats reports (see Appendix E and Appendix F) as well as visibility report provided by Internet Visibility Service™ (see Appendix G and Appendix H), as shown in Graph 4.1 and 4.2.

Graph 4.4: Overall visibility to search engines



From the overall visibility graph (see Graph 4.4), it is evident that the graphic-based website is 10% more visible than the text-based website. This indicates a noticeable difference in the visibility of the two websites to search engines when the different navigation schemes are used.

4.4 DESCRIPTIVE STATISTICS

To perform a *t*-test hypothesis with independent samples, different techniques can be used (Larson & Farber, 2003). Descriptive research involves comparing the mean of one group with the mean of another (Welman & Kruger, 1999). Furthermore, to conduct the test the following conditions are required:

- i. The samples must be randomly selected.
- ii. The samples must be independent.
- iii. Both populations must be normally distributed.

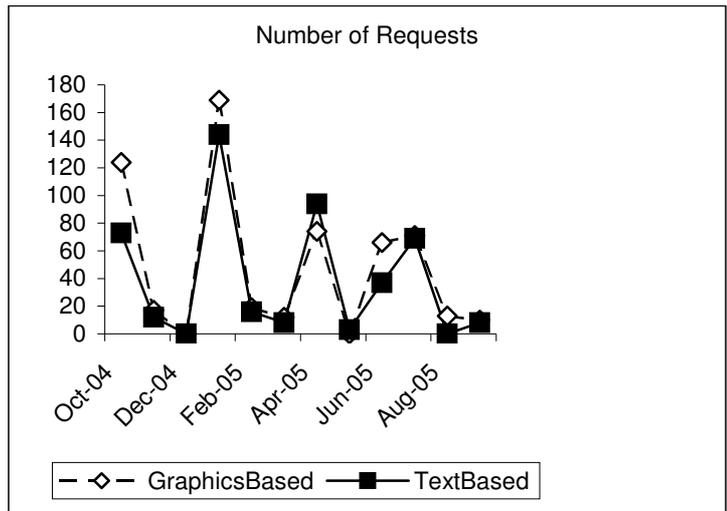
Descriptive statistics were used to further analyse the result. A *t*-test hypotheses testing was used to measure the difference between two sample means, because this test enables the user to determine whether

two groups have equivalent or different mean scores. Furthermore, the t-test determines whether or not an observed difference in the means of two groups is sufficiently large to be attributed to a change in some variable or if it merely could have taken place by chance.

The data obtained from AWStats (see Appendix G and H) was used with the SPSS program to produce detailed statistics in terms of mean, median, variance, standard deviation (see Appendix I) for preparing the calculation and comparative analysis of the two groups.

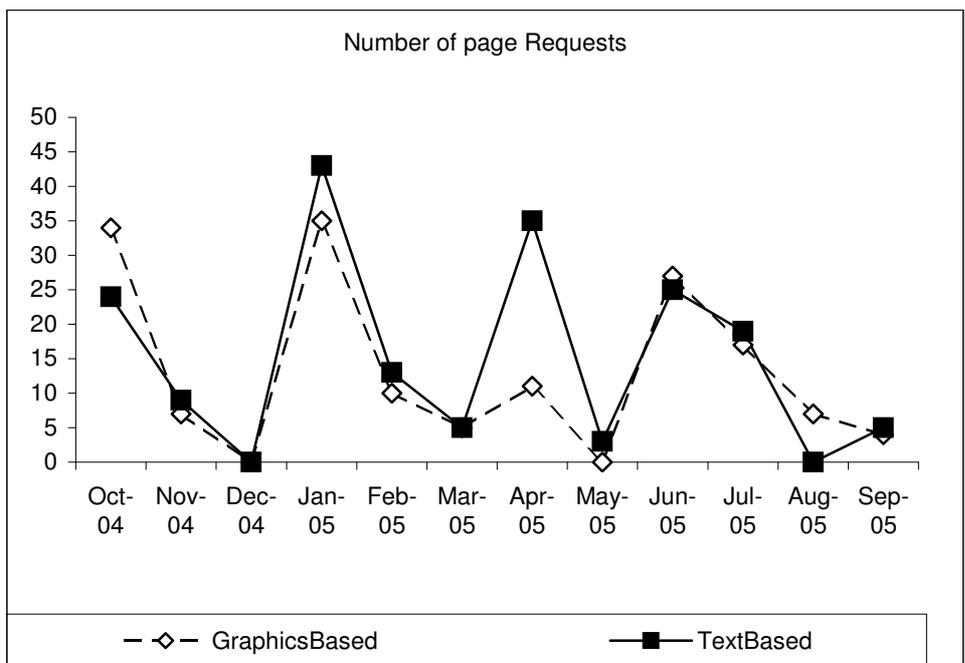
Table 4.12: Consolidated comparison of text vs. graphic websites						
Date	Number of requests		Number of page requests		Number of bytes transferred	
	Text Based	Graphics Based	Text Based	Graphics Based	Text Based	Graphics Based
01-Oct-2004	73	124	24	34	61772	179121
01-Nov-2004	12	17	9	7	20551	42075
01-Dec-2004	0	0	0	0	0	0
01-Jan-2005	144	169	43	35	102511	177563
01-Feb-2005	16	19	13	10	8924	35615
01-Mar-2005	8	12	5	5	2888	0
01-Apr-2005	94	74	35	11	45764	27934
01-May-2005	3	0	3	0	2298	0
01-Jun-2005	37	66	25	27	24520	97051
01-Jul-2005	69	71	19	17	80423	103410
01-Aug-2005	0	13	0	7	0	37421
01-Sep-2005	8	10	5	4	6680	17628

Graphs 4.5 to 4.7 show the monthly statistics for the number of requests, number of pages requested and the number of bytes transferred, respectively.

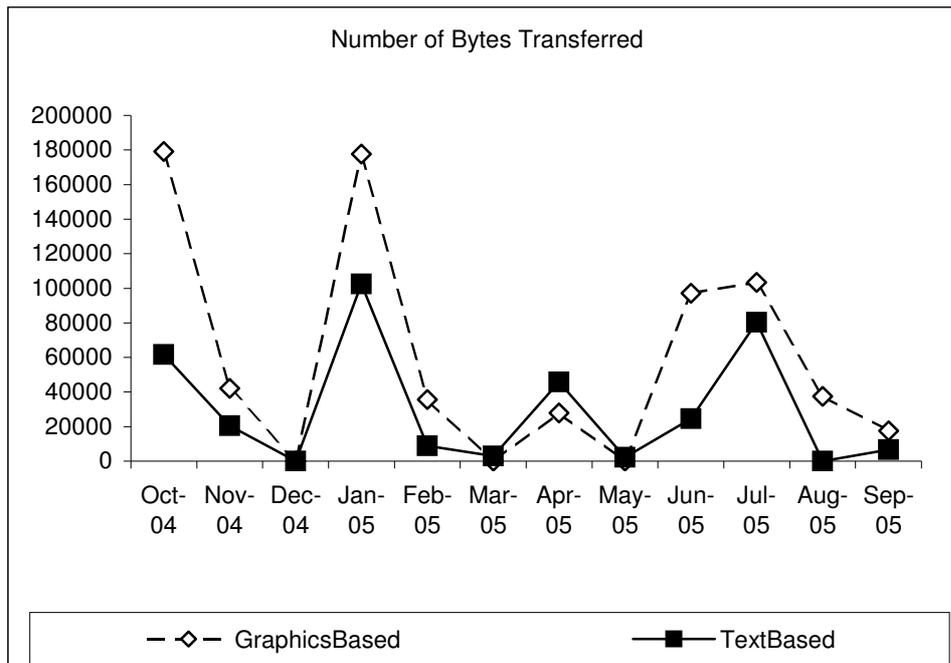


Graph 4.5: Consolidated comparison of number of requests

It appears from Graph 4.5 that the graphic-based website has a higher number of request (hits) for October, January, June and August, while the other site has a higher number of request in April, July and September. However, the overall pattern of the month-to-month data is similar. Similar patterns can also be observed for the number of page requests and the number of bytes transferred.



Graph 4.6: Consolidated comparison of number of page requests



Graph 4.7: Consolidated comparison number of bytes transferred

Excluding April, the graphic-based website has a higher number of bytes transferred as compared to the text-based website. This is due to the fact that graphics images are larger than text strings, and explains the larger difference between points on the two graphs.

The data from Table 4.12 was used to calculate the mean number of request (Hits) and mean number of pages requested for the two groups as shown in Table 4.13 and Table 4.14.

Web Type		Mean	Median	Std Deviation
Number of Request	Graphic	47.9167	18	54.2058
	Text	38.6667	14	46.3452
Number of Page Request	Graphic	13.0833	8.5	12.4496
	Text	15.0833	11	14.1900
Number of Bytes Transferred	Graphic	59,818.1667	36,518	64,889.444
	Text	29,694.25	14,738.5	34,989.5588

Although there is a difference between the means in terms of pages sent, pages requested and bytes transferred, the difference appears to be insignificant. The mean number of requests for Graphics (47.92) is higher than that of Text (38.67) but the mean page requests for Graphics (13.08) is less than that of Text (14.19). The mean number of bytes transferred appears to be higher for Text (34989.56) compared to 59818.17 for Graphics.

An independent sample t-test (not assuming equal variance) was performed on the data using the SPSS program. This showed that there were no significance differences (p-values > 0.05) in any of the variables between the graphics-based and the text-based web pages. This could be due to the relatively small sample size and the time frame used in this study.

	T	Df	p-value (2-tailed)	Mean Difference	Std. Error Difference
Number of requests	.449	22	.658	9.250	20.588
Number of page requests	-.367	22	.717	-2.000	5.449
Number of bytes transferred	1.415	17	.175	30123.917	21281.661

4.5 CONCLUSION

According to Graphs 4.1 and 4.3, graphic-based websites appear to be ranked higher than text-based websites when considering pages visited, hits, bandwidth, as well as overall average. This could transpose into the fact that graphic-based websites are more visible to search engines than text-based websites, although this difference is not significant.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

5.1 INTRODUCTION

The research hypothesis (See Paragraph 1.5) states that there is a difference in ranking and visibility of SMME websites to search engines between pages that use text-based navigation schemes and pages that use graphics-based navigation schemes. This hypothesis was tested using independent samples t-tests for the difference of means for the three variables: number of requests, number of page requests and number of bytes transferred. In all three cases there was no significant differences found between text-based and graphics-based websites. Therefore this research hypothesis must be rejected.

Two websites were developed with the same functionality, content and keywords, but different navigation architectures. In terms of the data extracted from the standard programs, it was evident that the navigation schemes used contributed to the search engine visibility of the websites, and that the graphic-based website was ranked higher than the text-based website. However, the results from the t-test show that there is no significant difference between the means of each of the variables.

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

It is therefore necessary for web developers to include a text navigation scheme so that the spiders have the means to record the words on the webpages. This transposes in having two forms of navigation on a website, namely: one that pleases target audiences visually, and one that the search engines crawlers can follow.

The basic factors impacting upon a webpage's ranking include the words in the title and keyword Meta-tags, word frequency in the document, and document length. Links to a website also boost website rankings, as well as keyword relevancy to page content.

Some possible reasons that a website may not be indexed include, from the literature reviewed:

- i. The website uses frames which are not indexable, as they require special formatting.
- ii. The server was not connected to the Internet, malfunctioned, was busy, or non-operational during crawling.
- iii. The search query entered does not reflect the website's content.
- iv. The host server contains the robots.txt exclusion tag, where this file has instructed the crawler not to index the website.
- v. The URL is submitted without the 'http: //' prefix.
- vi. The URL contains special characters such as: ?, =, %, &.
- vii. Pages requiring a cookie will not be indexed.

Meta-tags play an important role when a web search engine inspects a webpage and prepares to catalogue and rank it. Meta-tags are located in the header of webpage's HTML source. The most important meta-tags for search engine ranking are keyword meta-tags, description meta-tags, and robot meta-tags.

5.2 PROPOSED SOLUTION

The factors discussed below should be considered as contributing towards a successful e-commerce website.

5.2.1 Hypertext

An understanding of the benefits and disadvantages of using a hypertext navigation architecture will aid developers in their internal linking strategy. A good strategy is to use hypertext links as secondary navigation. With technologies such as Cascading Style Sheet (CSS), developers can

create attractive looking hypertext links as a primary navigation scheme. Finally, website developers might consider using a breadcrumb trail as an additional internal navigation scheme strategy to help targeted users keep track of where they are and to increase keyword rich anchor text.

5.2.2 Navigation Buttons

Graphical navigation schemes are as popular as hypertext navigation schemes. There are many benefits to using graphical navigation icons or buttons as primary navigation architecture - this include using high-impact fonts and colours that are not supported through regular HTML. Increasing the appearance of the links on a website could help to attract more clicks by the end user to the webpages of interest.

Claims have been made that search engines cannot index graphic images. These claims have been proven to be false. However, it is still recommended that, in all cases where graphics are used as navigation, it would be wise to use an alternate text tag within the source code of the image. The alt (alternative) tag gives the search engine the ability to assign meaning to the graphic image, in place of the anchor text. Hence, it is important to use keyword specific words that apply to the page linked to in the alternative tag.

5.2.3 Image Map Navigation

Image maps are assigned to single graphics that enable targeted users to access pages of a website by clicking on a portion (X and Y coordinate) of the graphic. Image maps are used for two primary reasons:

- i. To reduce file download size - one image takes less time to download than five images.
- ii. To simplify the coding of the links within complex graphics.

Image maps contain complex numbers and variables that prescribe where the user shall be taken. For this reason image maps are not search engine friendly - if used, developers should always use a search engine friendly secondary form of navigation, like hypertext links.

5.2.4 Flash and animation navigation scheme

Web designers can create, publish and transmit high definition, motion filled, colourful animations to a broad Internet base without running into the pitfalls of bandwidth limitations and browser support.

Flash websites tend to be one of the less popular types of websites in terms of search engine visibility. Most search engine spiders cannot read the contents within the SWF (file extension for published Flash files) files. Google and FastSearch are able to read the URLs within a Flash file, but nothing more at the time of writing. That is why it is important when creating Flash sites to also create a simple but text rich HTML website for crawlers and for WWW users that prefer the less graphics-rich media.

5.2.5 JavaScript or DHTML Dropdown or Pulldown Menus

Many sites utilize pulldown or dropdown menus to allow users to locate sections of a site quickly, and to allow targeted users to quickly jump from one section of the site to another section. The main problem with this type of navigation is that search engine spiders have difficulty reading the URLs within the JavaScript. In addition, usability experts tend to dislike JavaScript pulldown menus and recommend offering alternative navigation in addition to the JavaScript pulldown menu. It would be wise to use hypertext links as primary or secondary navigation when utilizing JavaScript pulldown menus for both usability and search engine visibility.

5.2.6 Framed navigation

Frames tend not to be search engine friendly for many reasons. One of the most obvious reasons is that it gives search engine crawlers a

challenge determining which file is the primary site content. However, search engines do not know how to index a frame-based site fully and it stays within the navigation frame, thereby missing out on all the content. The more frames utilized within a site the less search engine friendly the site is.

There is an acceptable standard available that can achieve the look of frames while making websites much more search engine friendly. This standard has been released by the World Wide Web Consortium (W3C) and is known as CSS or Cascading Style Sheets. If developers insist on using frames, there are some guidelines they should follow to help increase the search engine crawler's ability to index the pages.

As a result, to build a website that appeals to the user and to search engines it is imperative to design for maximum search engine visibility. According to the reports from the industry standard programs, the graphic-based website was ranked higher than the text-based website. These results are in contrast to the results gleaned from the literature review, which found that the text-based website should be ranked higher than the graphic-based website on crawler-based search engines.

5.3 CONCLUSION

The research in this dissertation had as its objective to determine the impact of the various kinds of website navigation schemes on website visibility. The literature review indicated that text-based links are easier for search engines crawlers to follow.

The research in this dissertation however shows that crawlers more readily follow graphic-based hyperlinks (see Table 4.11, and Appendix A and B). According to the statistical result however, there was little significant difference in visibility. This could be attributed to the small sample size used. This author is of the opinion that the research results seem to reject H_0 and accept H_1 . Interestingly the relationship appears to be contradictory to the perception created by the literature.

Under the proposed solution (see 5.2), the author has completed the discussion of the navigation schemes. There are many approaches to designing a navigation scheme. While each method has its advantages, the author recommends using secondary hypertext links, and building a simple site map that contains a link to each internal page of a website.

5.4 FUTURE RESEARCH

Future research on this topic could use a larger sample size and allow for a longer time frame for crawler visits. A larger sample of graphic-based and text-based websites could be obtained from the e-commerce audience, and a comparison of their visibility to a sample of crawler based search engines be conducted. The result could provide insight on whether there is a significant relationship between these two variables. Furthermore, future research on this field could look at the following navigation schemes and their impact on website visibility:

- i. Flash and animation.
- ii. JavaScript.
- iii. DHTML dropdown and pulldown menus.

LIST OF REFERENCES

Adam, S. & Dean, K. 1999. WebQUAL: An E-commerce audit. *Proceedings of the fifth Australian World Wide Web conference*, Australia, Southern Cross University, 11 – 16 June 1999:
<http://ausweb.scu.edu.au/aw99/papers/adam/>. [22 October 2004].

Agillo, I. 2000. A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources. *Journal of Management in Medicine*, 14(3): 13-28.

Alexander, J. 2002a. *How to Choose Keywords Before They Skyrocket in Popularity: Search Engines Workshop*.
<http://www.searchenginesworkshops.com/articles/response.html>. [27 June 2003].

Alexander, J. 2002b. *Ten Things You can do to Increase Visitor Response at your Website*.
<http://www.searchenginesworkshops.com/articles> . [27 June 2003].

Amazon. 2004. *Amazon*.
<http://www.amazon.com>. [28 November 2004].

Anon. 2000. BT technologist looks into the future of communications in new science Website. *Work Study*, 49(3): 22-29.

Baard, V. & van den Berg, A. 2004. Interactive information consulting system for South African small business – Part 1. *South African Journal of Information Management*, 6(2): 1-32.

Babbie, E. & Mouton, J. 1998. *The practice of social research*. 6th ed. Cape Town: Oxford University.

Barnes, J. & Vidgen, T. 2002. An interactive approach to the assessment of e-commerce quality. *Journal of Electronic Communication Research*, 3(3): 114-127.

Beatty, S. 2003. Fashion Tip: Get Online. *Wall Street Journal – Eastern Edition*, 242(87): 33-45.

Boggs, C. 2000. *Creating Dynamic E-commerce Website: Design factors*. <http://www.webdevelopersjournal.com/articles/Professional.html>. [21 July 2003].

Boyes, A. & Irani Z. 2004. An analysis of the barriers and problems to web infrastructure development experienced by small businesses. *International Journal of Information Technology and Management*, 3(4): 189-207.

Browne, G. & Jeremy, J. 2001. Website Indexing: Enhancing Access to information within Websites. *The Electronic Library*, 20(3): 66-77.

Bruemmer, P. 2002. Copywriting: Key to Search-Engines Visibility and Conversions. *Search Engines Guide*. <http://www.101searchengines.biz/>. [06 June 2003].

Byrne, E. 2004. Development through communicative action and information system design: a case study from South Africa. *South African Computer Journal*, 32: 25-33.

Chang, H. & Zhao, X. 2004. Effects of perceived interactivity on website preference and memory: Role of personal motivation. *Journal of Computer Mediated Communication*, 10(1).
<http://jcmc.indiana.edu/vol10/issue1/chung.html>. [28 April 2005].

Clarke, P. 2000. Internet as a medium for qualitative research. Paper read on conference on Web 2000 conference. Johannesburg, Rand Afrikaans University.

- Collin, S. 2002. Assessing and assuring the Quality of Website. *Online Information Review*, 10(3): 11-20.
- Craven, T. 2001. Description of meta-tags in locally linked webpages. *Aslib Proceeding*, 53(6): 203-216.
- Crowell, G. 2004. *Search Engine Watch: Retaining traffic after website redesign*.
[http: // Searchenginewatch.com/searchday/print.php/34711_342291](http://Searchenginewatch.com/searchday/print.php/34711_342291). [22 October 2004].
- Davidson, B.H. 2001. Database driven, dynamic content delivery: providing and managing access to online resources using Microsoft Access and Active Server Pages. *OCLC Systems & Services*, 17(1): 34-41.
- Dykehouse, S. & Sigler, T. 2000. Use of the World Wide Web, hyperlinks and managing the news by criminal justice agencies. *Policing: An International Journal of Police Strategies and Management*, 23(3) 267: 269.
- Eastman, C. & Jansen, B. 2003. Coverage, Relevance and Ranking: The impact of query operators on web search engine results. *ACM Transactions on Information Systems*, 21(4): 383-99.
- Elderbrock, D., Ezor, J., Dalton, L. and Weissberg, J. 1997. *Producing Web hits*. IDG Books Worldwide: USA.
- Fagan, C. 2001. Transaction to a unified system using Perl to drive library databases and enhance website functionality. *The Electronic Library*, 19(3): 168-179.
- Fontanella, J. 2000. The Web Based Supply Chain. *Management Review*, 3(4): 17-20.

Foot, A., Schneider, M., Dougherty, M. Xenos, M. & Carsen, E. 2003 Analyzing linking practices: Candidate website in the 2002 US electoral, Web sphere. *Journal of Computer Mediated Communication*, 8(4): 1-27. <http://jcmc.Indiana.edu/vol8/issue4/foot.html>. [22 April 2005].

Furnis, M. 2001. *The mission of Website: Building Search engines friendly Website*. <http://www.objs.com/survey/lang.htm>. [27 June2003].

Gardner, M. & Pinfield, S. 2001. Database-backed library website: a case study of the use of PHP and MySQL at the University of Nottingham. *Program: electronic library & information systems*, 35(1): 88-98.

Graham, S. & Inan, V. 1995. *Developing and maintaining e-commerce sites*. <http://www.objs.com/survey/lang.htm>. [27 June2003].

Greenough, R., Fakun, D. & Kay, J. 2001. Development of a digital manual for a manufacturing system – a case study. *Integrated Manufacturing Systems*; 12(6): 34-45.

Hof, R. & Sager, I. 2003. More Ways To Price-Shop Online. *Business Week*, Issue 3851: 14.

Holder, B. 2003. Forintek's value added pathfinder: A case study *The Electronic Library*, 21(1): 133, 143.

Hout, E. 1999. Step by step guideline for designing and documenting the navigation structure of multimedia hypertext system. *Information Management & Computer Security*, 7(2): 88-98. <http://docserver.emeraldinsight.com/deliver/cw/mcb/09685227/v7n2/s4/p88.htm> [23 August 2004].

Inan, H. 2002. Measuring the Success of Your Website: A Customer-centric Approach to Website Management. *Online information Review*, 26(6): 33-37.

Internet Visibility Service™ ('Real World Technology' <info@digitalphrase.com>). 27 June 2005. Internet Visibility Service™. E-mail to M. Ngindana (ngindanam@cput.ac.za).

Kaplan, M. & Carl, V. 1997. Software, Internet, network solutions diversity to serve users. *Control Engineering*, 46(5): 14-18.

Kerr, M. 2001. Tips and Tricks for Website Managers. *The Electronic Library*, 19(5): 15: 25.

Kleindl, B. A. 2001. *Strategic Electronic Marketing: Managing E-Business*. Cincinnati, OH: South-Western College Publishing.

Larson, R. & Farber, B. 2003. *Elementary statistics: Picturing the world*. New Jersey, Prentice Hall.

Li, W., Hsiung, W., Po, O., Candan, K. & Agrawal, D. 2002. Evaluation of architectural designs and implementation for database-driven websites. *Data and Knowledge Engineering*, 43(2): 151-60.

Li, X. 2003. A review of the development and application of the web impact factor. *Online Information Review*, 17(6).
<http://docserver.emeraldinsight.com/deliver/cw/mcb/14684527/v27n6/s3/p407.htm>. [23 September 2004].

Lide, R. & David, M. 2001. Provant launches new e-commerce Website, grows operating income 91.6% in 1Q. *Lifelong Learning Market Report*, 4(11): 26-33.

Lu, M. 1998. A framework for effective commercial Web application. *Internet Research: Electronic Networking Application and Policy*, 8(2): 166-173.

McGuigan, G. 2003. Invisible business information: the selection of invisible website in constructing subject pages for business. *Collecting Building*, 22(2): 68-74.

Mertz, D. 2001. Charming Python: Developing a full-Text indexer in Python: Paving the way to better searching. *Online Information Review*, 11(5): 21-25.

Mitra, A. 1999. Characteristics of the WWW Text: Tracing discursive strategies. *Journal of Computer Mediated Communication*, 15(1).
<http://jcmc.Indiana.edu/vol15/issue1/mitra.html>. [22 April 2005].

Mouton, J. 2001. *How to succeed in your Master's & Doctoral Studies: A South African guide and resource book*. Pretoria. Van Schaik.

Munafo, R. 2002. *Programming Languages: The Internet's current Opinion at MROB*. http://home.earthlink.net/~mrob/pub/lang_srom.html [27 June 2003].

Murata, T. 2003. Visualizing the Structure of Web Communities Based on Data Acquired From a Search Engine. *IEEE Transaction on Industrial Electronics*, 50(5): 38-50.

Nobles, R & Alexander, J. 2002. Internet Search Engines – Articles about Marketing and More. *Marketing on Internet Search Engines*.
<http://www.searchengineworkshops.com/articles.html>. [27 June 2003].

Ollins, D. 2002. *Professional Web Design – some good ideas for web design*. <http://www.webdevelopersjournal.com/articles/Professional.html> [21 July 2003].

Oppenheim, C., Morris, A. & Mcknight, C. 2000. Progress in documentation: the evaluation of WWW search engines. *Journal of Documentation*, 56(2): 190-211.

Paulussen, S. 2004. Online news production in Flanders: How Flemish online journalists perceive and explore the Internet's potentials. *Journal of Computer Mediated Communication*, 9(4).

<http://jcmc.Indiana.edu/vol19/issue4/paulussen.html>. [22 April 2005].

Pfattenberger, S. 2001. *How to search on search engines, A case study*.

<http://www.objs.com/survey/lang.htm>. [27 June 2003].

Position Research. 2002. *Search Engines Marketing Services including Link Building for Link Popularity Improvement and Better ranking*.

http://www.emarketing101.net/internet_marketing/link_popularity.html
[05 August 2003].

Rea-Smith, J.B., Alexander, E. & Ellinger, E. 2002. Insights from the introduction of an online logistics service system. *International Marketing Review*, 7(1): 1-21.

Rose, G., Khoo, H. & Straub, W. 1999. Current Technological Impediments to Business-to-Consumer Electronic Commerce. *Communication of the Association for Information Systems*, 1(16): 1-58.

Smith, A. 2001. Website Indexing: Enhancing access to information within websites. *The Electronic Library*, 20(3): 247-252.

<http://docserver.emeraldinsight.com/deliver/cw/mcb/02640473/v20n3/51005/p247.htm>

[23 August 2004].

Smith, C. 2002. Infrastructure needs for electronic commerce and the Internet, intranet and extranets. *South African Journal of Information Management*, 4(4): 1-17.

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

Struwig, F. W. & Stead, G. B. 2001. *Planning, designing, and reporting research*. Cape Town. Pearson Education.

Sullivan, D. 2001. How Search Engines Work: *How To Really Profit From SEO*. <http://www.searchenginesbook.com/presskit.html>
[27 June 2003].

Tansley, D. 2002. *Create Dynamic webpages using PHP and MySQL*. New York. Addison-Wesley.

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

Thelwall, M. 2000b. Commercial Websites: Lost in cyberspace. *Electronic Networking Applications and Policy*, 10(2): 150-159.

Thelwall, M., Binns, R. Harries, G. Page-Kennedy, T. Price, L. & Wilkinson, D. 2001. Custom interfaces for advanced queries in search engines. *Aslib proceedings: New information perspectives*, 53(10): 413-422.

Thelwall, M. 2002. Evidence for the existence of geographic trends in university Website interlinking. *Journal of Documentation*, 58(5): 55-65.

Thilmany, J. 2003. Web Search Measure Website Appeal In Seconds. *Mechanical Engineering*, 125(10): 10-14.

Thurrow, S. 2002. *Search Engine Visibility press kit*.
<http://www.searchenginesbook.com/presskit.html>. [27 June 2003].

Thurrow, S. 2003a. *Search Engine Visibility*. Indianapolis. New Riders.

Thurrow, S. 2003b. What can I do to optimize my website for search engines. <http://www.devwebpro.com/articles/0303b.html>. [27 June 2003].

Turner, T. & Brackbill, L. 1998. Rising to the top: Evaluating the use of the html meta-tag to improve retrieval of world wide web documents through Internet search engines. *Library Research and Technical services*, 42(4): 258 – 271.

Van Greunen, D. & Wesson, J. 2004. Exploring issues for information Architecture of Web-based learning in South Africa. *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)*, Port Elizabeth, South Africa, University of Port Elizabeth, 73-78.

Van Steenderen, M. 2001. Website management: making a website more visible. *South African Journal of Information Management*, 2(4): 1-12.

Watson, R., Pierre B., Leyland F. & George M. 2000. *Electronic Commerce: The Strategic Perspective*. Fort Worth. TX: The Dryden Press. 1-35.

Weideman, M. 2001. *Internet searching as a study aid for information technology and information systems learners at tertiary level*. [Unpublished Ph.D thesis]. Cape Town, University of Cape Town.

Weideman, M. 2005. FOIOTI: An implementation of the conceptualist approach to Internet information retrieval. *South African Journal of Libraries and Information Science*, 7(1): 2-25.

Weideman, M. & Ngindana, M. 2004. Website navigation architecture and their effect on website visibility: A literature survey. Fulfilling the promise of ICT, *Proceedings of SAICSIT 2004*, Stellenbosch, South Africa, 4 – 6 October 2004: 292-296. Stellenbosch University.

Welman, J.C. & Kruger, S.J. 1999. *Research Methodology for the*

Business and Administrative Science. Cape Town: Oxford University Press.

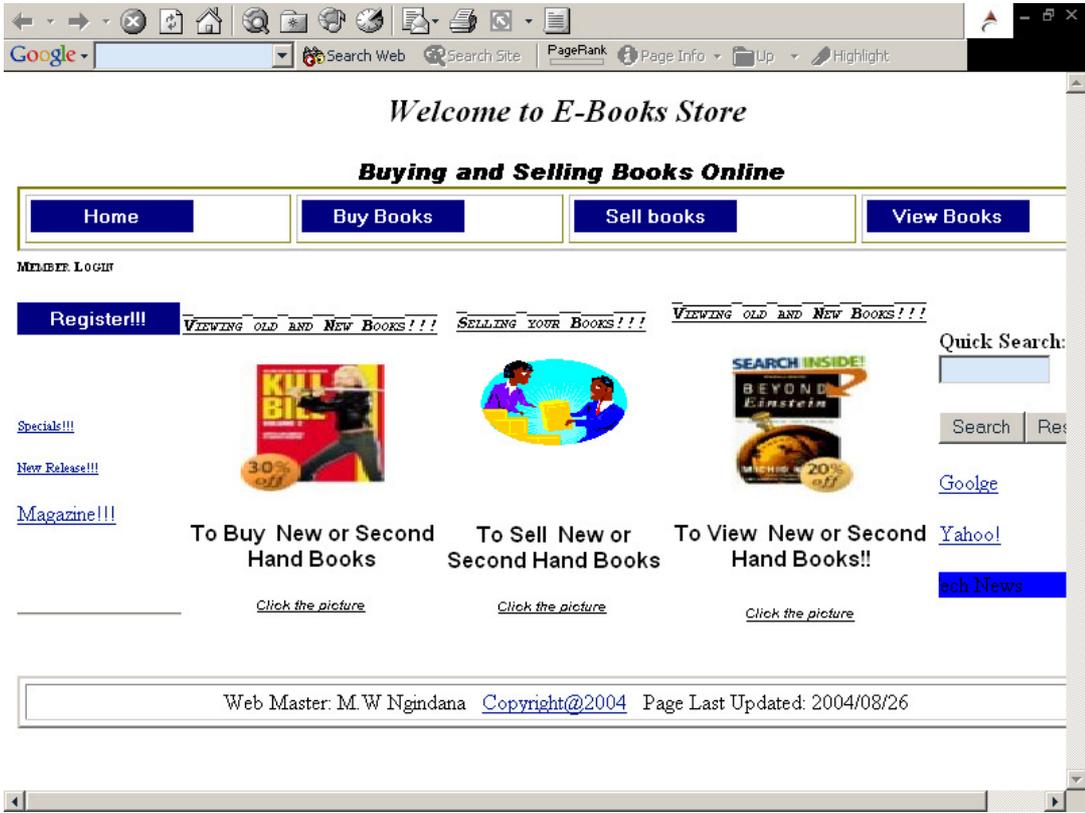
Whalen, J. 2003. *Search Engines Marketing vs Search Engines Optimization*. <http://www.searchengineguide.com/>
[18 June 2003].

Winn, W. & Beck, K. 2002. The persuasive Power of Design Element on an E-Commerce Websites. *Technical Communication*, 49(1): 17-19.

Xue, L., Harker, M. & Heubs, J. 2000. A review of the development and application of the Web impact factor. *Online Information Review*; 27(6): 25-38.

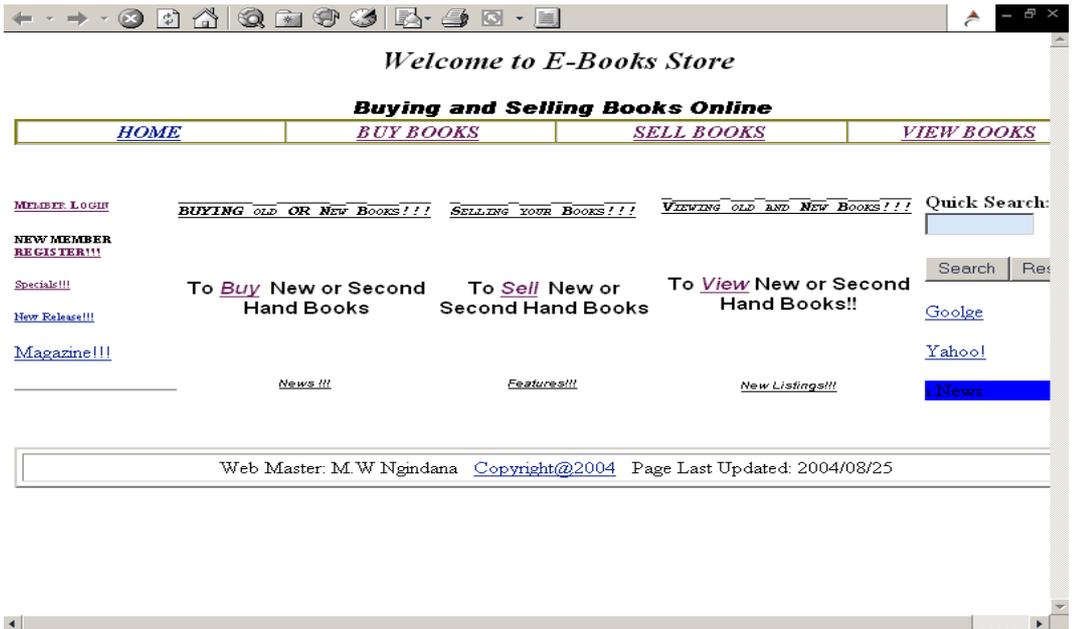
Yuen, L.T., Lee, Y.W., Lau, S.M. & Yen, J. 2001. From unstructured HTML to structured XML: How XML supports financial knowledge management on the Internet. *Library Hi Tech*, 19(3): 242-256.

APPENDIX A: Website layout: Graphic-based website



E-Books1 - a graphic-based website (www.ebooks1.co.za).

APPENDIX B: Website Layout: Text-based website



E-Books2 - a pure text-based website (www.ebooks2.co.za).


```

        <param name='url' valuetype='ref' value='Index.htm'>
    </applet>
</td>
<td width='25%' height='1'>
    <applet code='fphover.class' codebase='./' width='120' height='24'>
        <param name='color' value='#000080'>
        <param name='hovercolor' value='#0000FF'>
        <param name='textcolor' value='#FFFFFF'>
        <param name='effect' value='average'>
        <param name='bgcolor' value='#FFFF00'>
        <param name='text' value='Buy Books'>
        <param name='url' valuetype='ref' value='Info.htm'>
    </applet>
</td>
<td width='27%' height='1'>
    <applet code='fphover.class' codebase='./' width='120' height='24'>
        <param name='color' value='#000080'>
        <param name='hovercolor' value='#0000FF'>
        <param name='textcolor' value='#FFFFFF'>
        <param name='effect' value='average'>
        <param name='bgcolor' value='#FFFF00'>
        <param name='text' value='Sell books'>
        <param name='url' valuetype='ref' value='Info.htm'>
    </applet>
</td>
<td width='30%' height='1'>
    <applet code='fphover.class' codebase='./' width='120' height='24'>
        <param name='color' value='#000080'>
        <param name='hovercolor' value='#0000FF'>
        <param name='textcolor' value='#FFFFFF'>
        <param name='effect' value='average'>
        <param name='bgcolor' value='#FFFF00'>
        <param name='text' value='View Books'>
        <param name='url' valuetype='ref' value='Info.htm'>
    </applet>
</td>
</tr>
</table>
<table border='0' cellpadding='0' cellspacing='0' width='106%' height='290'>
<tr>

```

```
<td width='15%' height='1'><b><blink><span style='font-variant: small-caps'><font
size='1'><a href='Info.htm' style='border: 4 double #FF00FF'>Member
Login</a></font></span></blink></b>
```

```
<p>
<font size='2'>
<applet code='fphover.class' codebase='./' width='120' height='24'>
  <param name='color' value='#000080'>
  <param name='hovercolor' value='#0000FF'>
  <param name='textcolor' value='#FFFFFF'>
  <param name='effect' value='glow'>
  <param name='text' value='Register!!!'>
  <param name='url' valuetype='ref' value='info.htm'>
</applet>
</font>
```

```
<p>&nbsp;</p>
<p><a href='Info.htm'><font size='2'>Specials!!!</font></a></p>
<p><a href='Info.htm'><font size='2'>New Release!!!</font></a></p>
<p><a href='Info.htm'><font size='2'>Magazine!!</font></a></p>
```

```
<td width='23%' rowspan='2' height='47'>
  <p align='center'><b><i><span style='font-variant: small-caps; text-decoration:
underline'><font face='Arial' size='2'><sup>BUYING
  Old Or New Books!!!</sup></font></span></i></b></p>
  <p align='center'><a href='Info.htm'><img border='0'
src='B00005JMUA.01.30TLZZZZ.jpg' width='106' height='99'></a></p>
  <p align='center'><b><font face='Arial'>To Buy&nbsp;&nbsp;&nbsp; New or Second Hand
  Books!!!</font></b></td>
```

```
<td width='23%' rowspan='2' height='47'>
  <p align='center'><b><i><span style='font-variant: small-caps; text-decoration:
underline'><font face='Arial' size='2'><sup>SELLING
  Oold or new Books!!!</sup></font></span></i></b></p>
  <p align='center'><a href='Info.htm'><img border='0' src='Images/books.jpg'
width='78' height='94'></a></p>
  <p align='center'><b><font face='Arial'>To Sell&nbsp;&nbsp;&nbsp; New or Second Hand
  Books!!!</font></b></p>
```

```
</td>
<td width='23%' rowspan='2' height='47'>
  <p align='center'><b><font face='Arial' size='2'><i><span style='font-variant: small-
caps; text-decoration: underline'><sup>BUYING
```



```

        BOOKS</a></b></i></p>
    </td>
    <td width='30%' height='1'>
        <p align='center'><b><i><a href='Info2.htm'>VIEW
        BOOKS</a></i></b></p>
    </td>
</tr>
</table>
<table border='0' cellpadding='0' cellspacing='0' width='106%' height='290'>
<tr>
    <td width='15%' height='145'><b><blink><span style='font-variant: small-caps'><font
size='1'><a href='Info2.htm'>Member Login</a></font></span></blink></b>
    <p><b><span style='font-variant: small-caps'><font size='1'><blink>NEW
MEMBER </blink><a href='Info2.htm'>REGISTER!!!</a></font></span></b></p>
    <p><font size='1'><a href='Info2.htm'>Specials!!!</a></font></p>
    <p><font size='1'><a href='Info2.htm'>New Release!!!</a></font></p>
    <p><a href='Info2.htm'>Magazine!!</a></p>
    <p>&nbsp;</p>
    <hr>
</td>
    <td width='20%' rowspan='2' height='47'>
<p align='center'><b><font face='Arial' size='2'><i><span style='font-variant: small-caps;
text-decoration: overline'><sup>BUYING
    Old Or New Books!!!</sup></span></i></font></b></p>
    <p align='center'>&nbsp;</p>
    <p align='center'><b><font face='Arial'>To <a href='Info2.htm'><i>BUY</i></a>
    New or Second Hand Books</font></b></p>
    <p align='center'>&nbsp;</p>
    <p align='center'><b><font face='Arial' size='1'><i><u><a
href='Info2.htm'>News!!!</a></u></i></font></b></td>
    <td width='21%' rowspan='2' height='47'>
    <p align='center'><b><font face='Arial' size='2'><i><span style='font-variant: small-
caps; text-decoration: overline'><sup>SELLING
    Old Or New Books!!!</sup></span></i></font></b></p>
    <p align='center'>&nbsp;</p>
    <p align='center'><b><font face='Arial'>To <a href='Info2.htm'><i>
Sell</i></a>&nbsp;< New or Second Hand Books</font></b></p>
    <p align='center'>&nbsp;</p>
    <p align='center'><b><font face='Arial' size='1'><i><u><a
href='Info2.htm'>Features!!!</a></u></i></font></b></td>

```


**APPENDIX E: Visibility Report for Graphic-based website
from 'Real World Technology'**

From: 'Real World Technology' <info@digitalphrase.com>
To: Mongezi Ngindana
Date: Monday – June 27, 2005
Subject: Internet Visibility Services

Dear Mongezi,

Please find attached the analysis results for your site
<http://www.ebooks1.co.za>. Thank you for using our Internet Visibility
Services(tm)!

Kind Regards,
Ian Campbell

Head of The Internet Visibility Team

+27 (82) 513 6860

©2005 Real World Technology. All rights reserved.

INTERNET VISIBILITY SERVICES™

Introductory Analysis

INTRODUCTION

Hello! We would like to help you become more successful. We understand the frustrations and confusions of the online world. It often doesn't make any sense at all. And that's why we are here...

WHY THIS ANALYSIS?

Two-thirds of first-world populations use the Internet. This means your 'offline' business could be using the Internet to drive new business to it. Getting this to work is no accident, though. You have to know exactly what you are doing.

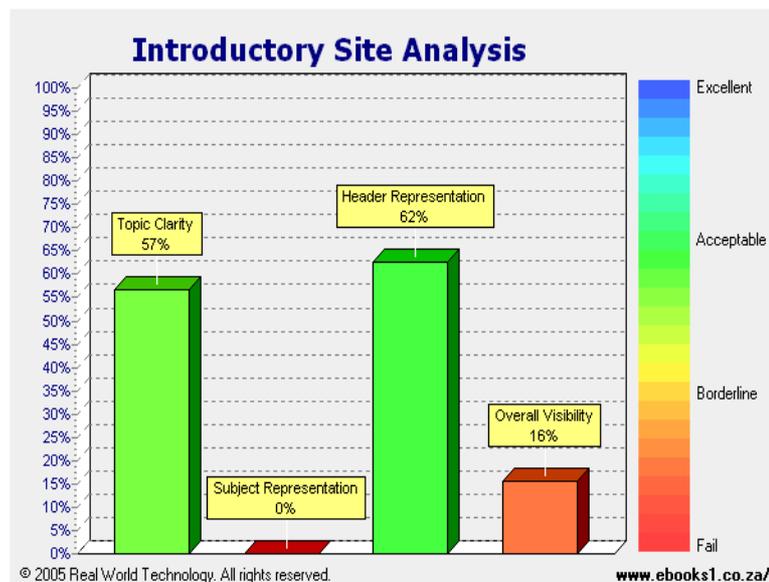
Considering the relatively low cost of producing a website, and considering that over 80% of people online are searching for services, we are sure you will agree there is a huge business potential to be tapped! That said, you are probably asking the million-dollar question:

Why isn't my website working like I expect it to?

YOUR ANALYSIS

Our Internet Visibility Services end the mystery of how to create a successful online business which either makes sales itself, or drives business to your offline business. This free analysis primarily shows you how visible you are to the search engines of the world. Why would you want to know this? Well, 85% of people use search engines to find what they want online. So if your ranking is say 12%, this means there is very little chance of potential clients finding you.

So let's see how your online business stands right now:



Graph E.1: Graphic-based website statistics (Source: Real World Technology, 2005)

Your site's content shows fair clarity, but could be improved. Improving this would be of advantage both from a legibility and search engine visibility point of view.

The representation of your site's topics to the search engines is severely inadequate. Even the best search engines will fail to properly recognise your site.

Your site's page headers are missing some components, but are good on the whole. This is a measure of structure, not content.

So far as visibility to the search engines is concerned, your site's visibility is very poor and must be improved.

The bottom line is your site is structurally in good shape. Our Internet Visibility Services contain all the tools and services you need to turn your website into a successful online business. They comprise the entire scope of scientific website creation, optimization, sales and marketing.

We are sure you'll agree things could be much better! We can help you with this. You won't need to learn technical jargon, or become a computer expert! We will help you step-by-step to turn your online business into one which works for you and helps your core business succeed better.

Thank you for using this introduction to our Internet Visibility Services. Simply reply now to this email and we will answer all your questions. We look forward to hearing from you!

Regards

The Internet Visibility Team

©2005 Real World Technology. All rights reserved.

**APPENDIX F: Visibility Report for Text-based website
from 'Real World Technology'**

From: 'Real World Technology' <info@digitalphrase.com>
To: Mongezi Ngindana
Date: Monday – June 27, 2005
Subject: Internet Visibility Services

Dear Mongezi,

Please find attached the analysis results for your site
<http://www.ebooks2.co.za>. Thank you for using our Internet Visibility
Services(tm)!

Kind Regards,
Ian Campbell

Head of The Internet Visibility Team
+27 (82) 513 6860

©2005 Real World Technology. All rights reserved.

INTERNET VISIBILITY SERVICES™

Introductory Analysis

INTRODUCTION

Hello! We would like to help you become more successful. We understand the frustrations and confusions of the online world. It often doesn't make any sense at all. And that's why we are here...

WHY THIS ANALYSIS?

Two-thirds of first-world populations use the Internet. This means your 'offline' business could be using the Internet to drive new business to it. Getting this to work is no accident, though. You have to know exactly what you are doing.

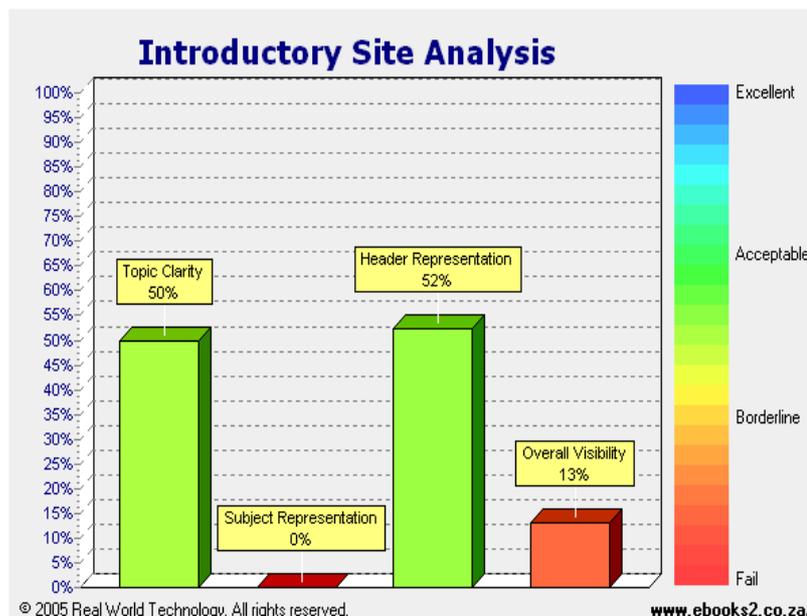
Considering the relatively low cost of producing a website, and considering that over 80% of people online are searching for services, we are sure you will agree there is a huge business potential to be tapped! That said, you are probably asking the million-dollar question:

Why isn't my website working like I expect it to?

YOUR ANALYSIS

Our Internet Visibility Services end the mystery of how to create a successful online business which either makes sales itself, or drives business to your offline business. This free analysis primarily shows you how visible you are to the search engines of the world. Why would you want to know this? Well, 85% of people use search engines to find what they want online. So if your ranking is say 12%, this means there is very little chance of potential clients finding you.

So let's see how your online business stands right now:



Graph F.1: Text-based website statistics

(Source: Real World Technology, 2005)

Your site's content shows fair clarity, but could be improved. Improving this would be of advantage both from a legibility and search engine visibility point of view.

The representation of your site's topics to the search engines is severely inadequate. Even the best search engines will fail to properly recognise your site.

Your site's page headers are missing some vital components, but are fair on the whole. This should ideally be corrected. This is a measure of structure, not content.

So far as visibility to the search engines is concerned, your site's visibility is very poor and must be improved.

The bottom line is your site is structurally in good shape. Our Internet Visibility Services contain all the tools and services you need to turn your website into a successful online business. They comprise the entire scope of scientific website creation, optimization, sales and marketing. We are sure you'll agree things could be much better! We can help you with this. You won't need to learn technical jargon, or become a computer expert! We will help you step-by-step to turn your online business into one which works for you and helps your core business succeed better.

Thank you for using this introduction to our Internet Visibility Services. Simply reply now to this email and we will answer all your questions. We look forward to hearing from you!

Regards

The Internet Visibility Team

©2005 Real World Technology. All rights reserved.

APPENDIX G: Statistics report for Graphic-based website from 'AWStats'

Monthly Report

The Monthly Report identifies activity for each month in the report time frame. Remember that each page hit can result in several server requests as the images for each page are loaded.

Note: Depending on the report time frame, the first and last months may not represent a complete month's worth of data, resulting in lower hits.

Table G.1: Website statistic report for graphic-based website (Source: AWStats, 2005)

Month	Percentage of requests	Number of requests	Percentage of page requests	Number of page requests	Percentage of bytes	Number of bytes transferred
1 October 2004	21.56%	124	21.65%	34	24.95%	179,121
2 November 2004	2.95%	17	4.45%	7	5.86%	42,075
3 December 2004	0.00%	0	0.00%	0	0.00%	0
4 January 2005	29.39%	169	22.29%	35	24.73%	177,563
5 February 2005	3.30%	19	6.36%	10	4.96%	35,615
6 March 2005	2.08%	12	3.18%	5	0.00%	0
7 April 2005	12.87%	74	7.00%	11	3.89%	27,934
8 May 2005	0.00%	0	0.00%	0	0.00%	0
9 June 2005	11.47%	66	17.19%	27	13.52%	97,051
10 July 2005	12.34%	71	10.82%	17	14.40%	103,410
11 August 2005	2.26%	13	4.45%	7	5.21%	37,421
12 September 2005	1.73%	10	2.54%	4	2.45%	17,628

Most active month January 2005 : **35 pages sent.** 177,563 bytes served. 169 requests handled.

Monthly average: **13.08 pages sent.** 59,818.16 bytes served. 47.91 requests handled.

This report was generated on October 4, 2005 12:16.
 Report time frame October 6, 2004 08:39 to September 27, 2005 14:30.

Web statistics report produced by:

Analog 4.13 Report Magic for Analog 2.04/linux

Quick Summary

The Quick Summary shows the peak entries for the entire website. This report gives the number of hits for each peak entry and the percent of the total hits for the website this represents for the reporting time frame.

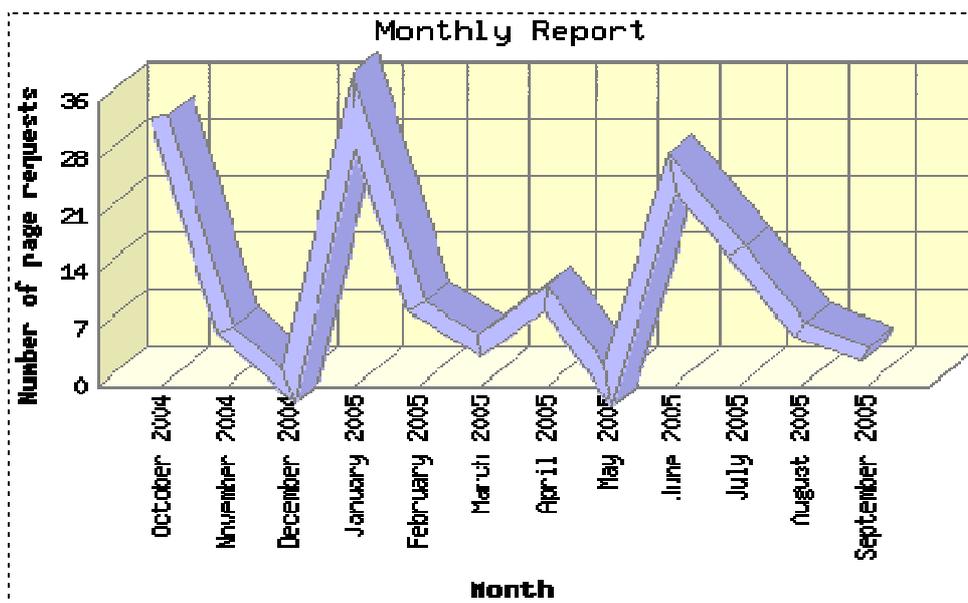
Table G.2: Quick summary of graphic-based website (Source: AWStats, 2005)

	Quick Summary	Peak Entry	Value	%Total
1	Most active day of the week	Friday	50	31.84%
2	Most active hour of the day	12:00 - 12:59	44	28.02%
3	Most active day	September 27, 2005	3	1.91%
4	Most active month	January 2005	35	22.29%
5	Most active week beginning	January 16, 2005	20	12.73%

This report was generated on October 4, 2005 12:16.
 Report time frame October 6, 2004 08:39 to September 27, 2005 14:30.

Web statistics report produced by:

Analog 4.13 Report Magic for Analog 2.04/linux



Graph G.1: Monthly Report of Graphic-based website (Source: AWStats, 2005)

APPENDIX H: Statistics report for text-based website from 'AWStats'

Monthly Report

The Monthly Report identifies activity for each month in the report time frame. Remember that each page hit can result in several server requests as the images for each page are loaded.

Note: Depending on the report time frame, the first and last months may not represent a complete month's worth of data, resulting in lower hits.

Table H.1: Website statistics for text-based website (**Source:** AWStats, 2005)

	Month	Percentage of requests	Number of requests	Percentage of page requests	Number of page requests	Percentage of bytes	Number of bytes transferred
1	October 2004	15.73%	73	13.26%	24	17.33%	61,772
2	November 2004	2.58%	12	4.97%	9	5.76%	20,551
3	December 2004	0.00%	0	0.00%	0	0.00%	0
4	January 2005	31.03%	144	23.75%	43	28.76%	102,511
5	February 2005	3.44%	16	7.18%	13	2.50%	8,924
6	March 2005	1.72%	8	2.76%	5	0.81%	2,888
7	April 2005	20.25%	94	19.33%	35	12.84%	45,764
8	May 2005	0.64%	3	1.65%	3	0.64%	2,298
9	June 2005	7.97%	37	13.81%	25	6.88%	24,520
10	July 2005	14.87%	69	10.49%	19	22.57%	80,423
11	August 2005	0.00%	0	0.00%	0	0.00%	0
12	September 2005	1.72%	8	2.76%	5	1.87%	6,680

Most active month January 2005 : **43 pages sent.** 102,511 bytes served. 144 requests handled.

Monthly average: **15.08 pages sent.** 29,694.25 bytes served. 38.66 requests handled.

This report was generated on October 4, 2005 11:50.

Report time frame October 6, 2004 08:40 to September 27, 2005 14:31.

Web statistics report produced by:

Analog 4.13

Report Magic for Analog 2.04/linux

Quick Summary

The Quick Summary shows the peak entries for the entire website. This report gives the number of hits for each peak entry and the percent of the total hits for the website this represents for the reporting time frame.

Table H.2: Quick summary of text-based website (**Source:** AWStats, 2005)

Quick Summary		Peak Entry	Value	%Total
1	Most active day of the week	Friday	61	33.70%
2	Most active hour of the day	12:00 - 12:59	54	29.83%
3	Most active day	September 27, 2005	4	2.20%
4	Most active month	January 2005	43	23.75%
5	Most active week beginning	April 3, 2005	32	17.67%

This report was generated on October 4, 2005 11:50.

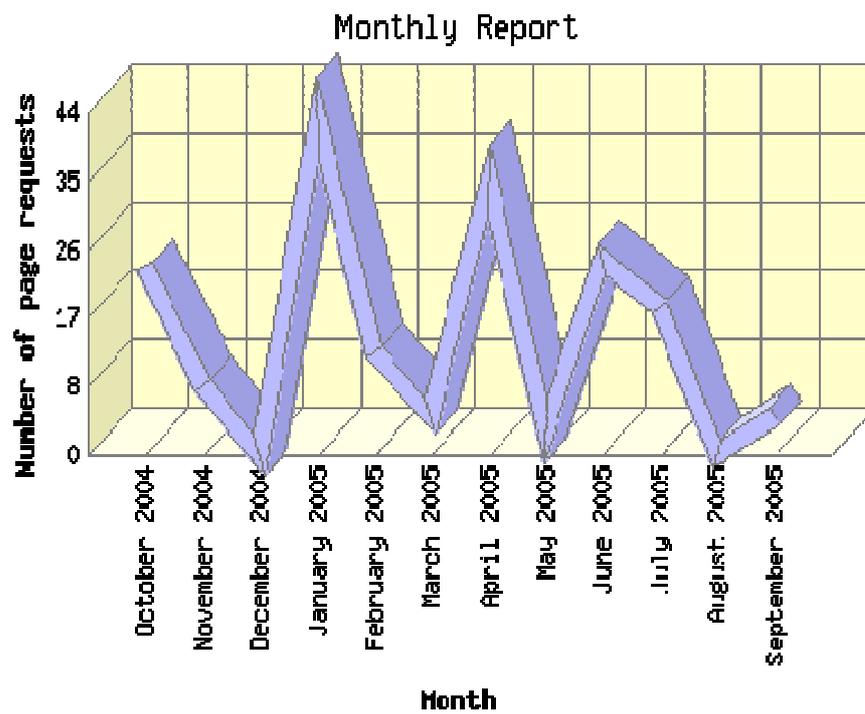
Report time frame October 6, 2004 08:40 to September 27, 2005 14:31.

Web statistics report produced by:

Analog 4.13

Report Magic for Analog 2.04/linux

Monthly Report



Graph H.1: Monthly report of text-based website (Source: AWStats, 2005)

APPENDIX I: Descriptive statistics for mean, mode and standard deviation.

Table I.1 Descriptive statistics, for mean median and std deviation

	WebType		Statistic	Std. Error		
Percentage of the requests	Graphics Based	Mean	0.0833	0.0272		
		95% Lower Confidence Bound Interval for Mean	0.0234			
		Upper Bound	0.1432			
		Median	0.0313			
		Variance	0.0089			
		Std. Deviation	0.0943			
		Minimum	0.0000			
		Maximum	0.2939			
		Range	0.2939			
		Interquartile Range	0.1092			
		Skewness	1.2570		0.6373	
		Kurtosis	0.8356		1.2322	
		Text Based	Mean		0.0833	0.0288
			95% Lower Confidence Bound Interval for Mean		0.0198	
	Upper Bound		0.1468			
	Median		0.0301			
	Variance		0.0100			
	Std. Deviation		0.0999			
	Minimum		0.0000			
	Maximum		0.3103			
	Range		0.3103			
	Interquartile Range		0.1461			
Skewness	1.2682	0.6373				
Kurtosis	0.8626	1.2322				
Number of requests	Graphics Based	Mean	47.9167	15.6479		
		95% Lower Confidence Bound Interval for Mean	13.4760			
		Upper Bound	82.3574			
		Median	18.0000			
		Variance	2,938.2652			
		Std. Deviation	54.2058			
		Minimum	0.0000			
		Maximum	169.0000			
		Range	169.0000			
		Interquartile Range	62.7500			
		Skewness	1.2565		0.6373	
		Kurtosis	0.8342		1.2322	
	Text	Mean	38.6667		13.3787	

	Based	95% Lower Confidence Bound	9.2203	
		Interval for Mean Upper Bound	68.1130	
		Median	14.0000	
		Variance	2,147.8788	
		Std. Deviation	46.3452	
		Minimum	0.0000	
		Maximum	144.0000	
		Range	144.0000	
		Interquartile Range	67.7500	
		Skewness	1.2682	0.6373
Kurtosis	0.8626	1.2322		
Percentage of the page requests	Graphics Based	Mean	0.0833	0.0229
		95% Lower Confidence Bound	0.0329	
		Interval for Mean Upper Bound	0.1337	
		Median	0.0541	
		Variance	0.0063	
		Std. Deviation	0.0793	
		Minimum	0.0000	
		Maximum	0.2229	
		Range	0.2229	
	Interquartile Range	0.1290		
	Skewness	0.9161	0.6373	
	Kurtosis	-0.5401	1.2322	
	Text Based	Mean	0.0833	0.0226
		95% Lower Confidence Bound	0.0335	
		Interval for Mean Upper Bound	0.1331	
		Median	0.0608	
		Variance	0.0061	
		Std. Deviation	0.0784	
Minimum		0.0000		
Maximum		0.2375		
Range		0.2375		
Interquartile Range	0.1175			
Skewness	0.7821	0.6373		
Kurtosis	-0.4280	1.2322		
Number of page requests	Graphics Based	Mean	13.0833	3.5939
		95% Lower Confidence Bound	5.1732	
		Interval for Mean Upper Bound	20.9934	
		Median	8.5000	
		Variance	154.9924	
		Std. Deviation	12.4496	
		Minimum	0.0000	
		Maximum	35.0000	

		Range	35.0000		
		Interquartile Range	20.2500		
		Skewness	0.9152	0.6373	
		Kurtosis	-0.5410	1.2322	
	Text Based	Mean	15.0833	4.0963	
		95% Lower Confidence Bound Interval for Mean	6.0674		
		Upper Bound	24.0992		
		Median	11.0000		
		Variance	201.3561		
		Std. Deviation	14.1900		
		Minimum	0.0000		
		Maximum	43.0000		
		Range	43.0000		
		Interquartile Range	21.2500		
		Skewness	0.7823	0.6373	
		Kurtosis	-0.4272	1.2322	
Percentage of the bytes		Graphics Based	Mean	0.0833	0.0261
	95% Lower Confidence Bound Interval for Mean		0.0259		
	Upper Bound		0.1407		
	Median		0.0509		
	Variance		0.0082		
	Std. Deviation		0.0904		
	Minimum		0.0000		
	Maximum		0.2495		
	Range		0.2495		
	Interquartile Range		0.1357		
	Skewness		1.0718	0.6373	
	Kurtosis		-0.0904	1.2322	
			Text Based	Mean	0.0833
		95% Lower Confidence Bound Interval for Mean		0.0209	
		Upper Bound		0.1457	
		5% Trimmed Mean		0.0766	
		Median		0.0413	
		Variance		0.0096	
		Std. Deviation		0.0982	
		Minimum		0.0000	
		Maximum		0.2876	
		Range		0.2876	
		Interquartile Range	0.1553		
	Skewness	1.1194	0.6373		
	Kurtosis	0.0629	1.2322		
Number of bytes transferred	Graphics Based	Mean	59,818.1667	18,731.9690	
		95% Lower Confidence Bound	18,589.3809		

	Interval for Mean	Upper Bound	101,046.9524
	Median		36,518.0000
	Variance		4,210,639,944
	Std. Deviation		64,889.4440
			0.0000
	Minimum		
	Maximum		179,121.0000
	Range		179,121.0000
	Interquartile Range		97,413.2500

	Skewness	1.0717	0.6373	
	Kurtosis	-0.0905	1.2322	
Text Based	Mean	29,694.2500	10,100.6156	
	95% Lower Confidence Bound Interval for Mean	7,462.9450		
	Upper Bound	51,925.5550		
	Median	14,737.5000		
	Variance	1,224,269,223		
	Std. Deviation	34,989.5588		
	Minimum	0.0000		
	Maximum	102,511.0000		
	Range	102,511.0000		
	Interquartile Range	55,324.5000		
	Skewness	1.1194		0.6373
		Kurtosis		0.063778

GLOSSARY

Absolute link

A hyperlink used to define a specific location of the webpage or document in terms of protocol, document name, directory, and server from which to obtain the page.

Breadcrumbs

Text link schematics that grow and shrink as the user navigates a website, informing him/her about their location on the website, as well as the pages visited leading to the current webpage.

Crawler-based search engine

Search engines which use automatic programs called 'crawlers', 'spiders' or 'bots' to visit webpages on the Internet, add their contents to the index, and allow users to query the index. Examples include Google and Inktomi.

Directories

A search engine whose main feature is to provide a hierarchical directory, created and maintained by human editors. Examples include Yahoo! and Looksmart.

Dynamic URL's

It is a universal resource locator that contains \$, %, =, &, in its name, which is generated dynamically during a request by a user. These characters are referred to as 'stop characters', because they force crawlers to stop interpretation of the URL at that point.

e-Commerce

This refers to a paperless exchange of routine business information using Electronic Data Interchange (EDI), e-mail, electronic bulletin boards and Electronic Fund Transfer (EFT).

Flash

It is a technology used to create visual flair, movies etc on websites. These technologies can add style to a website and help users to understand the navigation scheme better and enhance the user experience.

Frames

A navigation design technique used by website developers to divide the screen into sections. Each sections or frame represent a different webpage, and is typically used to provide easy-to-use navigation.

HTML

HTML is short for Hyper Text Markup Language, a basic Internet programming language that is understood by all World Wide Web clients, used for creating hypertext documents to be displayed by a browser.

Hyperlink

It is a word or set of words linked to a different location on the same webpage, a different webpage on the same website, or a webpage hosted elsewhere. If the user clicks on a hyperlink, he/she is transported to the destination location.

Image Map

It is single graphic image that represents links to different parts of a website. They provide users with a visual flair. Many search engines do not follow links inside image maps.

Index

A database created by search engine robots (software), used to retrieve documents or queries entered by users.

JavaScript

It is one of the most popular programming languages that enables website designers to add flair and interactivity to their websites. Some of the most

common uses for JavaScript includes: rollovers/mouseovers, pop-up windows and navigation menus.

JSP

JSP is the acronym for JavaServe Pages, a technology that offers a simple mechanism to create dynamic webpages that are both platform independent and server-independent.

Meta-tag

This is a webpage element used by the author of a website in the header section to describe or categorise the content of that webpage. This information is not seen when pages are viewed in browsers.

Navigation Button

This is a graphic picture or image, generally in a GIF or a JPEG format. It is a button used to navigate a website. It provides the user with a visual representation of how to navigate a website.

Pay Per Click (PPC)

This refers to a system which allows website owners to increase the traffic to their website. The charge is usually based on a bidding system. When a user clicks on the supplied link, the website owner pays a predetermined fee, regardless of the outcome of the visit.

Perl

Practical Extraction and Report Language (Perl), is a technology used to develop commercial and non-commercial websites, text processing and Internet services.

PHP

Hypertext Pre-Processor (PHP) is a server scripting language that works with HTML documents to produce dynamic content.

Relative Link

It is a link to a webpage, relative to the current webpage. It is normally shorter than an absolute link. Search engines can follow relative links to a webpage with ease.

Search Engine Optimisation (SEO)

It is a process of optimising a website so that it will be more visible to search engine crawlers. This process involves keyword research, website structure and architectural changes, building links to a website and creating meta tags.

Search Engines

Search engines are programs designed to assist users to find relevant information on the Internet. The term search engine is often used generically to describe both crawler-based search engines and human-powered directories. These two types of search engines gather their listings in radically different ways.

SMME

A Small, Medium and Micro Enterprise (SMME), is a small business that is defined by the number of employees (ranging from 5 to 150), and total annual turnover.

Spider

A program that travels across the Internet, visits one webpage after another, and indexes their contents.

Visibility

Visibility refers to the ease and effectiveness with which a search engine crawler can find and index a webpage. Webpage designers aim to design pages in such a way that they have high visibility to crawlers.

URL

An URL (Uniform Resource Locator) is an address referring to the location of a webpage or individual webpage element on the Internet.