11-1-2005

Search engine strategies : a model to improve website visibility for SMME websites

Rickard Chambers
Cape Peninsula University of Technology

SEARCH ENGINE STRATEGIES: A MODEL TO IMPROVE WEBSITE VISIBILITY FOR SMME WEBSITES

by

Rickard Chambers

THESIS

Submitted in 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

November 2005

DECLARATION

ii

I, the undersigned, hereby declare that the work done towards this qualification has been my own work, that it has not been submitted before for any degree or assessment at any other university or technikon. Furthermore, all the sources I have used or quoted have been indicated and acknowledged by means of complete references.

STUDENT:

Signature: _____   **Date: _____**
           R Chambers

ACKNOWLEDGEMENTS

"When you have come to the edge of all light that you know and are about to drop off into the darkness of the unknown, Faith is knowing one of two things will happen: There will be something solid to stand on or you will be taught to fly".

(Patrick Overton)

ABSTRACT

SEARCH ENGINE STRATEGIES: A MODEL TO IMPROVE WEBSITE VISIBILITY FOR SMME WEBSITES

The Internet has become the fastest growing technology the world has ever seen. It also has the ability to permanently change the face of business, including e-business. The Internet has become an important tool required to gain potential competitiveness in the global information environment. Companies could improve their levels of functionality and customer satisfaction by adopting e-commerce, which ultimately could improve their long-term profitability.

Those companies who do end up adopting the use of the Internet, often fail to gain the advantage of providing a visible website. Research has also shown that even though the web provides numerous opportunities, the majority of SMMEs (small, medium and micro enterprises) are often ill equipped to exploit the web's commercial potential. It was determined in this research project through the analysis of 300 websites, that only 6.3% of SMMEs in the Western Cape Province of South Africa appears within the top 30 results of six search engines, when searching for services/products.

This lack of ability to produce a visible website is believed to be due to the lack of education and training, financial support and availability of time prevalent in SMMEs. For this reason a model was developed to facilitate the improvement of SMME website visibility.

To develop the visibility model, this research project was conducted to identify potential elements which could provide a possible increase in website visibility. A criteria list of these elements was used to evaluate a sample of websites, to determine to what extent they made use of these potential elements.

An evaluation was then conducted with 144 different SMME websites by searching for nine individual keywords within four search engines (Google, MSN, Yahoo, Ananzi), and using the first four results of every keyword from every search engine for analysis. Elements gathered through academic literature were then listed according to the usage of these elements in the top-ranking websites when searching for predetermined keywords. Further qualitative research was conducted to triangulate the data gathered from the literature and the quantitative research.

The evaluative results provided the researcher with possible elements / designing techniques to formulate a model to develop a visible website that is not only supported by arrant research, but also through real current applications. The research concluded that, as time progresses and technology improves, new ways to improve website visibility will evolve. Furthermore, that there is no quick method for businesses to produce a visible website as there are many aspects that should be considered when developing "visible" websites.

RESEARCH OUTPUTS

The author produced the following research outputs during this study.

Output Type	Authors	Title	Institution / Event	Status
Journal Article	Weideman, M., Chambers, R.	A case study: website visibility of an e-commerce based cycling website.	South African Journal of Information Management (SAJIM)	Published in December 2005 on www.sajim.co.za
Book	Weideman, M., Kritzinger, W., Mbikiwa, F, Chambers, R. (Editors)	ICT Research Forum	Cape Town, South Africa. 01 May 2005, ISSN: 1814-9812.	Published
Conference Paper	Chambers, R., Weideman, M.	Search engine visibility: a pilot study towards the design of a model for e-commerce websites	7th Annual Conference on World Wide Web Applications	Published in Proceedings of WWW 2005 (http://www.zaw3.co.za)
Poster	Chambers, R., Weideman, M.	Website visibility: How visible do you want to be?	7th Annual Conference on World Wide Web Applications	Abstract published in Proceedings of WWW 2005 (http://www.zaw3.co.za)
Poster	Chambers, R., Weideman, M.	How visible are you?	South African Institute of Computer Scientists and Information Technologists (SAICSIT): Enablement Through Technology	Abstract published in Proceedings of SAICSIT 2005 (ISSN:1-59593-258-5 p.288)

SUPERVISOR:

Signature: _____   **Date: _____**

Prof. M. Weideman

TABLE OF CONTENTS

CHAPTER 3 – SEARCH ENGINE STRATEGIES

viii

CHAPTER 4 – RESEARCH METHODOLOGY

CHAPTER 5 – RESEARCH RESULTS AND ANALYSIS

CHAPTER 6 – CONCLUSION

REFERENCES

TABLES

FIGURES

GRAPHS

APPENDICES

CLARIFICATION OF TERMS

CHAPTER 1
BACKGROUND AND RESEARCH PROBLEM

1.1 INTRODUCTION

The aim of this thesis is to analyse search engine strategies and to identify possible elements required to develop an effective visibility strategy, which could contribute towards increased website traffic. In this thesis the author focuses on the extent to which SMMEs are or are not utilising the full potential of e-commerce offered through a highly visible website. The author describes this problem, identifies a potential model as solution and also explains how the model was formulated and verified. The objectives of this research study include the following:

- To investigate search services.
- To investigate to what extent SMMEs make use of e-commerce.
- To identify the most common index strategies used by websites occupying the top search engine positions.

SMME's are small to medium-sized companies – see full definition in the Clarification of Terms. According to Isaacs (2004), many small to medium-sized companies can dramatically increase their productivity with small improvements to their IT infrastructure. One of these improvements involves the adoption of the Internet. The use of the Internet by organizations opens up numerous business opportunities. These include overcoming geographical and cost barriers to new markets, improving service to customers, access to world-wide communication, streamlining internal processes, restructuring relationships and sharing scarce data (Simpande & Jakovljevic, 2003).

Due to the lack of knowledge time and funds, website owners will often simply launch their site, do nothing to ensure visibility and wait for hits which may never realize. According to Guenther (2004a: 47-48), this is an area on

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

Websites displayed on the first page of search engine results will enjoy a greater market share of search engine referred leads and will more often than not receive the widest exposure. According to Thelwall (2000: 150):

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

1.2 STATEMENT OF RESEARCH PROBLEM

The fact that SMMEs are not utilising the full potential of e-commerce results in the loss of advantages offered by a highly visible website. This ultimately culminates in a negative impact on bottom line profitability.

1.3 BACKGROUND TO RESEARCH PROBLEM

While large enterprises are restructuring and downsizing, SMMEs are playing an increasingly important role in South Africa's economy (Bester, 2003). According to the South African Department of Trade and Industry (2001), the small business sector made a valuable contribution to the economic development of South Africa in 2001. The research by Baard (2004) concluded that the small business sector contributed 53.9% to private sector employment and 34.8% to the total domestic gross product during the same period. Furthermore, that small businesses form 98% of the total business population in South Africa. However, between 70% and 80% of these small businesses face failure within the first three years of existence (Barron, 2000).

There are several reasons that contribute to these failures. One reason is that the majority of SMMEs are often ill equipped to exploit commercial use of the web due to limited knowledge, skills and resources at their disposal (Boyes & Irani, 2004: 191).

Information searchers would not only expect the item or content-rich page that they are searching to be in the top ten hits, but also that it is shown on the first screen. Similarly, businesses would also prefer that their webpages manifest in this format, most likely capturing the interest of the searcher (Rowley, 2001: 203-204). Every web developer should ideally invest in a search engine strategy based on the effective use of techniques and programming tools. For example, meta-tags or alt-tags could be used to increase the chances of the website being listed in high positions in search queries. In this respect, according to Constantinides (2002: 2633):

"A sound search engine strategy provides mechanisms for allowing the frequent evaluation of the site visibility."

In the majority of cases, except where there is no need for a webpage to be visible (e.g. page where content changes too rapidly), the developer's responsibility should also include sustained improvement of the visibility of the website. By achieving this, the developer could ensure sustained website visibility which, if supported by conventional advertising methods, should generate regular traffic to the website.

1.3.1   Type of Search Services

There are many methods for locating a website on the Internet. The most popular method is the use of a search service (Oppenheim, Morris, & McKnight, 2000: 192; Green, 2000: 125). Search services can generally be categorised into two types of sources, namely "directories" and "search engines". These services have their own search strategies for categorising websites, which determines where a site would be listed. To improve the understanding of the features required to design a website with higher

visibility, a clear understanding must be attained about how these two search services operate.

1.3.1.1 Directories

Nobles and O'Neil (2000: 6) defines a web directory as a service where a whole site, as opposed to a single page, is submitted by a developer and then evaluated by human editors. These human editors visit the submitted site to determine the ranking, and then categorise it according to subject or topic in a database.

1.3.1.2 Search Engines

A search engine is a search service, which makes use of crawlers to examine and index websites into a database of website listings according to their relevancy (Green, 2000: 125).

1.3.2 Factors Affecting Visibility

Due to search engines and directories being the most used search services on the Internet (Haltley, 2002 attest by Zhang & Dimitroff, 2004: 665), with each using its own indexing strategy, great emphasis should be placed on the importance of improving visibility techniques. By achieving visibility on the development of a website, both these services can be satisfied.

A number of authors (Rowley, 2001: 205-207; Srinivasan, Ruiz, & Lam, 1996:79-81; Thurow, 2003; Nobles & O'Neil, 2000; Weideman, 2004: 908) in the field of website visibility, collectively claim that:

- Visibility is a sustained process.
- Content must be accurate and compelling.
- Frames should be avoided where possible.
- Meta-tags should be included, even though they are not often used.

- Macromedia Flash should be avoided as it increases download time of content.
- Graphics should be minimised to improve download speed.
- An added site map provides additional hyperlinks.
- Care should be taken using JavaScript, as it is unreadable by crawlers.
- A website should be listed with the most popular search engines.

These are only some of the possible factors required to attain possible website visibility, and as time progresses and technology increases, new and improved ways to improve website visibility will evolve.

1.4   METHODOLOGY

A total of 300 websites were analysed to determine their ranking within six search engines, being AOL, Google, Lycos, MSN, Netscape and Yahoo. Only search engines were used. Evaluation was done on a further 144 websites by searching for 12 individual keywords within four search engines (Google, MSN, Yahoo, Ananzi), and taking the first three results of every keyword from every search engine for analysis. Elements gathered through academic literature (meta-tags, webpage titles, domain names, etc) were then listed according to the usage of these elements in the top-ranking websites, when searching for predetermined keywords.

The following serves as research and investigative questions for this research study:

Research question:

- *Can an effective website visibility strategy be developed to ensure that SMMEs draw website traffic, which could potentially lead to increased sales?*

Investigative research questions:

- To what extent do SMMEs require assistance with website visibility?
- How has the development of the Internet impacted upon search engine strategies?
- Which strategies are implemented by websites which are listed high in search engine results?

1.5 RESULTS AND CONCLUSION

The results from the research in this thesis returned that SMMEs are in need of improved visibility, as 93.7% of the 300 SMMEs analysed did not appear within the top 30 results of six leading search engines. Using quantitative methods, the research further identified 12 elements which could be used to improve the position of a website within the top search engines. Qualitative methods were then used to triangulate the results gathered from academic literature and quantitative approaches. These results pointed to ten elements to be considered when designing or maintaining a webpage. It should be noted that even though these elements are rated according to importance, their importance could vary as time progresses, technology increases and improved ways to improve website visibility evolve.

1.6 LIMITATIONS

The following limitations impacted upon the research project:
- The opinions on index strategies of the South African search engine representative may not correspond to opinions of other search engines in other countries.
- Only Western Cape SMMEs were used to conduct this research.
- As technology progresses, new indexing techniques develop in parallel. Some of the elements identified within this research could

be outdated or could even be considered as spam at any given point in time.

1.7   CLOSURE

There has been a strong motivation for businesses to produce websites and to be part of the Internet. These websites carry very little value without visitor traffic, which is achieved through several strategies. There are many aspects that should be considered when designing a visible website, which are both a time consuming and a sustained process. From the results gleaned from this research, it was possible to formulate an effective model consisting of possible elements necessary to achieve website visibility. This model provides possible methods for increasing website visibility to small business organizations, which lack the necessary funds, knowledge and skills to outsource their website visibility needs.

Furthermore, the model has the potential to provide SMMEs and other organizations with the essential techniques to achieve higher website visibility, which in turn should attract valuable customers to such organizations. It is also anticipated that, should this model be implemented, the following outcomes would culminate from it:

- Higher ranking in search engine results due to higher website visibility.
- Increased crawler friendliness without sacrificing user friendliness.
- Increased content rich webpages benefiting the search engine and the user.
- Webpages which load faster, thus providing a huge benefit to users operating on slower connections.
- Increased usability of websites to disabled users by implementing text that describes the graphics.
- Improved navigation offered to the user via a site map.

CHAPTER 2

BACKGROUND TO SMME'S AND SEARCH ENGINES

2.1   SMMES

2.1.1   Background to SMMEs

As large enterprises are restructuring and downsizing, small, medium and micro enterprises (SMMEs) have made a valuable contribution to economic development in South Africa (Boter, 2005: 244; South Africa Business Guidebook, 2003).

Research indicates that SMMEs in South Africa contributed 53.9% to formal private sector employment and 34.8% to the total domestic gross product in 2001. These businesses also form 98% of the total business population in South Africa (South African Department of Trade and Industry, 2001) and 80% internationally (Boyes & Irani, 2004: 192).

Although new SMMEs are continually being established, the majority of these enterprises seldom survive. Research shows that in South Africa, 70% to 80% of SMMEs face failure in their first three years of existence, costing the economy millions (Barron, 2000: 1; Streek, 2001 attest in Baard, 2004). These high failure rates are forcing South Africa to turn to economic strategies for the promotion of SMMEs (Rogerson, 2001: 271).

Reasons which can be attributed to the high failures have been identified in several research projects. Research by Baard (2004), and Fillis, Johannson, and Wagner (2004: 180) indicated that failures of SMMEs are partially due to internal factors, as a single manager operates the majority of small firms. A manager as sole decision maker sometimes lacks the required skills to manage critical tasks, such as adequate planning and financial control. Furthermore, failures in SMMEs are not always to be blamed on poor management.

The SMME sector also exists in a hostile environment having several constraints, one being that the operation within the global business environment has intensified as a result of competition (Baard, 2004).

2.1.2   Adoption of the Internet

In many instances, attaining a strategic advantage over competitors requires innovative use of information technology (IT) (Singh, 2002). According to Durkin and McGowan (2001: 15), SMMEs often have the potential for attaining a competitive advantage faster than a bigger firm, as they are much quicker and more flexible in connecting to the Internet when adopting the use of e-commerce effectively. If they are slow in the adoption of this technology, they could be confronted with a number of problems (Sparkes & Thomas, 2001: 336).

The growth of the Internet has produced an important information resource during the last decade, advancing at a much faster rate than was previously envisaged. As shown in Graph 2.1, the growth of the Internet took seven years to reach a 25% international market share. That is 70% faster than the development of the radio and 80% faster than the development of the telephone. This growth makes the Internet the fastest growing technology the world has ever encountered (Singh, 2002). Boyes and Irani (2004: 191) support this research by stating that the Internet had reached 50 million global users in five years as opposed to the 38 years of radio and 13 years of television.
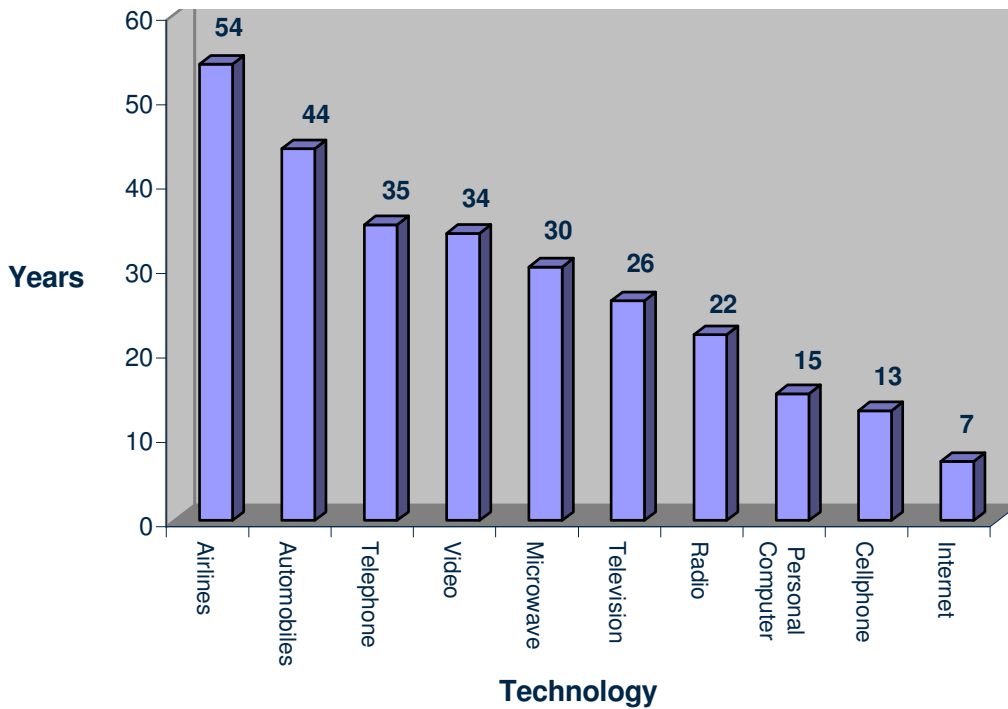
GRAPH 2.1: Time taken for technologies to reach a 25% market share (Singh, 2002)

Further research indicates that in 1989 the Internet had approximately 80,000 hosts' names registered. By 2005, this number had risen to 317,646,084 (see Graph 2.2) (Internet Systems Consortium, 2005).
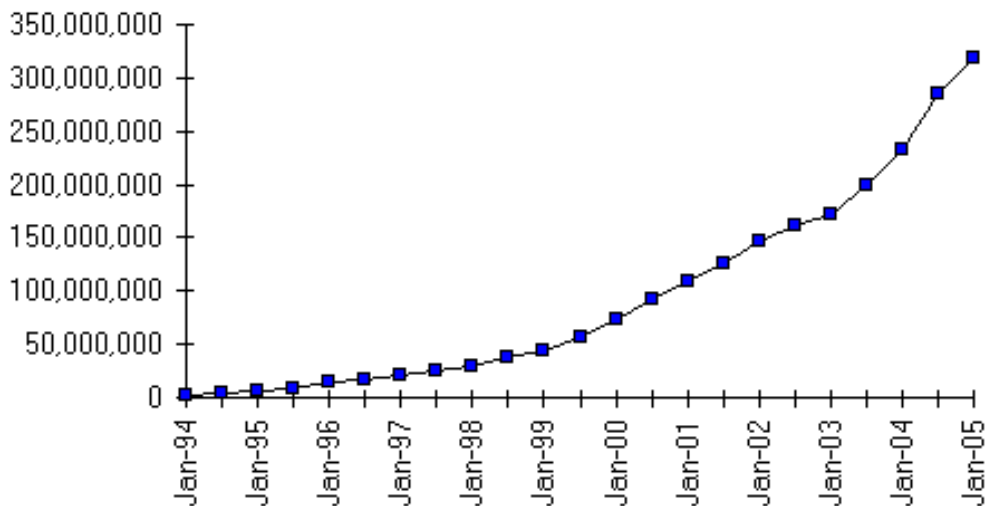
GRAPH 2.2: Internet domain survey host count (Internet Systems Consortium, 2005)

These statistics prove that the Internet as a global information economy has become an information source with no boundaries. Making effective use of the Internet can bring numerous business opportunities, enhancing global

competitiveness and attracting foreign investment for small, medium and micro enterprises (Simpande & Jakovljevic, 2003; Fillis *et al.*, 2004: 179; Sparkes & Thomas, 2001: 335).

Sparkes and Thomas (2001: 334) state that several SMMEs have come to realise the importance of what the Internet can provide in terms of improving customer relations. This is most often the basis for the majority of SMME marketing strategies. Large amounts of money are poured into these marketing strategies every year to attain new potential clients, which in the end could increase the company's customer base and ultimately its profitability. The same authors further state that a well developed website should have content-rich information with graphics, which draws the attention of the customer. This website could have a positive impact on the organization's strategies, which in turn could culminate in a competitive advantage (Sparkes & Thomas, 2001: 334).

Research conducted by Fillis *et al.* (2004: 181), Simpande and Jakovljevic (2003) and Boter (2005: 244-258) reported that by adopting e-business effectively into an SMME, the business could:

- Experience improved relationships with customers and suppliers.
- Attain control over distribution and marketing of products.
- Attain the ability, as a smaller firm to communicate globally to the same degree as its larger counterpart.
- Bypass traditional business barriers such as the physical distance between markets.
- Become more innovative.
- Experience a faster response to environmental demands.
- Become able to adapt or change faster to attain a competitive advantage.

Boyes and Irani (2004: 191) and Cool (2003) support the research carried out by Fillis *et al.* (2004) by stating that, when e-commerce is used effectively, it

has the potential of radically improving communication methods (customer and business related) at relatively low cost as opposed to conventional methods.

Singh (2002) summarised the benefits experienced by America's top 100 organizations by using the Internet. Results of this summary are depicted in Graph 2.3.
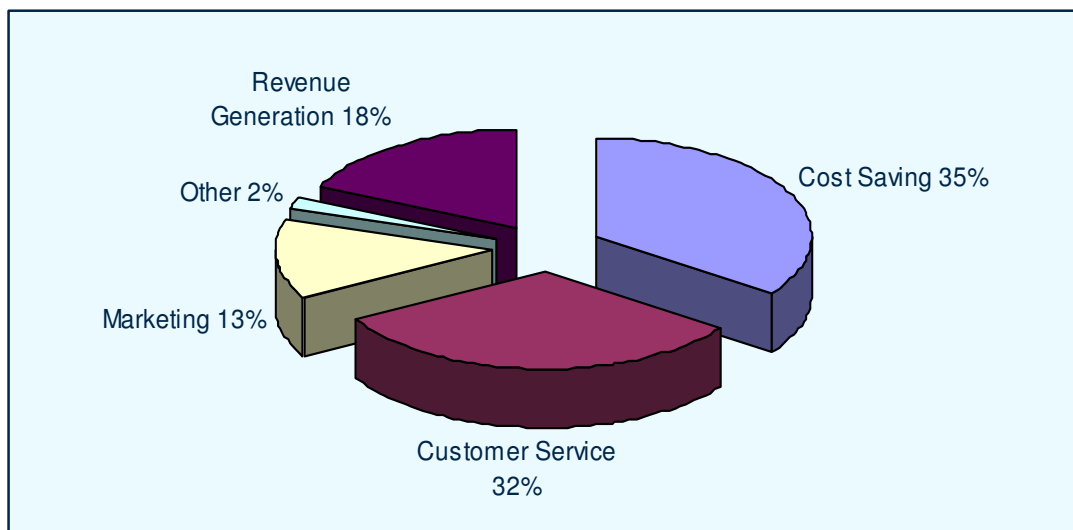
GRAPH 2.3: Benefits experienced by top 100 American organisations through using the Internet (Singh, 2002)

There are several reasons why businesses decide against the adoption of a new technology. One is the limited amount of time available to the SMME manager – most of his time is spent on daily operations and routines. The allocation of time to the exploration of new technologies is either very limited or is simply non-existent (Thelwall, 2000: 151). A further reason is that small business managers, as sole decision makers, often create ideas based upon entrepreneurial skills attained from practical experience. They therefore often lack the necessary managerial skills to transform their ideas into plans, strategies and ultimately into operational practice. The process of transforming ideas into plans and then executing such plans is believed to be one of the most difficult activities of a manager of an SMME, due to the high degree of uncertainty and complexity involved with the planning process (Baard, 2004; Boter, 2005: 244; Leacock, 2005: 355-366).

Fillis *et al.* (2004) list the following perceptions SMMEs have against the adoption of e-commerce. SMMEs have the perception that:

- Technical expertise is required to adopt e-business.
- An increased workload involved in establishing e-business processes is to be expected.
- A time constraint exists to establish e-business processes.
- A high number of resources are required to establish the e-business processes.

The reasons stated above are supported by research carried out by Bennett (1997: 324-344), where questionnaires had been distributed to exporting firms without websites to determine their reasons for not having a website. The author concluded that in general, companies felt less comfortable with the prospect of change and would rather stick to current practices with which they are familiar.

Managers of SMMEs furthermore often believe that Internet use is technically complicated and expensive, and that personal contact with customers through telephone conversations or visits by representatives is more important to secure sales. This type of thinking is often due to the fear of change. In this respect, according to Xavier (2005: 35):

"Change is uncertain… uncertainty breeds fear … fear prevents change from occurring."

Companies need to demonstrate flexibility, adaptability and agility in order to stay ahead of the game and to maintain a competitive advantage. In this respect, Xavier (2005: 35) is of the opinion that:

"Leaders who fail to plan for change will be left holding the ball long after the game is over – and they've lost."

Certain companies have taken the first step in the acquisition of new technologies by turning to ways of assigning individual employees to carry out the various managerial tasks. This approach allows the manager of the SMME to improve his/her managerial skills to manage the business more successfully. This approach has furthermore proved to be successful, but many SMMEs do not have the resources available to implement it (Baard, 2004). Other SMMEs have turned to outsourcing, as it often seems to be the only option for adopting e-business. One problem with outsourcing is that of finding reliable, cost-effective independent advice that will focus on a specific requirement (Baard, 2004).

SMMEs who adopt e-commerce in their business, often fail to develop the e-commerce technology as effectively as their larger counterparts and as a result, fail to exploit the benefits to their fullest extent. In a survey conducted by Smyth and Ibbotson (2001) in Fillis *et al*. (2004: 182), the results returned that Internet connectivity acquisition rates in SMMEs were higher than in larger corporate firms. Even though SMMEs acquisition rates were higher, the results also proved that their e-business activity rate was much lower. Reasons for this low implementation and growth rate were identified as the lack of the required skills, poor strategies, lack of trained staff and poor knowledge of the Internet start-up process.

Research carried out by Boyes and Irani (2004: 189), which is also supported by Fillis *et al*. (2004: 182) shows that, although the advantages offered to a small business through the use of the Internet are well researched and documented, small businesses are often the most ill equipped organizations to benefit from it. Boyes and Irani (2004: 193) emphasise the fact that:

"Unless further research is undertaken to analyse and document the problems being faced by Small Businesses when they connect to the web, and present this in a formal easily accessible by both the

academic community and SME community, many Small Businesses' future efforts to develop effective websites will prove to be unsuccessful."

Fillis *et al.* (2004: 184) formulated a conceptual framework, shown in Figure 2.1, to explain the acceptance and non-acceptance of e-business in SMMEs. This framework also highlights the potential dangers of non-adoption for those SMMEs with an existing negative outlook towards the adoption of e-business. It should also be noted that the framework depicted in Figure 2.1 does not necessarily contain all the factors impacting upon the decision to adopt e-business. The framework should be viewed as a guideline towards the understanding of the reasons behind e-business adoption and the impact of non-adoption (Fillis *et al.*, 2004: 184).
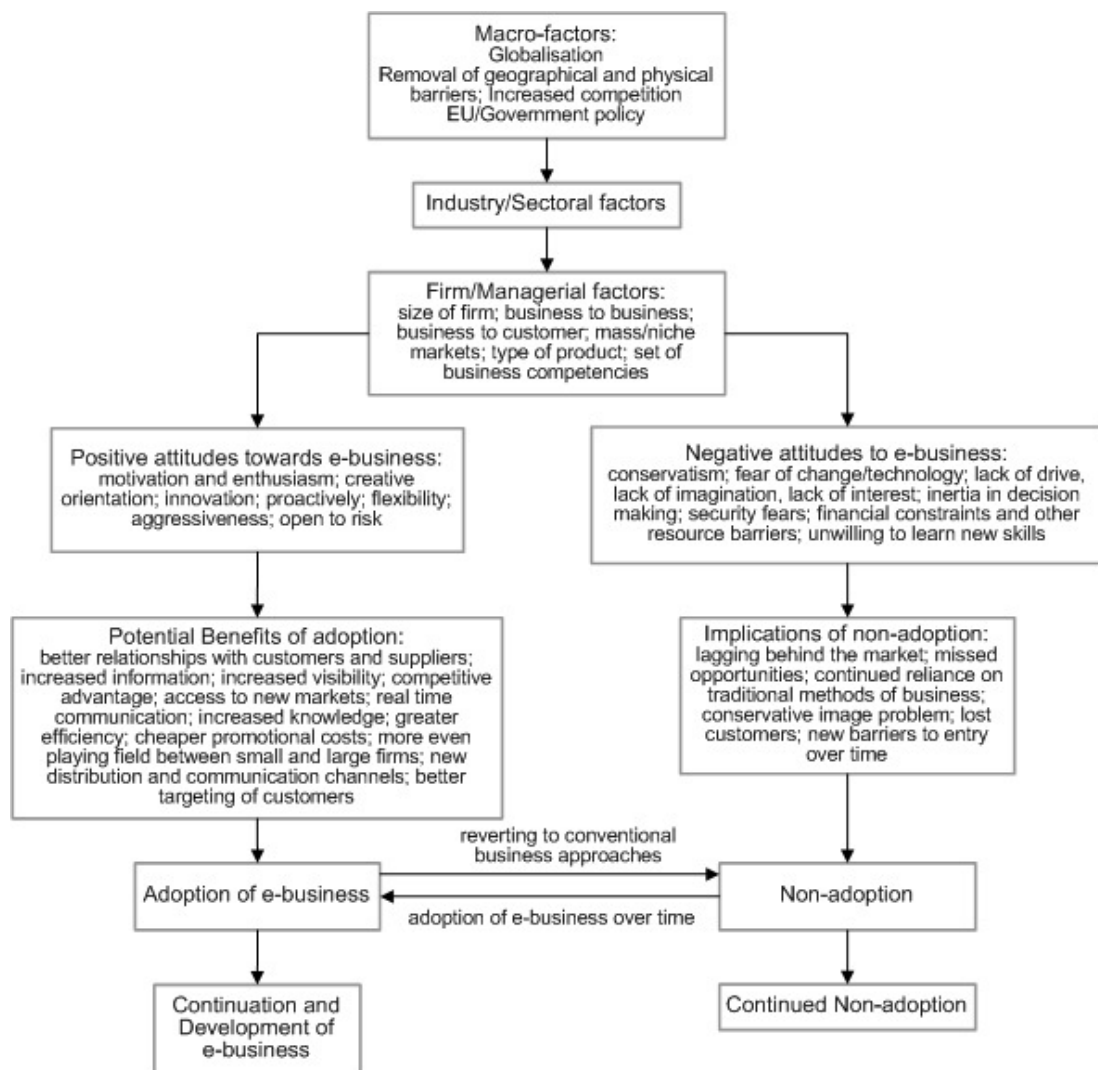
FIGURE 2.1: Conceptual model of e-business development (Fillis *et al.*, 2004: 184)

Key competencies required for the successful adoption of e-business include:

- Knowledge of the medium used to do business.
- The vision to predict the medium's usefulness in future business strategies.
- The ability to translate this vision into actual proactive business practice.
- Some technological awareness of how Internet technology operates to facilitate future business growth using existing and new technologies (Fillis *et al.*, 2004: 181).

Managers of SMMEs who have not successfully adopted the use of e-business, need to realise that by simply having a website, does not mean that anybody will visit the site (Thelwall, 2000: 149). A potential buyer starts a search for information in the hope of finding a solution to current needs. Those websites displayed on the first page of the returned results will enjoy a greater market share of search engine referred leads. Thelwall (2000: 150) states that it is critically important to carry out simple, non-technical tests on any newly designed website. For example, checking where the site ranks in the major search engines, by searching for relevant key words.

According to Guenther (2004a: 48), writing well-structured content and making use of several visibility techniques to improve a website's ranking, does not appeal to developers as it should. Furthermore, the improvement of website visibility is also an area where project teams tend to spend the least amount of time and thought.

A potential customer can visit a webpage in one of in three ways:

- By typing in its address.
- By following a link from another site.
- By following a link from a search engine results page.

Users usually know exactly what they want, but in the majority of cases, they do not know the address of the site containing the information. As a result, the users would turn to an easier method of searching for the required information on the web by utilising a search service.

It is therefore of importance to ensure that a site is listed high in the index of any search service. Research conducted by several authors confirms this fact. Jansen (2000) found that 58% of users view ten or fewer results per search query. Research done by Zhang and Dimitroff (2004: 665) shows that only 1% of users scan for pages containing the information they are looking for beyond the third page. Silverstein, Henzinger, Marais and Moricz (1999: 6) studied 993,208,159 AltaVista queries and found that 85% of users viewed ten or fewer results.

2.2 BACKGROUND TO SEARCH ENGINES

2.2.1 Indexing Techniques

Computers have become powerful tools that can perform numerical calculations in a billionth of a second and store vast amounts of information at the press of a button. These powerful computers have resulted in a global information economy having no boundaries. Technical constraints in the likes of including every single term of a given textual document in an index with the assistance of methods such as inverted indices, string and positional searches were unheard of two decades ago (Weideman & Kritzinger, 2003: 232).

Indexing is essential to information retrieval. It allows a searcher via an interface to search for certain keywords in a collection of documents. Keywords or phrases could then be used to retrieve relevant documents, without the user having to examine the whole document for relevant keywords (Wilson, 2002).

Even though several retrieval methods exist to index documents, effective use of a formal framework as designed by Blair, (1990: 27) can be developed to:

- Better understand some fundamental processes used by information retrieval systems.
- Make beneficial comparisons between the different methods used.
- Identify some of the major strengths and weaknesses in information retrieval design.

Blair, (1990: 27-70) describes 12 major retrieval models, which are analysed below:

2.2.1.1   Model 1

Model 1 is the simplest and also the most common of the twelve models, as it represents the basic retrieval characteristics of a typical library. Documents are indexed by making use of single descriptors e.g. Author Name, Book Title, etc. A user makes a request, which consists of a single descriptor e.g. Book Title. A document is then retrieved when the word entered by a user, matches one of the descriptors allocated to a document. The retrieved results are not ranked, and can be analysed by the user for relevancy. The problem with this method of retrieval is that by using a single descriptor to retrieve a document, the systems could become less useful the larger the database becomes (Blair, 1990: 28-29).

2.2.1.2   Model 2

Model 2 is very similar to model 1, however it represents an improvement on Model 1. In this model, a user can make a request, which consists of more than one descriptor, e.g. Author Name and Book Title. A document is retrieved if all the descriptors in the request match those in the index record of the document. By using more than one descriptor, a smaller number of documents will be retrieved and therefore it will increase the efficiency of the retrieval process. A problem with this model is that it is difficult to implement

on a manual system, and also does not provide room for mistakes (Blair, 1990: 29-31).

2.2.1.3 Model 3

Whereas models 1 and 2 retrieve only documents where the search query matches that of the descriptors assigned to a document, model 3 permits a document to be retrieved if only some of the descriptors match. In model 3, the user has the ability to specify how many descriptors should match the descriptors assigned to the document. For example, when the user does a search for the book title "Internet Basics" by author "Steve Lambert" with publisher "Random House Electronic Publications", the user can specify that only the title and author should match the descriptors assigned to the document. This functionality has several disadvantages. If a user should run a query containing a large number of descriptors and request a cut off value of only 2, a large number of request terms is generated which could cause processing problems. Retrieved documents are not ranked, and due to the different combinations of terms used to conduct the search, duplication of results could also occur (Blair, 1990: 32-34).

2.2.1.4 Model 4

Model 4 reduces the problem of the occurrence of duplication by ranking the retrieved documents according to the best match of the search query. One problem with this method is that it increases processing time. A further problem with this model is that it does not categorise terms according to their importance (Blair, 1990: 36-37).

2.2.1.5 Model 5

Model 5 represents an improvement on model 4, by allowing the user to indicate (with the use of numbers) the importance of some of the selection terms. The documents are then retrieved and sorted according to the importance levels assigned to them. One weakness of this model is that if a user should assign a very high weight (e.g. 200) to one descriptor and a lower weight (e.g. 10) to the rest of the descriptors, the user could cause the former to "overpower" the rest. Average users may also find this process of

assigning a weighting factor confusing, possibly introducing mistakes (Blair, 1990: 38-39).

2.2.1.6   Model 6

In this model, the binary assignment (assignment of AND, OR, NOT, etc.) specified by the user, is replaced by a manual or automatic weighting system. One method used to assign this weight, is by calculating the total number of times specific descriptors appear in the document. This model provides for more accurate retrieval as well as an easier interface to be used by the searcher. Again, processing time could increase due to operation of this weighting process. A further problem is the issue where a searcher and indexer could disagree about the weight assigned to certain descriptors (Blair, 1990: 41-42).

2.2.1.7   Model 7

Model 7 represents a combination of model 5 and 6, whereby retrieved documents are ranked according to weights assigned to descriptors by both searcher and indexer respectively. Due to the formula used to rank results according to the weight assigned by the indexer and searcher, increased processing time is expected (Blair, 1990: 42-43).

2.2.1.8   Model 8

In model 8, research was carried out by combining the weights of descriptors to produce the ranking of retrieved documents. Eventually a conceptual model was used to determine the value rating of a document. In this conceptual model, requests are treated as vectors and then used to calculate the combination of the weights. It is believed that this model could provide a basis for further research into information retrieval. The downside of this model is that it could cause researchers to focus only on this particular method (Blair, 1990: 44-45).

2.2.1.9   Model 9

Model 9 is one of the more popular models. It provides the user with the ability to make use of Boolean operators (AND, OR, NOT) to construct a

complex combination of searches, which will help improve the efficiency of the retrieval process. For example, a request for "Internet 'AND' Surfer" would return results containing descriptors, where a request for "Internet 'OR' Surfer" would return results containing the first, the second or both descriptors. The request "Internet 'NOT' Surfer" would return results containing just the first descriptor by excluding results which include the second descriptor. The shortcoming of this model relates to the combination of Boolean operators, which could become difficult to use for an average user (Blair, 1990: 46-47).

2.2.1.10 Model 10

In this model, the ranking is not just carried out by analysing the weight assigned to the descriptor, but also by analysing the entire text in a document. Text is analysed according to relevancy, excluding non-content words such as "and", "or", "the", "of", etc. and then retrieved with the use of Boolean operators. This method leads to a larger number of retrieval results than other models, and could result in the retrieving of a large number of different and irrelevant search terms. This method could cause the user to think of several possible search terms to use in finding relevant information within a document (Blair, 1990: 47-53).

2.2.1.11 Model 11

Model 11 provides the user with an option to allow the information retrieval system to include semantic words related to the descriptor provided in the query. An example would be a user providing the descriptor "Highly" and the retrieval system includes the words: extremely, very, exceedingly, very much, greatly, etc. The problem with making use of semantics, is that it may not guide the descriptors in the way the user intended and could therefore produce more problems as opposed to a single solution (Blair, 1990: 55- 61).

2.2.1.12 Model 12

In model 12, which is very similar to model 11, the user and indexer has the ability to decide which of the words in the thesaurus relate more strongly to the descriptor entered. The terms are then ranked according to equivalents,

providing the user with a ranked list of terms mapping the original query term. This method is also beneficial to the indexer, as it provides a list of suggested index terms that may be included with a particular selected term to index a given document. One deficiency of this model is that the thesaurus for every term assigned to a newly acquired document, would have to be updated when that document is entered into the database (Blair, 1990: 61-63).

These models provide for a perspective on how indexing occurs in the majority of information retrieval systems and potentially, search engines. It is of importance to note that all of these models provide indexing limited to text items. Information in other formats, such as graphics, sounds, etc. is not included due to the difficulty associated with analysing this type of information. The analogy can be drawn that the majority of systems weigh their index terms according to the number of times a specific term occurs in the text (Ru & Horowitz, 2005: 249).

Figure 2.2 illustrates a basic structure used by search services to index documents and websites. It should be noted that this is not a definitive model, but rather an illustration of how these services could operate.
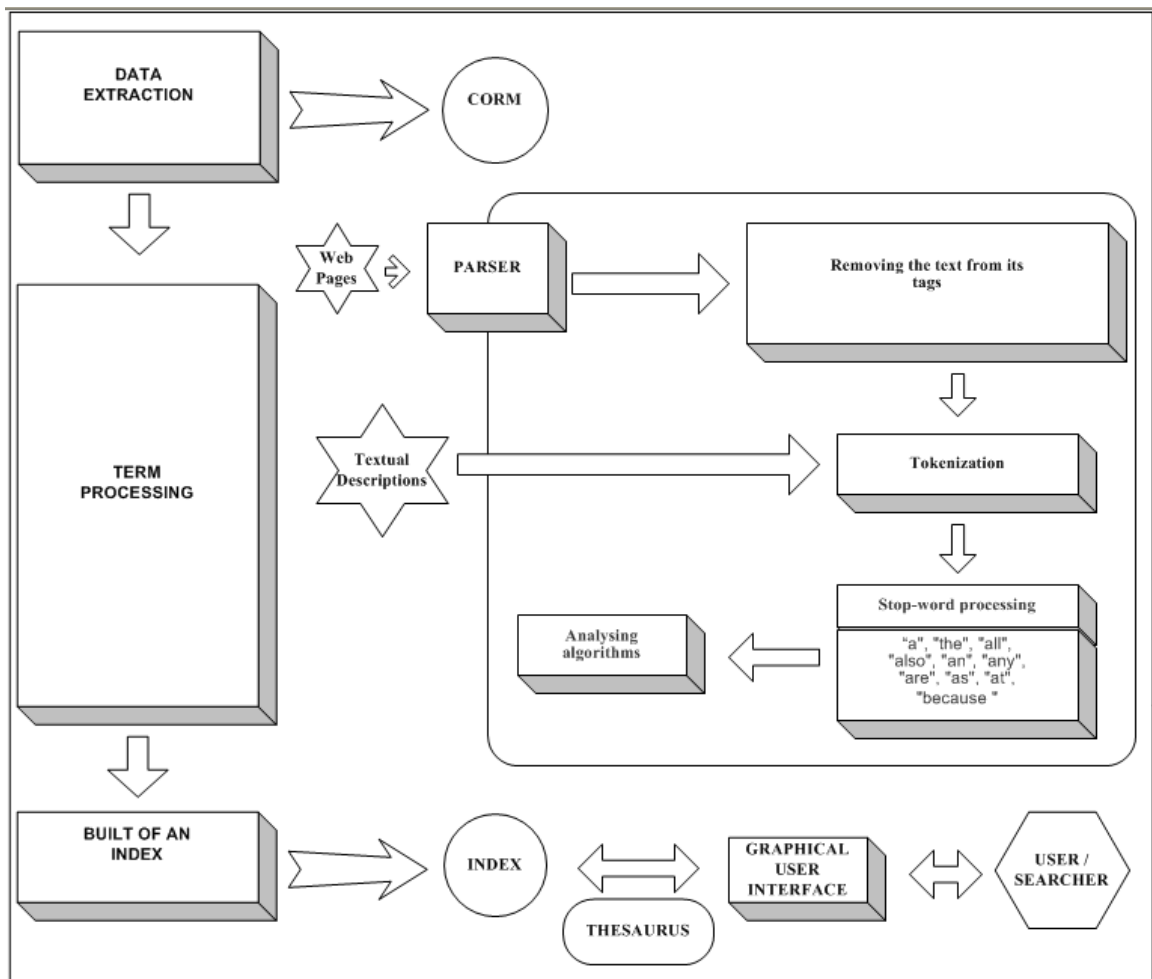
FIGURE 2.2: Illustration of the indexing process (Gandre & Lebbe, 2005: 3)

When a webpage is submitted, it is sent through a parser (algorithm or program) that is used to determine the structure of sentences in the webpage. Text is then removed from the tags contained within the ambit of the document and then "tokenised" by removing the stop words and splitting them up into a set of tokens. The index is then built by using individual algorithms and placing them in the indexing database. A database (see CORM in Figure 2.2) is used to store all the URLs of webpages and the textual descriptions, which require indexing.

When a user/searcher queries the index by making use of the user interface, a thesaurus could be interrogated, providing the user with the opportunity to perform a query extension. The query will then be optimised for index consultation. The results returned will be ranked according to relevancy for the user to analyse, providing an accurate indexing system which is an

essential part of the retrieving process. The Web is one area where the retrieval of information has become almost impossible without the aid of high quality search services guiding searchers towards relevant websites. This is where search engines and directories make a valuable contribution, by ensuring that the searcher's needs for certain information are satisfied (Wilson, 2002; Bennett, 1997: 326; Ru & Horowitz, 2005: 249).

2.2.2   Evolution of the web

The Internet, initially known as a network of networks (Arpanet), was developed during the 1960s specifically for computer science and engineering projects. It was originally designed as a vital communications link among researchers and defence contractors, and was unknown to outside users. In 1989, the USA government decided to stop funding the project and to make this network capability available to everyone. This capability led to its commercial successor, the Internet (Hoffman, 2002; Boyes and Irani, 2004: 191).

Currently, the Internet consists of an interconnected collection of international computer networks joining people and organizations together, and providing the "appearance" of a single network. Being composed of several linked networks, there is no overall Internet authority, and therefore it makes access easy for everyone, regardless of time or distance constraints (Poulter, 1997: 132).

Through the Internet, a user can access a global network of documents, offering several services and also many additional features of multimedia (e.g. sound, video and graphics). Furthermore, anyone with basic knowledge of computers can establish a presence on the Internet by making use of a Hyper Text Mark-up Language (HTML) (Bennett, 1997: 324) which is discussed in Paragraph 3.3.3. Furthermore, the Internet as a low cost communication tool now also permits SMMEs to become global marketers early in their development (Kiani, 1998: 7).

The Internet in its early stages represented a text based interface, with most of its functionality coming from command line use, similar to DOS (Disk Operating System). It was only in 1995, that a browser using the WWW (World Wide Web) with a graphical interface appeared on the market (Notess, 2003: 54). This browser allowed users to access webpages through a graphical user interface. Webpages were then developed to contain text, graphics and other multimedia resources and then linked to Internet using the compact URL addressing scheme (Poulter, 1997: 133).

Search services became a huge success soon after the introduction of the graphical interface of the browser. It provided a much needed service where terms could easily be used to search for specific information. These services generated mass appeal by offering quick and relevant responses (Hubbard, 2004). However, according to Green (2000), many search providers started to neglect their search indexes or web directories during 1996 and 1997, due to their focus on attaining as many users and businesses as possible.

Currently relatively cheap communications technology, software and the growth of Internet service providers offer Internet access to both individuals and organizations at a relatively low cost. Web searching services have become known as everyday tools for information seeking needs (Spink & Xu, 2000). Internet access is now available in the majority of public libraries and in community centres, providing access to information previously only accessible by large organizations (Bester, 2003).

The majority of users rely on search engines to find relevant information on the Internet (Guenther, 2004a: 47). This is due to the freedom of access and ease of use of this method for finding information on the Internet (Weideman, 2002). Therefore, any web author whose income depends on the amount of traffic drawn by a website, should have a clear understanding of how search engines operate and also how a webpage should be written to achieve a higher position in the index of the majority of search services.

The majority of search engines make use of two core indexing services, one being human editors (referred to as a "search engine" throughout the literature) and the other being retrieval software commonly referred to as crawlers. The distinction between these two indexing services became somewhat blurred with the evolution of Meta-search services, which made use of both search engines and web directories to retrieve data. As a result, both these services will be referred to as a search engine, throughout this thesis.

2.2.2.1   Web directories (Manual indexing)

Green (2000) defines a web directory as, "a pre-defined list of websites, compiled selectively by human editors through categorisation, according to subject or topic". A web directory provides a service which allows a user to navigate through several listings and an option to search the entire directory. According to Green (2000), the major web directories also make use of search engine indices to provide secondary results, whenever their human compiled indices fail to produce results matching the user's query. Yahoo serves as an example of a directory. It uses human indexers to examine documents and to identify their principal themes (subjects) using a controlled vocabulary. Web owners need to submit the address of the site in order to have it reviewed and registered, so that it can be found in a search. Once indexed by web directories, it will remain listed within that directory, unless it is removed, for example as a result of excessive use of spam etc. (Green, 2000; Thelwall, 2000: 152; Weideman, 2001: 60).

One possible disadvantage of web directories according to Hubbard (2004), is that they cannot keep up with the capabilities of their automated counterparts. The reason for this being that directories are driven by human power, which is subject to a time factor. Hubbard (2004) is of the opinion that although directories cannot do full-text indexing, their combined efforts to maintain quality subject indices are sufficient to meet most searching needs.

Hubbard (2004) praises the use of directories by stating:

> "There is a potential need for human control in the world of indexing, due to the amount of spamming and potential difficulties computers encounter to comprehend the human language. The human language is filled with synonyms, polysemy, homonyms, spelling variations and slang which make automatic retrieval software rather blunt tools in their attempts to process natural language."

Poulter (1997: 137) states that the two biggest problems with directory search engines are:

- Their ability to keep up with the ever increasing flow of new WWW pages which need individual consideration.
- Preserving an ordered indexed structure as the amount of webpages continually expands.

2.2.2.2 Search Engines (Automated indexing)

According to Green (2000), a search engine is a search service that uses retrieval software called spiders or crawlers that examine websites and then index them into a database of website listings according to their relevancy. Search engines use their own indexing software and strategies to continuously traverse the web, searching for the most up-to-date content possible. The indexing software (also referred to as spiders or crawlers) is responsible for traversing through webpages following links between pages. The pages found are then analysed and parts are copied back to the site running the indexing software and added to the database for the purpose of including them in the search engine results (Poulter, 1997: 134).

Even though the majority of search engines use different algorithms to rank a site, they operate in terms of similar principles. All search engines primarily strive to retrieve and display relevant results (webpages) which contain

words or terms that match the user's search query (Green, 2000; Guenther, 2004a: 47).

Oppenheim *et al.* (2000: 191) are of the opinion that although search engines search a vast amount of information at impressive speeds, they are criticized on issues such as the retrieval of duplicate and irrelevant records due to spamming techniques.

2.2.3   Different types of search services

2.2.3.1   Meta-search engines

A Meta-search engine is a service that returns results derived from a single user request that is submitted to several search engines simultaneously. The result is then analysed and duplicates are usually removed. Meta-search engines do not have a database of websites to be maintained on a sustained basis. Their methodology is to convert the users' search request to a format that is recognised by other search engines (Green, 2000; Guenther, 2004a: 48; Mamma.com, 2005). In layman's terms, a Meta-search engine produces similar results than the use of multiple search engines simultaneously.

Hubbard (2004) states that meta-search engines could offer timesaving features for users doing extensive searches on obscure topics. The author is further of the opinion that when a user is searching for basic information on general topics, it is often not worth using meta-search engines, as this information can be found easily on most other search engines. Research conducted by Henzinger, Motwani and Silverstein (2002) describes this method of retrieval as "incomplete". The reason given for this statement is that, "it is hard to tell whether a user clicking on a page actually ends up finding that page relevant or beneficial".

Some of the most popular meta-search engines include:

- **Dogpile**   The word Dogpile represent a sports term used to describe players piling on top of one another in celebration. Dogpile

was initiated a month after the meta-search engine Mamma.com in 1996. When a user types in a query, Dogpile makes use of a virtual retriever named Arfie, to retrieve the best results from the Internet's top search engines. The retrieved results are then analysed by Dogpile's meta-search technology to remove duplicates. The purified results are then ranked to ensure that the best results appear first on the list (Dogpile.com, 2005).

- **Mamma**   Mamma became one of the first meta-search engines and quickly established itself on the Internet. In February 2004 Mamma.com received an honourable mention in the best meta-search category in the 4th annual SearchEngineWatch Awards (Mamma.com, 2005). Mamma.com also employs a voting service (rSort) for its search results. Instead of simply eliminating the duplicates, it uses the duplicates to rank results. Every duplicate returned is considered a vote for that result. Pages voted higher will eventually appear higher on the result list (Mamma.com, 2005). According to Mamma.com (2005), one big advantage of the rSort service is that it reduces search engine spam, as it is difficult for spammers to spam more that one engine simultaneously, which means that results containing spam will tend to receive fewer votes from multiple sources.

2.2.3.2   Popularity based analysis

Popularity based analysis evolved in April 1998 with the introduction of the new technology of Direct Hit. This new service differs from other search engines in the sense that Direct Hit focuses on a "user-controlled ranking algorithm", based on websites that users have visited ("click popularity") and the time spent on a website ("stickiness") (Green, 2000).

Direct Hit is similar to the meta-search engine concept. It is not a separate search engine with its own index that can be queried directly. Instead, by

using its own database it provides a second level analysis of search results returned from other popular search engines. The "click popularity" of a website is measured by the number of clicks received by each site in a search engine's results page. "Stickiness" is a measurement of the amount of time a user spends at a site. It is calculated according to the time that elapses between each of the user's clicks on the search engine's results page. The results are then re-ranked according to the most popular websites that match the search term displayed in the top of the results list (Green, 2000; Sullivan, 2002).

According to Sullivan (2002), although the results returned by Direct Hit originally looked good, the technology was not well maintained or improved. Spamming became a problem for Direct Hit, as the popularity of the website was largely determined by its "click popularity". This facilitated the manipulation of the results if a person had a good understanding of how this search service operated. It was reported that some companies made use of software that would search for a certain keyword on a Direct Hit based search engine. The software would then analyse the returned results for a company's site and click on it. After remaining on the company's site for a specified amount of time, the program would go back and repeat the process, thus improving the rating of the specified webpage (Searchengines.com, 2003).

2.2.3.3 Natural language searching

The majority of search engines automatically ignore frequently used words (called stop words) such as "or", "to", "not", etc. They do not consider the context of the search terms, i.e. the syntactical relationships between the search terms and other vocabulary within their indices. Furthermore, they search for exact matches and therefore fail to consider semantics or use thesauri (Green, 2000).

The first search engine that addressed these limitations and supported natural language was launched in 1998 and became known as Ask Jeeves. The name was later shortened to Ask. It operated in such a way that when a

user requested information, the engine would match the user's query against a database of seven million template questions. Should no relevant information be found, the engine returned the closest alternatives from the database and the user was then asked to select the most appropriate. It also conducted a Meta search across AltaVista, Go (InfoSeek), Lycos and Yahoo! (Green, 2000).

Another natural language search engine, the Electronic Monk, was launched a few weeks after Ask Jeeves had been launched. Instead of matching the query against a database of template questions, Electronic Monk analysed the query, using natural language algorithms, which also made use of thesauri to consider words with the same meaning. The natural language search was then converted into a complex Boolean query and submitted to AltaVista (Green, 2000).

2.2.3.4 Links-based analysis

Links-based analysis is a technique used to examine the relationships among webpages by ranking them according to hyperlinks. By using this technique, the search engine can identify reliable sources of topic-specific information, highly relevant to user queries (Green, 2000).

One famous search engine that makes use of this technique is Google, which was developed by two students at Stanford University and launched in 1998. The software used to index webpages is known as PageRank, a system responsible for determining the rank of webpages.

PageRank interrogates the web and analyse how webpages link to one another by using its link structure as an indicator of the value of an individual page. If for example page A has a hyperlink pointing to site B, Google perceives it as a vote for site B. In layman's terms, Google rates the target site B higher because site A believes that site B was good enough to be referenced on site A. To avoid the problem of Doorway pages and Link farms (to be discussed in Paragraph 3.3.7.2), Google furthermore analyses the page that casts the vote (page A) for relevancy e.g. if page A has a high

rating, then it would assist page B to receive a higher rating as well (Google, 2005).

Google not only makes use of PageRank, but also combines it with sophisticated text-matching techniques to proven pages that are both important and relevant to a search query (Google, 2005). Links-based analysis is also used by some other search engines, such as Excite and HotBot as part of their algorithm strategy. But according to Green (2000), Google is the only search engine that is exclusively focused on link-based searching.

2.3   CLOSURE

The small business sector makes a valuable contribution to global economic development, however, due to its high failure rate it is of importance to seek ways to assist it. The use of e-commerce has proven to be one area which could provide potential benefits. It provides the ability to access a global information infrastructure, with no boundaries in a global information economy.

It has also become evident from the literature that small businesses often lack IT skills, and may be particularly vulnerable in accepting e-commerce. A lack of general knowledge about the Internet, which may include search engine strategies, website design etc. was identified as possible reasons. It was also highlighted, that although the use of the Internet provides a huge advantage, it does not guarantee success to any company.

Furthermore, by making effective use of the Internet, SMMEs can:

- Bypass several traditional business barriers such as the physical distance between markets.
- Improve old ways of communication.

- Improve the sharing and communication of information across the entire organization.
- Satisfy the need of organizations to exchange information with suppliers and customers to meet information requirements.

Previous research shows that most users only view the top 20 or 30 results produced by search engines. The literature summarized also shows that to use e-commerce effectively, a company needs to ensure that it has a visible website, ranking in the top 30 results in the major search engines. To effectively create a visible website one should have at least general knowledge of how search engines operate. In Chapter 3, the known strategies used by search service algorithms to index webpages will be discussed in detail.

CHAPTER 3
SEARCH ENGINE STRATEGIES

3.1 SEARCH ENGINE STRATEGIES

A well developed website may be completely ignored if potential customers cannot find it. "There are two distinct elements of the Web: the visible and the invisible" (Green, 2000).

Search engines evaluate webpages for relevancy according to the engine's individual algorithms. Guenther (2004a: 48) emphasises this point by stating:

> "Understanding how queries are processed and relevancy established is important for those seeking information, but it is even more important for website developers and content writers, who can directly effect how their websites and web documents are indexed by the different search sites."

It is clear that search engine indexing strategies hold significant challenges. As a result, site objectives should be established clearly before developing a webpage to provide a solid foundation from which specific design elements can be identified (Guenther, 2004b: 54).

There appears to be no shortcut to ensure high search engine rankings and a number of approaches can be employed for this purpose. Achieving sufficient visibility through a search engine requires careful attention to Web page content and search engine registration. Several techniques exist to make a website more visible for the majority of search engines. Some general ranking principles are sometimes discussed in search engines' help sections, and other methods are obvious from research conducted on search results (Notess, 1999: 85). The exact methods however, used by search engines are closely guarded due to the nature of the company's competitive strategies and problems with spam (refer to Spamming in Paragraph 3.3.7).

By improving and implementing methods to increase the web traffic to a website, a web author can also build an improved customer relationship that could later be used effectively to indirectly boost sales revenues.

3.2 SEARCH ENGINE REGISTRATION

Several sources claim that the first and most effective way to improve the visibility of a webpage is to have it listed with the most popular search engines and web directories (Rowley, 2001: 205-207; Srinivasan *et al.*, 1996:79-81; Thurow, 2003: 9-13; Nobles & O'Neil, 2000; Weideman, 2004; Guenther, 2004a: 48). Thelwall (2000) emphasises this point by stating that those sites that fail to register with the leading search engines are missing the potential of what the Web can provide. Table 3.1 reflects the top ten search engines of 2005 including the percentage of searches conducted on each of these search engines.

TABLE 3.1: Top ten Search engines in 2005 (Nielsen/NetRatings, 2005)

Search Engines	Jun-05	Jul-05
Google	47.00%	46.20%
Yahoo!	22.30%	22.50%
MSN	12.50%	12.60%
AOL	5.50%	5.40%
My Way	1.80%	2.20%
Ask Jeeves	1.80%	1.60%
Netscape	0.90%	1.60%
Dogpile	0.80%	0.90%
iWon	1.00%	0.90%
EarthLink	0.80%	0.80%

A site that is registered with a search engine also allows search engine crawlers and human indexers to visit and index these pages, which could later be used by a searcher to search for content of which the URL is unknown (Van Steenderen, 2001).

A website according to Van Steenderen (2001) can be registered in various ways, but three of the well known methods include:

- Submission by hand, where an author submits a website manually to each search engine.
- Submission with the assistance of software, where the author makes use of software to automatically submit a website to multiple search engines.
- Submission with the assistance of free submission sites, where the author makes use of a web service to automatically submit a website to multiple search engines.

The last two methods mentioned above provide arguably the fastest way to be registered with a large number of search engines. Developers are however cautioned when making use of these services, as difficulties can be encountered, such as submission firms which provide a temporary service only to collect e-mails to use later for e-mail spamming (Galon, 1999: 130). Should a developer utilise the services of automated submission firms, it is recommended that the developer should seek the assistance of trustworthy companies.

3.3   FACTORS AFFECTING WEBSITE VISIBILITY

Although search engine registration is possibly the first and most effective step to achieve high website visibility, it is not a guarantee to draw large volumes of traffic to a particular website. A developer should also adopt Web development standards for HTML use, metadata, and content structure prior to and during the development of a website.

According to Thurow (2003: 19), to achieve the best results in the indexing process of any of the major search engines, a developer should include the following components within the context of a website:

- Text component.
- Link component.
- Popularity component.

Very few webpages would entirely satisfy each one of these three requirements, however, by including some of the basic elements of these components in the design phase of a website, traffic could be drawn in the early stages of the website's existence. Effective usage of the elements should therefore also make future acquisition of customers less expensive and furthermore provide early economic value (Rajgopal, Kotha, & Venkatachalam, 2000: 3). Furthermore, with visibility elements in place, the web author could start shifting his attention to ways of retaining website users.

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

A website is usually developed as either a "static" or a "dynamic system", or a combination of both. Static pages are manually created, updated and registered (see Paragraph 3.2) by a developer, usually using HTML. This is probably the most cost effective method of providing users with holistic company information on a global scale, by making use of text and images (Green, 2000; van der Westhuizen, 2001).

Dynamic pages are most often automatically created, by making use of computer software to automate certain application tasks. Examples of such software include JavaScript (see Paragraph 3.3.12), VBScript, ASP, PHP, etc. Dynamic systems are usually implemented for websites, which require more complex functionality than a static website, e.g. interactive e-commerce based webpages. Dynamic pages are used on websites where users request or submit information from a static page, which is then supplied or processed by a database onto a dynamic page. The problem with dynamic webpages is that they usually consist of a "dense collection" of data and information,

which average search engines find difficult to analyse (Green, 2000; van der Westhuizen, 2001).

The URL which these dynamic pages produce is also not very clear or understandable due to its interaction with the database. For example, the URL of a static webpage could be "http://www.mwe.co.za/cycling.htm", however a page produced with PHP for example, could be displayed as "http://www.mwe.co.za/cycling/asp?products_no=5&product_price=asc". This URL could make it very difficult for a search engine crawler to determine the focus of the website and therefore could ultimately cause the particular webpage to receive a lower ranking (Thurow, 2003: 153). The content of static webpages contains several elements, which could make a valuable contribution to where a website is listed in search engine indices and are discussed below.

3.3.1   Frames

Frames allow webpages to be split into a number of separate rectangular areas in the browser window, each independent of the other. The problem is that only the "starting page" can be referenced correctly within an Internet link.

Previously, frames were viewed as a beneficial feature to implement on a webpage. It allowed pages to be split into separate areas in the browser window, which kept the user from continuously scrolling down to look for what he/she requires. Currently, frames have become less popular than when they were first introduced (Thelwall, 2000: 152).

The easiest way to determine if a page is using frames, is to scroll down towards the bottom of the page. If one portion of the page moves, and the other portion remains static, it is a page containing frames (Adams, 2003: 38). An example of where developers make use of frames is when a menu is created within a frame in the top left corner, thus allowing the menu to stay in that corner regardless of how far the user scrolls down on the page.

38

Frames present several problems, as crawlers often ignore text contained in frames, which significantly reduces the chance of users visiting it, and therefore should be avoided where possible (Thelwall, 2000: 157; Galon, 1999: 71; Thurow, 2003: 141). It is also known that some websites could index each frame as a distinct page, which would culminate into the site being displayed partially. This means that if the content matching a query is reflected in a pane, visitors clicking on those links will see only the pane, not the full page, as it was originally designed.

A site containing frames viewed in older browsers could also produce several problems when printing a web document. One of which is when a user prints information from a webpage containing frames, the printer often prints the information in the frame last clicked upon (Adams, 2003: 38). For example, in Figure 3.1, should a user click on frame 1 and then click on the print button, the printer would only print the information in frame 1. If the user wants to print the information in frame 3, the user should first click on frame 3 and then print the page.

FIGURE 3.1: Layout of website containing frames

Some techniques do however exist to increase the visibility of a site which contains frames. One of these techniques include making use of the "<NOFRAMES>" tag, which basically operates the same way as the "<NOSCRIPT>" tag discussed in Paragraph 3.3.12. Making use of this tag, provides an alternative way of viewing a page if a browser does not support

the frames functionality. Thurow (2003: 141), recommends making use of this tag whenever designing a site containing frames, as it would provide an area in the HTML code, to allocate valuable information containing keyword-rich phrases for the crawler to analyse.

3.3.2 Hypertext / Anchor Text

Hypertext, also known as Anchor text, has basically the same functionality as a banner (see Paragraph 3.3.9). Hypertext is a highlighted area, which provides access to a different location within the original document, or another document, image, movie or other resource (Henzinger *et al.*, 2002: 5). When well structured on a website, hypertext could also provide users with the ability to decide for themselves what level of detail they need (Guenther, 2004b: 56).

Notess (1999: 85) gives the following example of the use of hypertext:

> "…a Web page that mentions both the publisher Bowker and offers a link to Bowker's Web site from the word "Bowker," has "Bowker" as the hypertext."

Rowley (2001: 208) is of the opinion that the domain and brand name of a website need to be made "memorable" to the user. One way to achieve this, is to include one or both of these elements in hyperlinks throughout a webpage. The use of established brand names in a URL would not only significantly increase the chances of a customer being able to use existing brand knowledge to locate a website, but would also provide keyword-prominent links.

Hypertext can be seen as an effective way to improve the visibility of a website due to the linking capability to other webpages or websites (see Paragraph 3.3.11). A further advantage of hypertext is the small amount of data that is required to display them, compared to the size of a banner link, which usually consists of graphics (Weideman & Haig-Smith, 2002: 285).

According to Henzinger *et al.* (2002: 5), some search engines benefit from including hypertext analysis, due to the resemblance between the hypertext description and the hyperlink. This resemblance makes it difficult for spamming techniques to be successful, as it could impact upon the perception of the user as well.

A site map is yet another method that could be used to boost the effectiveness of hyperlinks. Many search engines allow a web developer to submit only the main page of a site (Nobles & O'Neil, 2000: 94), which limits keywords and hyperlink elements drastically. By providing a site map, a developer not only provides the crawler with an easy navigation path throughout the entire website, but also assists in increasing the usability of the website to the user.

3.3.3   HTML

The structure of data plays a major role in how data is retrieved in a search. According to Blair, (1990: 122):

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

To understand the features affecting visibility and what search engines can use to index webpages, developers should know the structure of an HTML document. A typical HTML document (webpage) generally consists of two parts: A "HEAD", and a "BODY" section. The header section is normally used to provide information about the content of a webpage, which is not displayed as part of the webpage. The body section contains several meta-tags, which determine how the browser will display the webpage. All tags within the document should be opened (<>) when used, and closed (</>) again when completed. For example, when using the tag <Title> one must close it again

with the tag </Title>. A basic HTML document structure example is listed below.

```
<HTML>
<HEAD>
  <TITLE>Title of Website</TITLE>
</HEAD>
<BODY>
  ---Code which determines how the webpage appears to the user----
</BODY>
</HTML>
```

Webpages written in HTML are neither free text nor well-structured data, as reflected in Figure 3.2.

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

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

FIGURE 3.2: Example of HTML code

HTML coding provides limited structural information. It is typically used to control the layout of the webpage. The code may seem dysfunctional by looking at it, but it is in fact a particularly valuable source of metadata. The majority of metadata is usually invisible to the user and therefore provides opportunities for several spamming techniques (see Paragraph 3.3.7). The

value of the metadata controlling the layout of the webpage in HTML is that it is more difficult to use spamming techniques without affecting the experience of the user (Henzinger *et al.*, 2002: 2).

Although webpages might seem to be unstructured, they do provide some structure exhibited through the use of HTML coding. According to Henzinger *et al.* (2002: 9):

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

According to Henzinger *et al.* (2002: 9), the search engine can determine the weight of a phrase in several ways by analysing the data provided within the HTML code which is discussed in the paragraphs to follow.

3.3.4   Keywords

A study by Duncan and Fichter (2004) to establish how users search and navigate websites returned that:

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

The use of effective keywords is extremely important in describing the content of a webpage, as search engines extract information (e.g. position and frequency of keywords) from the initial content of a webpage to effectively categorise it (Guenther, 2004a: 48). One way of choosing

appropriate keywords is to visualize carefully how users would conduct a search and then select and apply these keywords accordingly throughout the site. Website authors should also try to avoid repeating a keyword more than twice as it could be viewed as a potential spamming technique (Craven, 2001: 203).

Choosing effective keywords for a website "…is not simply a question of what data elements are to be included, in what order and with what punctuation, font and capitalisation, but also of how the values of those data elements are to be derived" (Craven, 2002).

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

 Galon (1999: 53) disagrees with Henzinger *et al*. (2002: 9) by stating that a developer should include misspelt words in Meta-description and Meta-keyword tags, compensating for keywords that are often misspelt by users. Galon, (1999: 53) is furthermore of the opinion that by using this technique, a webpage could reach a top position in the results page when a keyword is misspelt and searched for. This statement could be questioned as Google, Yahoo and several other search engines now also provide a spellchecker, which automatically checks the spelling of a word. When a word is misspelt, the search engine would suggest an alternative spelling based on occurrences of all words on the Internet, which could decrease the use of websites returned with misspelt keywords. As a result, this spellchecking capability could make the statement by Henzinger *et al*. (2002: 9) more appropriate.

Research indicates that search engines consider words closer to the top of a webpage and words at the beginning of a paragraph to be more important

than words, which appear towards the bottom of a page or towards the end of a paragraph (Guenther, 2004a: 47; Kritzinger & Weideman, 2005). Website authors should therefore make use of the inverted pyramid writing style. This style consists of a top-down approach, where content is displayed in the order of most relevant to least relevant, benefiting the author of a site (website visibility) and the user visiting the site.

In the example listed below, a description was drawn from a cycling website and then changed to contain an improved keyword prominence:

ORIGINAL

"The South African Cycling website is provided to cover all topics of Cycling in South Africa: Touring, Important events, Racing, Pedal Power organizations, Local Provincial Bodies, National Federation, Mountain Biking and Club information"

IMPROVED VERSION

"This South African cycling website provides cover on all topics of cycling in South Africa: touring, important events, racing, pedal power organizations, local provincial bodies, national federation, mountain biking, road deaths and club information"

It should be noted that in the example above the improved version contains a limited number of capital letters and very few stop words. Users often type words in a search engine in lowercase and search engines also often ignore stop words rendering them useless (Thurow, 2003: 50).

3.3.5   HTML Naming Conventions

Prominent keywords should be used throughout a website as stated in Paragraph 3.3.4. These keywords should be included in the HTML page names (URL). All page names, except the entrance page name (index.html or homepage.html) should contain keywords in their naming conventions, not exceeding 30 characters, which provide content rich information about the

webpage (Galon, 1999: 56). Examples include changing a page containing cycling news, with the URL of news.html to cycling-news.html which now contains both keywords cycling and news, telling the indexer and user that the page contains news on cycling. It should also be noted that hyphens as apposed to underscores should be used when separating keywords in filenames. Some search engines would see "cycling-news.html" as two keywords e.g. "cycling" + "news", while it would see "cycling_news.html" as "cyclingnews" (Galon 1999: 56).

3.3.6   Meta-tags

Meta-tags are non-displaying HTML tags used to provide optional information about a webpage and its content, quality, condition, and other characteristics (Alimohammadi, 2003: 239). According to Henzinger *et al*. (2002: 7), meta-tags are currently the primary way to include and organise metadata (data about data) within HTML. This is also one retrieval method, due to its keyword richness, that is still being included by the majority of search engine algorithms when it comes to indexing web content (Guenther, 2004a: 47).

Examples of meta-tags include: the title tag, description tag, keyword tag, robot tag, etc. These meta-tags should be placed near the upper part of an HTML document, between the <HEAD> tag and the </HEAD> tag as reflected in Figure 3.3.

```
<HTML>
<HEAD>

<TITLE>South Africa Cycling - news, fun rides, tours, clubs and much more</TITLE>

<META NAME="DESCRIPTION" CONTENT=""South African Cycling website provide cover on
                          all topics of Cycling in South Africa: Touring,
                          Important events, Racing, Pedal Power Organisations,
                          Local Provincial Bodies, National Federation,
                          Mountain Biking, road deaths and club information" ">
<META NAME="KEYWORDS" CONTENT="road bikes, roadbikes, cycle, cycles, road cycling,
                          roadcycling, ride, SA cycling, SA-Cycling, sacycling,
                          team, cycle club, cycling club, cycling, South Africa
                          cycling, cycle touring, cycle tour, WP cycling, best
                          cycling tours, fun ride, funrides, cicling, bicycle,
                          bicycling, rapport, giro del capo, giro, cape argus,
                          david cowie, bike touring, RIP, hill, cycle news,
                          cycling news, pedal power, PPA, WPPA, track cycling,
                          rules of cycling, Tour de France, professional cycling,
                          pro cycling , procycling, track, fixed-gear, ride">
</head>
<body>
      ----Code which determine how webpage appear to user-----
</body>
</html>
```

FIGURE 3.3: Example of the use of meta-tags in a webpage

Although the use of meta-tags is one possible method to increase the weight and relevancy of a webpage, it does not guarantee a top place in the search results. This is primarily due to the spamming techniques (see also Paragraph 3.3.7) developed to manipulate search engines into indexing sites not relevant to a search request (Guenther, 2004a: 47).

In a study by Smith (2002), who surveyed 33 e-journals, it was found that only 19.6% of the surveyed pages made use of meta-tags. This raises the question whether meta-tags are still used or required? Although the addition of meta-tags does not always improve visibility of webpages in several search engines, they should still be included for those search engines that still support them. By providing these tags the developer also provides a permanent record of work carried out. Furthermore, a collection of keywords is also provided, which could later be used when submitting a site to a search engine.

In many cases, visibility can effectively be increased with the use of these indexing features built into HTML (see Paragraph 3.3.3). Dublin Core and some of the well known meta-tags, such as the title, keyword and description meta-tags built into HTML are discussed below.

3.3.6.1   Dublin Core

The Dublin Core metadata technology was developed in 1995 by a selection of 52 researchers and professionals from librarianship, computer science, text encoding, and other related areas. The participants produced a set of metadata elements called Dublin Metadata Core Element Set (or for short, Dublin Core). The primary objective of these metadata elements was to provide improved information discovery in the networked environment, clear enough to be understood by the average user (Weibel, Godby, Miller & Daniel 1995; Srinivasan *et al.*, 1996:81). Dublin Core metadata also attracted a broad spectrum of communities, including libraries, museums, government agencies, archives and commercial organizations.

Some of the goals set out by Dublin Core technology were:

- To identify content and its rightful owner.
- To agree on the terms and conditions of its use and reuse.
- To be able to share this information in a reliable way that makes it easier to find.
- To retrieve the date of a publication or data of a find (in the case of artefact), etc. (Bearman, Miller, Rust, Trant, & Weibel, 1999).

Examples of how some of these elements could appear in a website are depicted in Figure 3.4:

```
<!-- Start Dublin Core - Do Not Modify Tags in This Block -->
<!-- Dublin Core Meta Tags generated by TagGen - The Meta Tag Management System -->
<meta name="DC.Title" content="Bridges: Minnesota's environmental information search">
<meta name="DC.Subject" scheme="AACR2" content="Foundations Project (Minn.)">
<meta name="DC.Subject" scheme="LIV-MN" content="Environmental information -- Minnesota">
<LINK REL=SCHEMA.DC HREF="http://www.bridges.state.mn.us/servlet/lexico">
<meta name="DC.Description" content="Search interface to Minnesota state agency online envir
<meta name="DC.Creator.PersonalName" scheme="AACR2" content="Olson, Robert">
<meta name="DC.Publisher.CorporateName" scheme="AACR2" content="Minnesota Dept. of Natural R
<meta name="DC.Contributor.Editor" scheme="AACR2" content="Quam, Eileen">
<meta name="DC.Contributor.Editor" scheme="AACR2" content="Pellinen, Brent">
<meta name="DC.Contributor.PageDesigner" scheme="AACR2" content="Wadsworth, Lynn">
<meta name="DC.Date.Creation" scheme="ISO 8601" content="1998-07-31">
<meta name="DC.Date.Modified" scheme="ISO 8601" content="2002-10-15">
<meta name="DC.Type" content="Text">
<meta name="DC.Format" scheme="HTML" content="text/html">
<meta name="DC.Identifier" scheme="URL" content="bridges.state.mn.us/index.html">
<meta name="DC.Language" scheme="ISO 639-1" content="en">
<LINK REL=SCHEMA.dc HREF="http://purl.org/dc/documents/rec-dces-19990702.htm">
<!-- End Dublin Core - Do Not Modify This Block -->
```

FIGURE 3.4: Example of how elements could appear in a website (Bridges, 2003)

According to Weibel *et al.* (1995), the developer could provide the following information to the indexer and/or user, by including the Dublin Core meta-tags within a website:

- Metadata, in a format that automated resource discovery tools could collect.
- Basis for a more detailed cataloguing record, if the need arises.
- More detail (e.g. origin) of the website to the user.

A study by Weideman and Kritzinger (2003: 235) to determine the number of websites that do make use of Dublin Core concluded that 0 out of a randomly selected sample of 200 websites implemented Dublin Core in their code. This

raises the question of how beneficial these tags really are in the commercial world.

3.3.6.2   Meta-title tag

Research by Craven (2003: 8) to determine the relevancy of meta-tags clearly showed that the title tag carries the highest weight of all the meta-tags. As a result, developers should ensure that the most relevant keywords are used within the title tags, containing no duplicates, not exceeding ten words and finally containing as few as possible non-content (stop) words, such as "and", "or", "the", "of", etc. By providing a meaningful description for the title, the developer ensures that both the indexer and the users get a good idea of what the page is about (Galon, 1999: 57).

The title tag contains four important functions that include:

- The title text carries the highest weight of the metadata in the majority of search engine algorithms (Craven, 2003: 8).
- The title text is the first text to be shown in the search results, describing your webpage (see Figure 3.5).
- The title text is saved when a user adds a webpage to bookmarks and favourites (see Figure 3.6).
- The title text appears on top of the website page when it is displayed (see Figure 3.6).

FIGURE 3.5: Example of how the title tag appears in search engine results

An effective title with keyword-rich content would not only boost the visibility of a website, but would also improve its usability. The title would appear in several areas for the user to see (as indicated in Figure 3.5 and Figure 3.6),

which could provide the necessary information to satisfy the need of the searcher.

FIGURE 3.6: Example of how a title tag would appear in a webpage and favourites

Placing a website's address within a description, should be avoided as it fails to provide a distinct description of the website. If the user knows the address, the question could be asked, why would the user search for it in a search service? Furthermore, Galon (1999: 57) recommends website authors to include the business name (when appropriate), at the end of the description due to its uniqueness. Rather include keywords relating to the product or service offered by the website more towards the beginning of the title description, allowing more weight to be allocated to it.

3.3.6.3   Meta-description tag

A description tag is used to provide a short description of a particular webpage. Some search engine algorithms still analyse these descriptions for indexing, while other search engines in some cases use this description to display a summary of the site below the title, when it is displayed in the

results (see Figure 3.7). According to Craven (2003), the description should be no more than 200 characters. Furthermore, it should be concise and it should not match the text in the title tag.

FIGURE 3.7: Example of how the description tag would appear in search results

To compile a description tag that best describes the nature of a particular webpage, one could use keywords drawn from the title or meta-tagged keywords, keywords from the upper section of text on the page and also text and passages emphasized by header tags (Craven, 2003). It is also advised to make use of lower case as opposed to uppercase, as users usually make use of lowercase when typing in terms and also keeping in mind that some search engines are case sensitive (Thurow, 2003: 78). Developers should be careful not to use keywords more than twice in a sentence, as it could be perceived as a possible spamming technique.

Of all the pages on a website, the home page tends to be less likely to retain its description in the meta-tag, due to the frequent redesigning of home pages. Usually, when using various webpage editing software, the descriptions are easily overlooked due to their invisibility when previewing the page (Craven, 2001). As a result, the responsibility of a website author should not only be to create effective Meta-description tags, but also to regularly revise these description tags as pages change over time.

3.3.6.4   Meta-keyword tag

The Meta-keyword tag is used to list key words, which would accurately describe the content of the site. Alternatively these keywords also assist search engines to categorise websites, by providing words under which the site is searchable. An example of how this element can be implemented is depicted in Figure 3.3.

To compile a list of keywords, an author should make use of distinct lists of keywords, which relate to the specific page on the site as opposed to using one broad set of keywords for every page. Alimohammadi (2003: 241) also recommends making use of acronyms, synonyms, Americanisms, related words and also a thesaurus, when creating a keyword list.

Making use of commas only improves the readability of the words to the developer and has no advantage in the indexing process. According to Thurow (2003: 80), search engines do not take commas into account when evaluating meta-tags in a webpage. For example, the text "bike touring, hills, cycle news, cycling news, pedal power, PPA, track cycling" is seen by a search engine as "bike touring hills cycle news cycling news pedal power PPA track cycling." Note the repetition of the words "news" and "cycling" in the example above.

Due to the increased sophistication of spamming techniques used by websites (see Paragraph 3.3.7), Meta-keyword usage has deteriorated dramatically and is in many cases ignored. Automated indexing search engines currently focus more on text throughout the webpage, which includes text in alt tags, description tags, header tags, etc. Even though keyword tags are rarely used, they should still be included for those few search engines that still utilize them in the ranking process which could lead to a higher listing.

3.3.6.5   Meta-header tag

Research by Craven (2003) to determine the relevant weight of meta-tags, indicates that the H1 (heading 1) and H2 (heading 2) tags are the second and third most highly weighted (after the title tag) of all the meta-tags. As with Microsoft Word, HTML has also built in styles for headings to differentiate among importance levels of text that are usually used to break up text into paragraphs. The different options allow the designer to develop large and bold text in an HTML document, marking the beginning of a new paragraph or section (Henzinger *et al.*, 2002: 9).

In HTML, there are six levels of headings. H1 is the most important; H2 is slightly less important, down to H6 which is the least important of the six importance levels. Some search engines recognise the use of header tags as a safe method to weight keywords, due to its connection with a heading of a paragraph. Henzinger *et al.* (2002: 9) states that the higher the importance of the headings, the more weight a search engine could assign to a given webpage. For example, text in an <H1> (as listed below) would appear prominently on a webpage and therefore some search engines could see it as safe to weight the text highly in the heading.

As an example, the code:

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

will be displayed as:

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6

It is recommended that the author include a number of distinct keywords in several header tags, raising the importance of the keywords by displaying the keywords in bold.

3.3.6.6 Meta-robot tag

Robot tags in a webpage tell the crawlers of the search engine how deep they can index the webpage and how often to return for re-indexing. This method is most frequently used when a developer wishes to block pages, which would make little contribution to the visibility of the website. Examples where a developer could use the robot tag to block robots include:

- A page containing mostly script which is unreadable by a crawler.
- A page containing excessive use of graphics, which is unreadable by a crawler.

There are several methods available to control the access of robots. Two of these examples are discussed below.

One method for blocking a crawler from indexing a webpage includes adding a robot meta-tag within the HEAD of an HTML document. An example is listed below:

<META-NAME ='robots' CONTENT='noindex, nofollow'>

In the above example, the "noindex" directive informs the robot not to index the page. The "nofollow" directive informs the robot not to follow the links contained inside the page. The "index" directive does the opposite of the "noindex" directive and the "follow" directive does the opposite of the "nofollow" directive (Galon, 1999: 138).

The second method to control how a crawler indexes a site, is with the use of a robots.txt file. When a search engine robot examines a website, it would usually first search for a file with the name "robots.txt" which contains the instruction of how the webpage should be indexed. If no such file were found, it would proceed in checking for the robot meta-tag in the HTML code; and if such a tag is not found, normal indexing would occur (Thurow, 2003: 109).

The use of the "robots.txt" file also allows the developer to add all indexing commands inside one file as apposed to every webpage, as with the method described above.

In this robot.txt file, the developer has the ability to indicate which part of the site should not be indexed (e.g. files containing Script, password protected folders, folders which contain only images) by specific search engine robots.

An example of commands within a robots.txt file follows:

> User-agent: *
> Disallow: /members/
> Disallow: /protected/
> Disallow: /private/
> Disallow: /clients/
> Disallow: /scripts/
> Disallow: personal_details.html

The "User-agent" specifies the robot name. In the example above the asterisk indicates the use of all robots. The asterisk could also be replaced with the name of a search engine robot, for example "googlebot" which would specifically focus on Google's retrieval software (crawler). The "Disallow" specifies the directory or file which should be excluded from the indexing process. The use of "/members/" specifies a directory while "personal_details.html" specifies a file.

3.3.6.7   Alt tags

Alt (alternative) tags are used to display text in the place of an image on a Web page if graphics are turned off. The alt text will also display if a user places his/her mouse pointer over an image for a few seconds. Currently, automated crawlers can read only text elements within a webpage and are unable to read multimedia elements, as discussed earlier. For this reason, it is of importance to apply alt tags where possible, accurately describing the graphics on the webpage (Hubbard, 2004).

By implementing ALT text within a webpage, the developer ultimately caters for users who turn graphics off, to increase loading speed. Without the use of ALT tags, a site could become impossible to navigate when graphics are turned off. The use of ALT tags also provides the ability to cater for users with a visual disability. Voice-output screen readers (benefiting the blind) will not "*read*" a non-text element (Oppenheim *et al.*, 2000: 204). Ultimately, the improvement of techniques to allow the blind to access a website, could also play a major role in improving customer relationships which could lead to a competitive advantage.

3.3.7   Spamming

Since the use of search engines is regarded to be a very popular way of finding information on the web, website authors more often than not turn to manipulative ways to get a higher placing in the top results of search engines, as their income depends on traffic volumes. This process, to deliberately manipulate search engine ranking algorithms by making use of unethical techniques to achieve higher website visibility, is often called search engine spam (Henzinger *et al.*, 2002: 2; Notess, 1999: 84). Search engine spam in search engine context differs from traditional unsolicited e-mail spam and exists in many forms. Thurow (2003: 220-228), provided a broad list of sixteen tactics that are used as search engine spam:

- Keywords unrelated to a site.
- Redirects.
- Keyword stuffing.
- Mirror/duplicate content.
- Tiny Text.
- Doorway pages.
- Link Farms.
- Cloaking.
- Keyword stacking.
- Gibberish.

- Hidden text.
- Domain Spam.
- Hidden links.
- Mini/micro-sites.
- Page Swapping (bait & switch).
- Typo spam and cyber squatting.

According to Henzinger *et al.* (2002: 3) and Thurow (2003: 220-228), spamming approaches can generally be divided into four categories, as discussed below.

3.3.7.1 Text Spam

Initially, Meta-data was designed to allow web authors to both describe their webpage and to specify certain keywords related to their webpage (see Paragraph 3.6). This Meta-data was ranked highly by search engines, as it was meant to describe exactly what the site was about. Spammers quickly created techniques to modify the text in such a way that the search engine rated the page higher than it deserved (Henzinger *et al.*, 2002: 3).

An example of this kind of spam is keyword stacking, where one or more keywords are repeated several times, in such a way that it does not disturb the user. These keywords are then usually placed in areas unnoticeable by the user. Some of these areas include presenting text in the smallest size and in the same colour as the background, or using them in meta-tags (see Paragraph 3.6), noframes tags (see Paragraph 3.3.1), noscript tags (see Paragraph 3.3.12) etc. (Thurow, 2003: 221; Notess, 1999: 85).

A further method used by sites containing pornographic content to attain top search engine visibility, is to place words on their webpages that are not related to the actual content of the pages. An example would be where a page contains names of popular celebrities in order to make the page appear relevant when a user searches for such celebrities (Henzinger *et al.*, 2002: 4). This spamming technique is often used to attract users searching for

popularity of the word and not necessarily because the word relates to the content of the webpage.

3.3.7.2 Link Spam

The links-based analysis technique used by some search engines, e.g. Google (see Paragraph 2.2.3.4) has been countered by an effort of spammers to manipulate the link analysis systems. An example would be a "link farm", where developers try to create multiple websites with the sole purpose of linking them to one another to increase their ranking position. These links, which usually appear on the bottom of a webpage, more often than not use exclusively incoming links to increase the visibility of the webpage (Thurow, 2003: 225; Henzinger *et al.*, 2002: 4).

A further form of link farm that is more sophisticated is known as "doorway pages". These pages consist entirely of links pointing to sites and are usually not intended to be viewed by users conducting a search (Henzinger *et al.*, 2002: 4), which is why search engines consider these types of webpages as spam.

Google tries to avoid this kind of spamming by adding another step to its ranking algorithm. Google weighs hyperlinks from high-ranking sites higher than hyperlinks from poorly ranked sites. Combining this source authority with the anchor text references, can achieve highly relevant results (Notess, 1999: 85).

3.3.7.3 Cloaking Spam

The cloaking technique is very popular in websites containing intense multimedia content. In such a webpage, the developer often creates a separate text page to assist the search engine ranking on the site. However, cloaking is not always used for this reason. Spammers also make use of cloaking to manipulate search engines without affecting the content read or used by the user (Henzinger *et al.*, 2002: 4). According to Thurow (2003: 227), all the major search engines consider cloaking to be spam.

An example of how this spamming technique is used is the implementation of an auto-refresh tag in the header of a webpage. The auto refresh tag is used to display a certain page (e.g. welcome screen) for a certain number of seconds, and then, after the time has elapsed, the page loads another page (e.g. homepage) containing other information. Spammers soon made use of this method to mislead search engines by setting the time sequence to a split second, causing the user to see only the second page. The first page was then loaded with content, which was basically unreadable by the user but most relevant for the search engines (Galon, 1999: 70).

3.3.7.4   Domain Spam

Domain Spam (also called Mirrored Sites), is closely related to content duplication. It is the process where a developer purchases multiple domain names and then places websites with identical content in these different domain names. By using this technique, spammers anticipated to obtain multiple listings in search engine rankings through the boosting of the link popularity of the particular website, which could ultimately result in the increase of traffic (Thurow, 2003: 227). The use of this technique should be avoided at all times, as it could result in getting the site banned from a search engine's index. According to Thurow (2003: 227), search engines have developed techniques to remove all mirror sites from their index.

Spamming appears to be a never-ending battle between the spammer and search engines. As soon as new search engine techniques develop, new spam techniques develop as well. To stay on top, search engines have to continuously develop and improve techniques for detecting and fighting spam. Due to this, search engines often fail to publish their anti-spam techniques in the hope of making it more difficult for the spammers to succeed (Henzinger *et al.*, 2002: 6).

59

Some countermeasures against spam that have been identified by Henzinger *et al.* (2002: 1-9) include:

- Search engines ignoring text with same colour as background.
- Search engines detecting certain distinct patterns of link-farms and doorway pages and then rating them lower.
- Search engines ignoring text in auto-refresh tags.

These spamming techniques (see Paragraph 3.3.7) have caused search engines to constantly turn to new ways of ranking webpages. Currently, the textual content of a webpage is still analysed by all search engines, but the amount of weight assigned to this textual content is questionable. By evaluating spamming techniques, the investigative question can be answered, which reads: How has the development of the Internet affected search engine strategies? Spamming techniques should be avoided at all times as it could result in the website being automatically removed from a search engine's index, and black listed (Van Steenderen, 2001).

3.3.8 Multimedia (Graphics, Flash, PDF)

Graphics, Macromedia Flash, sound, video and animation, are all elements which could contribute to increasing the user friendliness of a website. A developer should however proceed with caution when implementing these features, as they are not always visible to search engines' crawlers to analyse (Thurow, 2003: 122).

3.3.8.1 Graphics

The use of graphics should be limited to minimise download time, and also to ensure that there is a certain amount of text on the site. When graphics are used on a site, the developer should at all times ensure that all graphics include descriptive and keyword-rich alt text as discussed in Paragraph 3.3.6.7.

3.3.8.2 **Flash**

Macromedia Flash is a Scalable Vector Graphics (SVG) technology designed to allow a developer to animate graphics, text, sound, create interactive experiences, etc. Macromedia Flash is currently the most widely used software to bring animation and interactive content to a website (Goh & Wang, 2004: 144). Henceforward, "Macromedia Flash" will be referred to as "Flash". Flash is one way to improve interactive experiences of users on a website. It provides ways to deliver images with magnitude and direction as opposed to bitmap images, which provide images based on grouping of pixels. These vector graphics (Flash) have several distinct advantages, making it a powerful tool to use. Some of the advantages include:

- Vector graphics can be coded to be scalable causing the image to resize as the browser window is resized.
- Vector graphics load faster than similar pages containing JavaScript, PHP, and ASP.
- Vector graphics allow interactivity between user and code.
- Vector graphics do not require programming skills as opposed to Java applets.

One example where Flash was successfully implemented to improve customer relationships, was on the USA Mini Cooper website (MINIUSA.COM, 2005), (see Figure 3.8). After the initial load of the Flash file on this site, the users have the ability to choose their own colour, style, model etc. and then have the car with all their chosen preferences, including the price, displayed on the screen.

FIGURE 3.8: Example of how Flash is used in a webpage (MINIUSA.COM)

Even though Flash in several instances improves the interaction between the user and the website, a developer should also study the visibility implications involved when making use of this technology.

Goh and Wang (2004: 144) reported on the experiences faced by the National Archives of Singapore in their online virtual exhibition of the "Colours in the Wind". The authors found several disadvantages with the use of Flash, which are listed below:

- Flash software needs to be licensed, and therefore forces the website developer to purchase the Flash software and thereafter also the upgrades to newer versions.
- Developers have to cater for users who fail to upgrade their browsers to support the latest versions of Flash.

- In cases where the development of a site is outsourced, updates to a site could not be made without the source code as opposed to HTML code, which is accessible immediately at time of download.
- Flash files appear in binary format, which is not readable by search engine crawlers, thus affecting the visibility index in some search engines.

The use of Flash however, has several benefits for the developer and the user in terms of website visibility. Although a few of the major search engines can find some of the links embedded inside a webpage containing Flash, textual content is very difficult to extract (Thurow, 2003: 147).

There are however ways to make use of Flash and still have a visible website. Thurow (2003: 151) recommends the coding of two separate websites, namely one that consists of Flash and a further without Flash, and then providing an option for the user to choose which one to access. By making use of this technique, the developer not only provides a Flash page with a high interactive design, but also a page with content rich keyword phrases.

According to Gerhart (2001), a website containing Flash can still be made visible by keeping the following in mind:

- Not making use of Flash in the navigational structure of a site.
- Not including all important information in Flash.
- Being careful not to make Flash too big, causing sites to load slower on computers with slow connections.

3.3.8.3   PDF (Portable Document Format)

Search engines have become increasingly efficient at indexing document types in addition to standard web formatted documents. For example, google.com can index 14 different types (see Table 3.2) of documents

63

(Google.com, 2004) while Yahoo can search seven different types (see Table 3.2) of documents (Yahoo.com, 2004). Even though several search engines have started to read and index non-HTML file formats, none of these search engine crawlers will follow links appearing within the majority of these formats. As a result, links which a developer want search engines to follow, should be placed in regular HTML pages.

TABLE 3.2: Different file types indexed by Google and Yahoo

	File Types	Google	Yahoo
1	Adobe Portable Document Format (pdf)	✔	✔
2	Adobe PostScript (ps)	✔	
3	Lotus 1-2-3 (wk1- wk5, wki, wks, wku)	✔	
4	Lotus WordPro (lwp)	✔	
5	MacWrite (mw)	✔	
6	Microsoft Excel (xls)	✔	✔
7	Microsoft PowerPoint (ppt)	✔	
8	Microsoft Word (doc)	✔	✔
9	Microsoft Works (wks, wps, wdb)	✔	✔
10	Microsoft Write (wri)	✔	
11	Rich Text Format (rtf)	✔	
12	Shockwave Flash (swf)	✔	
13	Text (ans, txt)	✔	✔
14	Standard HTML (htm, html)	✔	✔
15	RSS or XML feeds (xml, rdf, rss)		✔

The PDF (portable document format) is one of the documents that search engines have been programmed to index. PDF is a file format that captures all the elements of a printed document as an electronic image, which can be viewed, navigated, printed, or forwarded to someone else. Any computer with adequate software can also read PDF files without incurring platform conflicts.

PDF is the preferred format for many companies and educational researchers, as it allows the document to preserve its look and feel, irrespective of which platform is used. However, companies that do decide to make use of PDF files, should take care when designing the document. PDF documents can in most cases only achieve top search engine rankings if

formatted correctly. Thurow (2003: 167) suggests the following formatting techniques:

- PDF documents should contain actual text and not a picture of text.
- Keyword rich text should be included in PDF documents.
- Keep the size of the PDF document limited, to decrease download time.
- If the PDF document is fairly large, consider creating an abstract of the PDF document on an HTML page.

3.3.9   Banner advertising

Banner ads, which are more often than not also part of paid advertising, were believed to be the most used advertising medium on the Internet (Van Steenderen, 2001; Weideman & Haig-Smith, 2002: 285). This advertising medium involves bringing a URL to the attention of potential customers, thereby increasing the chance that they would visit the website being advertised. Banner advertisements are usually placed on high traffic sites at an additional cost or freely on websites of strategic partners (Rowley, 2001: 208). Webpages of the search engines are also known to sell areas on their sites for banner advertising. Banner advertising not only emphasises a URL in order to entice the audience to visit the website, but also provides a hyperlink to the advertised website, which is important in achieving higher visibility (see Paragraph 3.3.2). According to Weideman and Haig-Smith (2002: 285), standard banner advertising is slowly losing ground and largely being replaced by more sophisticated forms of banner advertising, e.g. animated banners with full motion graphics and sound.

3.3.10   Awards

Awards involve design experts evaluating websites according to pre-defined criteria. These awards could range from graphic awards to content awards and are yet another way of improving a website's visibility. Receiving awards

hold several advantages according to Van Steenderen (2001), not just to the user, but also to the developer. Some of the advantages are listed below:

- Awards confirm the content of the site.
- Awards assist others in determining the quality of a site.
- Awards provide a hyperlink to the winning site, which is an element discussed in Paragraph 3.3.2.

Some examples of websites which assign awards include:

- **www.marketme.com:** Focuses on the uniqueness, usefulness, design, layout and the clear and simple navigation of the site.
- **www.webthrower.com:** Focuses mainly on the design of a website, but also looks at elements such as loading time, functionality, content, originality etc. If a site is chosen for site of the month, a screenshot of the winning site is listed on webthrower.com.
- **www.ultrawebaward.com:** Focuses on overall impression, content, presentation/design and technical operation.

3.3.11 Link popularity

As discussed earlier, several search engines focus mainly on the link popularity element. The link popularity of a webpage can be defined as the number and quality of links pointing to a specific webpage.

Two basic methods to increase the link popularity of a site include:

- Links from other websites linking to the developer's site using hypertext.
- Linking to pages within the website using hypertext.

According to Henzinger *et al.* (2002: 6), a search engine typically assumes that when a site provides a link to another site or page, it is because the

author believes that the other site or page has relevant and information rich content. Finding companies to link to a developer's webpage is not as simple as it might seem. According to Thurow (2003: 113), the majority of search engines have developed new techniques to measure the relevancy of links pointing to a webpage. This is mainly due the excessive use of spamming to manipulate the search engine indexing process. Search engines are now analysing the popularity of one site pointing to another to determine the weight given to the site being pointed to. For example, a popular high quality site "A" providing a link to site "B", would cause site "B" to receive a much higher weight than a low quality site "C" such as a link farm or a doorway page (see Paragraph 3.3.7) linking to site "B".

It could therefore be worthwhile browsing only for other high quality sites with similar interests and request that they provide a link to the site (Van Steenderen, 2001; Galon, 1999: 126). A developer should however be cautious not to fall into the trap of adding links to link farms, or doorway pages with no content relating to the developer's site. These methods are often viewed as spamming techniques (see Paragraph 3.7) and could culminate into a site being excluded from an index or being ranked poorly.

3.3.12   JavaScript

JavaScript is one of several script languages usually used to automate certain application tasks, e.g. a site's navigation scheme. An example where this technology is implemented on a website is depicted in Figure 3.9, where JavaScript is used to automate the menu by implementing dropdown functionality. When the user moves the mouse pointer over one of the menu options e.g. "our company", a dropdown menu list would appear and in some cases a second window would appear as illustrated.

FIGURE 3.9: Example of how JavaScript operates in a website (Coca-Cola, 2003)

According to Thurow (2003: 123), JavaScript has become so widespread that it's now considered to be part of many design packages. JavaScript can make a positive contribution to a website's usability and navigation. The problem however is that should this technology not be carefully implemented by using the correct design techniques, it could culminate in websites becoming less visible to the majority of search engine crawlers.

A prominent search engine expert strongly disapproved of the use of JavaScript, Flash and session ID's in an interview conducted by Dickson and Marshall (2004: 28-35). A web crawler views a website as seen through a text browser and not as viewed by the user. The major search engines may look for URL's in JavaScript, but the average crawler may choose or be programmed to ignore it (Dickson & Marshall, 2004: 28-35). An example of how JavaScript code could appear within a webpage's code is shown in Figure 3.10.

```
<html>
<head>

<title>Coca-Cola</title>

<meta name="Keywords" content="The Coca-Cola Company, Coca-Cola, CocaCola,
                               Coca Cola, Coke, Benefit and Refresh Everyone">

<script type="text/javascript" language="JavaScript" src="/js/menu_data.js"></script>

<script language="JavaScript">
        function checkBrowser(){
          this.ver=navigator.appversion
          this.dom=document.getElementById?1:0
          this.ie5=(this.ver.indexof("MSIE 5")>-1 && this.dom)?1:0;
          this.ie4=(document.all && !this.dom)?1:0;
          this.iemac=((!this.pc) && (this.ie5 || this.ns5)==1)
          this.ns4=(document.layers && !this.dom)?1:0;
          this.bw=(this.ie5 || this.ie4 || this.ns4 || this.ns5)
          return this;
          }
</script>

</head>
<body>
        -- Code which determine how code appears to user --
</body>
</html>
```

FIGURE 3.10: Example of how JavaScript could appear in HTML (Coca-Cola, 2003)

Even though JavaScript could decrease search engine visibility due to its language type, several design techniques could be implemented to improve search engine visibility of a website using JavaScript. Some of these techniques include:

Placing the JavaScript code into a separate file:

Search engine crawlers mainly focus on keywords which appear in HTML documents. As a result, JavaScript code in an HTML document has little value improving the ranking position on a webpage. One way of minimising the amount of JavaScript code within an HTML document, is to move the JavaScript code to a text editor and save the file with a ".js" extension. Now the developer only indicates the filename of the file containing the JavaScript code on the webpage (see Figure 3.10). By making use of this technique, the developer not only minimises the amount of unreadable code, but also improves the downloading speed of the site. When a webpage which contains JavaScript code in a separate file loads, the separate file is placed in the computer's cache, causing the webpage to load faster after the initial visit (Thurow, 2003: 130). The developer could then further improve the use of this technique by placing all the JavaScript files within a single directory and then ordering the crawler not to scan the JavaScript code (see

Paragraph 3.6.6). By keeping crawlers from crawling, JavaScript files could also improve the indexing process. According to Thurow (2003: 131), an extended download time could cause the crawler to identify the page as a page containing potential spam.

Making use of the <noscript> tag:

A further design technique to improve the visibility factor, is to make use of the <noscript> and </noscript> tags which are placed between the body tags. These tags allow the developer to include alternative keyword rich content, visible to browsers not supporting JavaScript. This tag would also provide alternative content to the user who prefers to search the web with JavaScript disabled.

Unfortunately, this tag has also fallen prey to spamming due to the ability to hide irrelevant text within. The abuse of this tag has resulted in several crawlers allocating less weight to text in these tags, or in some cases even ignoring it completely (Thurow, 2003: 132). The use of this tag has therefore become questionable. When it is implemented, care should be taken to avoid the site being labelled as a site which contains a piece of code containing spam.

3.4   CLOSURE

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

Every search engine makes use of its own search algorithm to undertake the difficult task of effectively ranking the increasing website submissions faced by search engines. It is therefore a requirement to understand how keywords on a webpage will affect the indexing process. Rowley (2001: 209) and Galon (1999: 125) collectively claim that, if these keywords were implemented correctly throughout a website, it should enhance the likelihood of a webpage receiving a higher index and thus resulting in a competitive advantage. The

use of meta-tags could improve the position of a website on the indices of several search engines. Ignorance of designers not including this meta-tag functionality could culminate in potential income opportunities being lost.

It has also become evident throughout the literature that websites can make use of elements such as frames, JavaScript, graphics, etc. and still be visible. It is imperative for a developer to keep a site as simplistic as possible for the search engine crawler to visit. Simplicity can be evaluated by viewing a developed site through a text editor, and scanning the code for keyword rich sentences.

CHAPTER 4
RESEARCH METHODOLOGY

4.1   INTRODUCTION

Through the literature study it has become evident that there is no clear-cut approach to significantly increase the visibility of a website. It is a complex process, which involves the preference of the website's owners, technical issues and ethical aspects. Making a website visible for search engines is a time-consuming and sustained process.

To bridge the gap of the plethora of diverse factors impacting upon website visibility, a model was developed to serve as a foundation on which SMMEs can build to improve the visibility of their websites.

Some of the factors included in the model are:

- The ranking of elements, from the most used elements to least used elements.
- The implications these elements could have on a website if applied correctly.

In the following paragraphs the author defines the research question, investigative questions, sample size and other relevant factors. Reasons for choosing certain design methods are also stated and motivated.

4.2   RESEARCH QUESTIONS

The research question forming the crux of this thesis, reads as follows:

Can an effective search engine visibility strategy be developed, to ensure that SMMEs draw website traffic which could potentially lead to increased sales?

The following investigative questions will be researched in support of the research question:

- To what extent do SMMEs require assistance to increase their website's visibility to attain higher rankings?
- How have the development of indexing techniques, spamming techniques, and the development of the Internet affected search engine strategies?
- Which website visibility elements are used by websites listed in the top 30 results of leading search engines?

4.3 SURVEY ENVIRONMENT

Search engine visibility has become an essential part of web development strategies around the world, due to the increase in competition. This increase has ultimately culminated in businesses seeking improved techniques and strategies to boost their web competitiveness. Furthermore, the majority of SMMEs often lack the necessary skills to improve their position in the global online market. Some SMMEs turn to outsourcing, which can result in the problem of finding reliable, cost-effective, independent advice which will address their specific need (Baard, 2004). It is therefore evident that there is a need to assist SMMEs in utilising the benefits of presenting a visible website.

It was therefore a requirement to determine how visible SMME websites are on the major search engines, and to develop a model from the results, to assist SMMEs to improve the visibility of their websites.

4.4 MODEL

Cooper and Schindler (2003: 55) define a model "...as a representation of a system that is constructed to study some aspect of that system or the systems as a whole."

Three major functions of modelling according to Cooper and Schindler (2003: 55) include:

- **Descriptive models:** Portray the behaviour of elements in a system where the theory is inadequate or nonexistent.
- **Explicative models:** Explanation of the entire system in detail, unravelling of complexities, or improvement of understanding of their key concepts.
- **Simulation models:** A simulation which, through mathematical equations, attempts to reveal the process relations among concepts.

The model that will be developed does not have to portray a mirror image of all design techniques, neither does it have to provide all the visibility elements. It should however portray the most essential elements required to assist developers in developing a well optimised and visible website. This model should be self-explanatory and sorted in terms of most relevant to least relevant. Methods to implement the elements within a website should be applicable without major modifications to the layout of the website. To comply with the above requirements, a descriptive model was considered to be the best choice.

4.5 SAMPLE

It has become evident from the literature review that SMMEs make a sizeable contribution to the world economy (Department of Trade and

Industry 2001). Not only do SMMEs provide much needed jobs, but also provide financial support to many people. Unfortunately, the average lifespan of an SMME is less than three years, which is of serious concern for both the South African and other governments. One of the reasons for these high failures is global competition. It was therefore decided to select a sample of SMMEs to investigate the extent of the visibility of their websites on the World Wide Web. Furthermore, from this sample, a list of categories will be extracted and used to identify visibility elements used most often by the websites in the top positions.

4.6   RESEARCH METHODS

To fully understand the quality and nature of website data, the specific requirements for each statistical method must be understood before adopting a particular research approach. The following is a short summary of several methods for analysing and collecting data.

4.6.1   Quantitative Approach

Struwig and Stead (2001: 7-8) define quantitative research as:

> "…a form of conclusive research, involving large representative samples and fairly structured data collection procedures."

With quantitative research, the emphasis is placed on the methodology, since it relies on the measurement and analysis of statistical data to determine relationships between entities, which could ultimately culminate in quantifiable conclusions.

Due to the size of samples analysed through quantitative approaches, it is essential to fully understand the nature of the elements required to produce high quality outputs, before starting a survey of a quantitative nature. Furthermore, should an error be identified on a data collection instrument and

this is realised only after execution, there is little a researcher can do to fix it (Eldabi, Irani, Paul & Love, 2002: 65).

Struwig and Stead (2001: 7-8) and Cooper and Schindler (2003: 148) respectively claim that the most common methods used to conduct quantitative research involve exploratory, descriptive and experimental approaches.

4.6.1.1 Exploratory research

According to Cooper and Schindler (2003: 148), a study may be viewed as either exploratory or formal of nature. Struwig and Stead (2001: 7-8) define exploratory research as an investigation with the assistance of the development of hypotheses, or questions into a problem about which little is known for future research. Formal research begins where exploratory has left off by continuing to test the hypotheses or answer the research questions posed (Cooper & Schindler, 2003: 148).

Website visibility is an area where multiple research outputs have already been produced. However, very little research has been carried out to determine how well SMMEs apply the elements required to achieve a high website ranking within the major search engines. For the purpose of this research, an exploratory research design will be deployed to identify how well SMMEs make use of elements required to achieve higher website rankings.

4.6.1.2 Descriptive research

According to Struwig and Stead (2001: 8-9), descriptive research attempts to describe the way things are. It is also used to summarise, organise and simplify data.

Typical questions asked in descriptive studies pertaining to SMME website visibility could include:

- Why do SMMEs fail to produce well-optimised websites?

- Where does the average SMME website appear in the major search engines?

The main difference between exploratory and descriptive research is that in descriptive research a complete and accurate description of a situation can be produced. Research methods, which are sometimes used to conduct descriptive research include case studies and statistical methods (Cooper & Schindler, 2003: 148; Struwig & Stead, 2001: 8-9).

I.   Case Studies

Case study research is a technique usually used for an in-depth contextual analysis of a few conditions, which could often be achieved by answering who, why and how questions. This research method is generally used in the beginning of a research process to improve the understanding and relevance of a specific research topic. According to Lubbe (2003), if the results derived from the findings, extending the boundaries of existing knowledge of the research area, the output could be seen as a valid form of research. Even though case studies are not often used in research projects, researchers have begun to implement them more regularly due to the intensity of the research on a relatively small number of cases.

According to Lubbe (2003), the case study methodology was designed to support or reject hypothetical proposals and not for the measurement of occurrences of events. Therefore, due to the diversity of the large number of SMMEs required to determine the effectiveness of their websites, case study research proved not to be feasible for this research.

II.   Statistical Studies

According to Wikipedia.com (2005), the term "Statistics" is defined as:

"…the science and practice of developing human knowledge through the use of empirical data, expressed in quantitative form…"

Statistical studies differ from case studies in the number and thoroughness of the cases studied. With statistical studies, the researcher ultimately examines a small set of variables within a large number of cases with the assistance of different statistical methods (Cooper & Schindler, 2003: 150).

Wikipedia.com (2005) lists the following steps for an analysis:

- Planning.
- Summarising.
- Interpretation of observations.
- Prediction or forecasting of future events based on a possible mathematical model of the system being observed.

Some authors believe that, due to the large number of cases in a statistical analysis, a small number of extreme cases are less likely to distort findings (Struwig & Stead, 2001: 9). Statistical studies can therefore in some cases produce a bigger overall picture than case studies could have produced. A negative aspect of statistical studies is the compromising of detailed insights, which case studies can provide.

4.6.1.3 Experimental research

Experimental research involves attempts to control and/or manipulate variables in the research study. Struwig and Stead (2001: 9) define experimental research as

"…the extent to which a set of independent variables influence other dependent variables…"

According to Cooper and Schindler (2003: 150):

"…experimentation provides the most powerful support possible for a hypothesis of causation…"

Lubbe (2003) supports this statement by stating:

"…it is important to bear in mind that the case study methodology is not designed to measure the frequency of occurrence of events, but rather to support or reject theoretical propositions…"

Several design methods exist for conducting experimental research, e.g. randomised post-test-only and pre-test-post-test control group design. The use of these design methods largely depends on the research question and the extent to which the researcher controls the variable. Due to the size and simplicity of the sample in this research project, experimental research was not selected for this research project.

4.6.2   Qualitative Approach

Qualitative refers to the meaning or definition of something being described, as apposed to the exact numerical measurement of that something, as in the quantitative approach. In layman's terms, qualitative refers to meaning of a research area, while quantitative assumes the meaning of the research area and refers to a measure of it (Cooper & Schindler, 2003: 152). Qualitative research is often used in studies that involve the management science, sociology, anthropology, social work, education, history, etc.

Several authors believe that the qualitative approach appears to serve as a methodology of verification rather that discovery (Eldabi *et al*., 2002: 65; Struwig & Stead, 2001: 7; Cooper & Schindler, 2003: 152).

To conduct qualitative research, a researcher could make use of a variety of approaches. Some of these approaches as identified by Cooper and Schindler (2003: 150) and Struwig and Stead (2001: 12-15) include:

- **In-depth interviewing:** Dialogue between two or more people where questions are usually asked in an unstructured manner to obtain information for assessment.

- **Participant observation:** Observing first what participants experience and taking note of anything the participants do. An example would be observing a shopkeeper communicating with a customer.

- **Films, photographs and videotape:** Analysis of films, photographs and videotape to obtain information.

- **Case Studies:** In-depth exploration of a few events or conditions for the purpose of attaining an understanding of the issues being investigated.

- **Elite or expert interviewing:** Very similar to an in-dept interview in that the sample will potentially consist of influential or well-informed people in an organization or community.

From the methods stated above, expert interviewing will be deployed for this research project. The elements found in the literature review will be compared with the results gathered from the statistical study. Findings will then be triangulated against information gathered from a personal interview with a representative of a leading South African search engine. Information regarding search engine algorithms will be verified from the search engine enterprise under consideration and then interpreted. Lubbe (2003) states that accuracy of data can be enhanced when there is a physical presence of an interviewer. A further method to improve the accuracy of the interview is to weigh, test and sift false statements and personal opinions from the questions before conducting the interview. Cooper and Schindler (2003: 325) state that:

> "…the greatest value of personal interviews lies in the depth of information and detail that can be secured…"

It is believed that far more information can be secured from personal interviews than from telephone and self-administered surveys (Struwig & Stead, 2001: 87; Cooper & Schindler, 2003: 325). The interviewer has the ability to interview the respondents in their natural work environment, in a

comfortable time frame (Wood, 1998). In certain circumstances, additional information can also be gleaned by the interviewer through observation, or by asking more questions.

Interviewing could however contain several disadvantages. Interviews can become a costly and time-consuming method for collecting data, depending on the size of the sample and the availability of the interviewees.

4.7   RESEARCH DESIGN AND METHODOLOGY

To conduct this study an experimental study with multiple qualitative and quantitative methods for collecting data will be conducted, with the unit of analysis being website visibility. Detailed steps as depicted in Figure 4.1 will be followed.

FIGURE 4.1: Model explaining the processes involved in conducting this research

Detailed descriptions of the subsections as depicted in the methodology of Figure 4.1 are supplied in Chapter 5 under Paragraphs 5.2, 5.3 and 5.4.

4.8   PHASE 1 – DESIGN OF CRITERIA QUESTIONS AS INSTRUMENT: CRITERIA 1

The purpose of this phase was to design and then test a search criteria list to evaluate websites ranking in the top results of major search engines. In so doing, the investigative question, "Which strategies are implemented by websites listed high in search engine results?" could be answered. The search criteria was called Criteria 1 (see Appendix A), and requested information from the evaluator on the basic design features regarding website visibility.

The search criteria were designed to contain the majority of the elements identified through the literature study. Questions included in the criteria list were set to determine which elements were used by the top four websites in the results, and which were returned from a keyword search. Each time a site made use of an element specified on the criteria list, its importance rating was increased. After the analysis process, the elements which appeared in the top websites were ranked according to their importance. The ranked list was then used to compile a model to indicate the most important factors to bear in mind when designing or updating a website.

The following aspects were considered during the selection phase of the criteria list:

- Only those elements, which would require the evaluator to go no further than the home page, should be included.
- Situations where the evaluator needed to evaluate the backend code of the homepage should be kept to a minimum.
- Questions, which require the evaluator to evaluate pages, which link from the homepage, should be removed.
- Precise and clear instructions on how to answer questions should be included.
- Only dichotomous (yes/no) questions should be used.

- Leading questions should be avoided.

The author identified a list of categories from which keywords could be derived. Keywords were compiled in context with SMMEs. The Cape Chamber of Commerce (referred to as CCofC) was used as source for the keywords. The CCofC serves the interests of business by satisfying the needs of its members through representing their views and providing services of value (Cape Chamber of Commerce, 2005a). All companies belonging to the CCofC were divided into 38 categories. These categories were manually extracted from the websites and placed into a spreadsheet document as indicated in Appendix E.

A pilot study was conducted to test the effectiveness and relativity of the questions listed in criteria 1 (see Appendix A). A total of four websites were analysed by analysing the websites listed first, after searching for one keyword within three International and one South African search engines.

4.9   PHASE 2 – EDITING AND TESTING OF CRITERIA QUESTIONS AS INSTRUMENT: CRITERIA 2

After conducting Phase 1 (see Paragraph 4.8), several issues became clear regarding the design of Criteria 1 (see Appendix A). Issues identified were then used to edit and improve the design of Criteria 1 (see Appendix A). The following problems emerged from Phase 1 (see Paragraph 4.8).

4.9.1   Clarity

Criteria 1 (see Appendix A) contained several terms, which could be termed unclear to the average respondent. Due to the high technical aspect of this research process, a definition list was added to assist participants in the analysis process. The definitions list, (see Appendix D) was compiled from the literature, which contained views of several authoritative references.

4.9.2   Layout

Some questions were found to be very unclear and difficult to interpret. An extra page was added to the questionnaire, which stated the detailed steps involved to successfully conduct the study. This page was named "Instructions to Analyst" (see Appendix F).

4.9.3   Questions

Some of the questions were found to be irrelevant to the research project and therefore removed from the set criteria.

4.9.4   Instructions

Instructions within questions made the questions difficult to read. Instructions where changed to bold font and positioned above the questions, to improve the readability of the questions.

These changes were then applied and tested (see Appendix B) by involving five learners in a training session. The five selected learners were in the process of completing their bachelor degrees in Information Technology at the time of the survey. A 45 minute meeting was set for each learner with the purpose of explaining the requirements, and to test the design of Criteria 2 (see Appendix B).

All businesses listed by the CCofC at the time of this study were divided into 38 categories. These categories were manually extracted from the category list on the CCofC website and placed onto a spreadsheet (see Appendix E) to use as keywords in the pilot study. Keywords 30 to 35 on the extracted category list (see Table 4.1 and Appendix E) were selected to test the design of Criteria 2 (see Appendix B).

TABLE 4.1: Keywords chosen to test criteria list

No	Keyword
30	Mining and Quarrying
31	Pharmaceuticals, Toiletries and Cosmetics
32	Printing, Packaging and Stationery
33	Property
34	Public Services Sector
35	Safety and Security

The 45 minute session with the learners was divided into three time segments as indicated in Table 4.2. The training session was used to explain the analysis process. Key terms, visibility elements and instructions were discussed in this session. The practical session was used to test the criteria list. In this session each learner was provided with one of the selected keywords and requested to conduct a search within three leading international (Google, Yahoo, MSN) and one national (Ananzi) search engines by using the keyword. The website listed first in each search engine was then selected for analysis. During the practical sessions conducted with the five learners, 20 websites were analysed and results from criteria were recorded onto a spreadsheet (see Appendix G). The final session was used to gather feedback (verbally) regarding the design and layout of Criteria 2 (see Appendix B).

TABLE 4.2: Time allocated for training session

Task	Time
Training session	15 minutes
Practical session	25 minutes
Feedback	5 minutes
Total time	**45 minutes**

4.10　　PHASE 3 – FINAL CHANGES TO CRITERIA QUESTIONS AS INSTRUMENT: CRITERIA 3

The third phase was used to implement the final changes identified through Phase 2 (see Paragraph 4.9). Sections affected on the criteria list include:

4.10.1 Domain Names (see Appendix C, Section 1)

In this section, two questions (1.1 and 1.2) were removed (see Criteria 2 in Appendix B) as it was discovered that these questions would be of little value to the research process. The two remaining questions were rephrased and examples were added in bold to provide a better understanding.

4.10.2 HTML Naming Conventions (see Appendix C, Section 2)

An extra question was added to this section to provide more information on the given element.

4.10.3 Hypertext / Anchor Text (see Appendix C, Section 4)

An example was added to question 4.1 on Criteria 3 (Appendix C) to provide a better understanding of what was required. The bold instruction phrase was also altered slightly to avoid confusion.

4.10.4 Meta-tags (see Appendix C, Section 6)

4.10.4.1 Title tag (see Appendix C, Section 6.2)

A question was added (see 6.2.5 on Criteria 3 in Appendix C), and another was removed (see 6.2.4 on Criteria 2 in Appendix B) in this section. Instructions were made bold and questions were rephrased to provide a better understanding of what is required from the participant.

4.10.4.2 Meta-description tag (see Appendix C, Section 6.3), Meta-keyword tag (see Appendix C, Section 6.4), Meta-header tag (see Appendix C, Section 6.5) and Alt tags (see Appendix C, Section 6.7)

In these sections instructions were made bold and questions were rephrased to provide a better understanding of what is required from the respondents.

4.10.5 Macromedia Flash (see Appendix C, Section 8)

Some of the questions listed in this section were difficult to determine, as there are different levels of occurrence of the element in question. One question (see 8.1 on Criteria 3 in Appendix C) was then altered to prevent occurrence of confusion.

4.10.6 Banner Advertising (see Appendix C, Section 10), Link Popularity (see Appendix C, Section 11) and JavaScript (see Appendix C, Section 12)

In these sections some spelling mistakes were identified and corrected. Instructions were made bold and questions were rephrased to provide a better understanding of what is required from the respondent.

4.10.7 Referencing

On the top left corner of the criteria list, three words were added, requiring the participant to specify the keyword used, the ranking of the website analysed and the search engine used. These were added to simplify the analysis process in case of possible errors occurring.

4.11 CLOSURE

In this chapter several measuring methods and approaches were identified. A criteria list (see Criteria 1 in Appendix A) was developed through research identified in Chapter 2 and 3. The criteria list was then tested, improved (see Criteria 2 in Appendix C) and then re-tested. Final changes were then applied to the criteria (see Criteria 3 in Appendix C). Criteria 3 (see Appendix C) was now deemed ready to be used in the analysis process.

CHAPTER 5

RESEARCH RESULTS AND ANALYSIS

5.1 INTRODUCTION

The aim of this chapter in the first instance is to determine, with the assistance of software, where SMMEs rank in the major search engines (see Phase 4 in Paragraph 5.2). In the second instance, to identify which visibility elements (see Phase 5 in Paragraph 5.3) websites in the top positions make use of, and also how they make use of these visibility elements to achieve top results. In the third instance, to report on an interview with a representative of a leading South African search engine to verify the importance of elements identified in Phase 5 (see Paragraph 5.3).

5.2 PHASE 4 – PROOF THAT SMMES REQUIRE ASSISTANCE

The purpose of this phase is to prove that SMMEs are in need of assistance in utilising the full potential of a visible website. The first step was to identify a list of SMMEs which operate within the Western Cape. This list was drawn from the CCofC website (Cape Chamber of Commerce, 2005b).

The CCofC is considered a valuable and reliable source to obtain a list of SMMEs from, for the following reasons:

- The CCofC has a mission to satisfy and serve the needs of its members, through representing their views and providing services of value.
- For an SMME to be listed under the CCofC, it is liable for an annual subscription fee.
- SMME details are verified for updates annually at renewal of the subscription.
- The CCofC provides over 50 value added services to members.
- The CCofC has more than 4000 SMMEs listed.

The CCofC allows any user to search through its database by providing the user with multiple options to refine the search. These options include company name, area, category, business classification, keywords, number of employees and also an option to specify whether the company is an exporter, importer, neither or both. Options were selected as indicated in Figure 5.1, with 3,040 results being returned.

FIGURE 5.1: CCofC members' selection page (Cape Chamber of Commerce, 2005)

Judgement sampling was used to remove all companies without active websites and those which employ more than 200 people, therefore eliminating those companies without websites and reducing companies to only SMMEs. Due to the size of the remaining SMME sample, the SMMEs were then further randomly reduced to a sample of 300 SMMEs with active websites. The recorded detail of the 300 SMMEs which were used to

determine the visibility of SMMEs in the Western Cape, is reflected in Appendix H.

5.2.1 Analysis

The ranking position of the 300 website samples within six search engines, were analysed by making use of keyword verification software provided by the company, Marketleap (Marketleap, 2005). This verification tool was used to determine if a site appears in the top three pages (or 30 results) of a search result for a specific keyword. In all instances a maximum of five keywords were extracted from the company description provided on the CCofC website, to determine the ranking of the site (see Figure 5.2) in six (AOL, Google, Lycos, MSN, Netscape and Yahoo) of the top 20 search engines.

An example of the results of the keyword verification software is shown in Figure 5.2 and Figure 5.3. Results returned from Hotbot were removed from the analysis, due to its partnership with Lycos at the time of this research. The remainder of all the results returned was summarised in a spreadsheet (see Appendix H).

FIGURE 5.2: Screenshot of Marketleap keyword verification software

FIGURE 5.3: Screenshot of Marketleap results page

91

5.2.2   Results

The ranking position of the 300 websites analysed, is listed in Table 5.1 and Graph 5.1. The results returned that only 6.3% of the SMMEs analysed, appeared in any of the top 30 results of the six search engines (AOL, Google, Lycos, MSN, Netscape and Yahoo), when searching for keywords extracted from the SMMEs' descriptions. The results further returned that 0.7% of SMMEs analysed were listed in the top 30 results of all six search engines. The remaining 5.7% of SMMEs were listed in some of the search engines (see Appendix H) within the top 30 results.

TABLE 5.1: Webpage ranking statistics

Webpage Ranking Statistics	Quantity	Percentage
Total SMMEs	300	100.00%
SMMEs Top 30 in all six search engines	2	0.70%
SMMEs not in any search engine	281	93.70%
SMMEs within Top 30 in some SE	17	5.70%

GRAPH 5.1: Website ranking within AOL, Google, Lycos, MSN, Netscape and Yahoo

The results returned clearly indicated that there is a definite need for assistance to enhance website visibility for SMME websites and as thus, they provide an answer to the investigative question which reads as follows: *To what extent do SMMEs require assistance with search engine visibility?*

5.3 PHASE 5 – ANALYSIS OF WEBSITES RANKING IN TOP POSITIONS

In this phase, 320 websites were analysed to identify the elements used or avoided by website owners to achieve top positions. This study was conducted by making use of:

- Twenty-three keywords from the category list extracted from the CCofC website (see Appendix E).
- Six learners, each in the process of completing their bachelor degree in Information Technology. Five of these students were used in the training session discussed in Paragraph 4.9.
- Criteria 3 (see Appendix C) as formulated in Chapter 4.

The five learners (trained during the pilot run), were each provided with a template which contained a spreadsheet with four keywords and four open columns (see Figure 5.4), to record the URLs of webpages analysed.

No	Keyword	Rank	Google	Yahoo	MSN	Ananzi
1	Agriculture and Fishing	1				
	Agriculture and Fishing	2				
	Agriculture and Fishing	3				
	Agriculture and Fishing	4				
2	Airconditioning, Refrigeration and Cold Storage	1				
	Airconditioning, Refrigeration and Cold Storage	2				
	Airconditioning, Refrigeration and Cold Storage	3				
	Airconditioning, Refrigeration and Cold Storage	4				
3	Automotive	1				
	Automotive	2				
	Automotive	3				
	Automotive	4				
4	Banking Sector	1				
	Banking Sector	2				
	Banking Sector	3				
	Banking Sector	4				

FIGURE 5.4: Screenshot of software package given to first five learners

The sixth untrained learner was provided with a separate package as listed in Appendix I. This package was used to glean all information recorded by the five trained learners on Criteria 3 (Appendix C) on the day of the analysis. The layout contained the following elements:

- The first column was used to count the number of websites analysed.
- The second and third columns represented the keyword numbers, as listed in Appendix E.
- The fourth column represented the search engine used to conduct the search.
- The fifth column represented the headings on the criteria list.

The whole analysis process was set to take place within a set timeframe. Websites not analysed in that timeframe were eliminated from the study. The five trained learners were scheduled for the analysis process while the sixth learner was scheduled to record the results on a spreadsheet.

In the three hour session, four of the trained learners completed queries for only two keywords. The fifth learner completed queries for one keyword. As a result, of the 320 expected outcomes, only 144 were achieved in the allocated timeframe. Results recorded by the sixth learner were then further refined into separate sections containing the questions and the average result as listed in the in paragraphs to follow.

5.3.1 Results (Domain Names)

Domain Names (see Appendix C, Section 1) from the criteria list were used to determine how the domain names of top webpages relate to the content of that page and how it was structured. Results from Section 1 in the study are depicted in Table 5.2 and Graph 5.2.

TABLE 5.2: Results from Domain Names (Section 1) on criteria list

1. Domain Names	%
Domain names related to products, services or company name	68.8%
Domain names which are short (2 or fewer keywords)	83.3%
Domain names which are easy to pronounce and spell	79.2%
Domain names which are descriptive	65.3%
Domain names which are memorable	68.8%

GRAPH 5.2: Results from Domain Names (Section 1) on criteria list

5.3.2   Analysis (Domain Names)

In the online marketplace, the domain name plays an ever increasing role in the visibility of websites. A well chosen domain name assigned to a website can have a major influence on how users find and view the site. Furthermore, a well branded domain name can also have an effect on how staff, suppliers, business partners, the trade, regulators and providers of capital view a company.

Authors such as Callan (2004), Clark, Chou and Yen (2001: 225), Galon (1999: 46), Gorman (2000: 161) and Rowley (2004: 132) claim that website authors should try to keep domain names:

- Short and simple.
- Meaningful, by making use of keywords.
- Relating to the business.
- Easy to remember.

Results of the analysed websites (see Table 5.2 and Graph 5.2) map to the elements in the above list. The majority (68.8%) of websites analysed, had a domain name containing words related to the company, products, services or website. Certain domain names were short (83.3%), pronounceable (79.2%), descriptive (65.3%) and memorable (68.8%). These results are strongly supported in literature as demonstrated in Paragraph 3.3.2.

5.3.3 Results (HTML Naming Conventions)

HTML naming conventions (see Appendix C, Section 2) from the criteria list were used to determine how the HTML naming conventions of top webpages related to the content of that page and also how it was structured. Results from Section 2 in the study are depicted in Table 5.3 and Graph 5.3.

TABLE 5.3: Results from HTML naming conventions (Section 2) on criteria list

2. HTML Naming Conventions	%
Naming Conventions which are short (2 or fewer keywords)	67.4%
Naming Conventions which are easy to pronounce and spell	70.1%
Naming Conventions which are descriptive	68.8%
Naming Conventions which are memorable	54.9%
Naming Conventions which are related to products, services or company name	63.9%
More than two keywords which contain hyphen and/or an underscore	45.4%

GRAPH 5.3: Results from HTML naming conventions (Section 2) on criteria list

5.3.4   Analysis (HTML Naming Conventions)

HTML naming conventions in Section 2 (see Appendix C, Section 2) were aiming to determine how websites listed in the top positions saved their filenames which form the page names in the URL, e.g. www.sa-cycling.com/contact-us.html. According to Galon (1999: 56), search engines more often than not first look for keywords in a URL when indexing a page. It is therefore of importance to ensure that the correct keywords are used throughout the URL. Galon (1999: 56) further states that designers:

> "…are voluntarily giving up a good spot that could be used to advantage for an important and useful keyword…"

Nobles and O'Neil (2000: 4) also stated that by placing keywords in the URL, a developer could experience added relevancy to those words in certain search engines.

Against the above background, it is recommended that page names should comply with the rules provided in Paragraph 5.3.2 which require keywords to be:

- Short and simple.
- Meaningful.
- Relating to the business.
- Easy to remember.

Results of the analysed websites (see Table 5.3 and Graph 5.3) map to the elements in the above list. The majority (63.9%) of websites analysed had file names containing words related to the company, products, services or website. Furthermore, those containing file names representing more than one keyword (45.4%), made use of an underscore and/or hyphens to separate the words. By making use of these two parameters, the web developer makes it easier for users to remember the name of the website.

Furthermore, by making use of a hyphen to connect words, the developer allows the search engine to view a file name as distinct keywords, e.g. "apple-tree.html". This is seen by the search engine as "apple" + "tree" where as with "apple_tree.html", the search engine reads "appletree" (Galon, 1999).

Of the file names analysed, the majority were short (67.4%), pronounceable (70.1%), descriptive (68.8%) and memorable (54.9%). These results are strongly supported in literature as demonstrated in Paragraph 3.3.5.

5.3.5   Results (Frames)

Frames (see Appendix C, Section 3) from the criteria list were used to determine how often top webpages make use of frames to structure their content. Results from Section 3 in the study are depicted in Table 5.4 and Graph 5.4.

TABLE 5.4: Results from Frames (Section 3) on criteria list

3. Frames	%
Webpages containing visible frames	5.6%
Webpages making use of <noframes> tag	0.0%

GRAPH 5.4: Results from Frames (Section 3) on criteria list

5.3.6   Analysis (Frames)

Nobles and O'Neil (2000: 180) emphasised that:

- Very few webpages ranking in the top results make use of frames.
- Only extremely complex websites need to use frames.
- Frames increase the file size and the total number of words that make up a website.

Galon (1999: 71) stated:

"I wouldn't recommend using frames on the entrance page…"

These statements were all supported by the results (see Table 5.4 and Graph 5.4) produced from the analysis process on Frames from the criteria list (see Appendix C, Section 3). Of all the websites analysed, only 5.6% contained frames of which none made use of the <noframes> meta-tag. These results confirm the notion that the use of frames should be avoided if possible.

5.3.7   Results (Hypertext/Anchor Text)

Hypertext / Anchor Text (see Appendix C, Section 4) from the criteria list, was used to determine to what extent top webpages make use of hypertext (anchor text). Furthermore, it was used to determine how hypertext relates to the content of that page and how it was structured. Results from Section 4 in the study are depicted in Table 5.5 and Graph 5.5.

TABLE 5.5: Results from Hypertext / Anchor Text (Section 4) on criteria list

4. Hypertext / Anchor Text	%
Webpages containing hyperlinks	100.0%
Hypertext which accurately describes the products and services	91.0%

GRAPH 5.5: Results from Hypertext / Anchor Text (Section 4) on criteria list

5.3.8 Analysis (Hypertext/Anchor Text)

The use of hyperlinks is viewed by several authors as an effective and efficient way to improve the visibility of a site, due to its linking capability (Henzinger *et al.*, 2002: 5; Rowley, 2001: 208; Weideman & Haig-Smith, 2002: 285). Some search engines also rank keywords appearing in hypertext higher due to the resemblance between the hypertext description and the hyperlink (Henzinger *et al.*, 2002: 5). As stated earlier, spammers often turn to data manipulation to increase their rankings. With hypertext, the text description represents the destination of a link and could therefore affect the perception of the user if the link and text carry no resemblance. Therefore, this resemblance makes it difficult for spammers to manipulate search engine algorithms by making use of inadequate hypertext links.

Statistical results confirm the importance of hypertext. Results (see Table 5.5 and Graph 5.5) show that 100% of the websites analysed made use of hyperlinks. Of these hyperlinks, 91% had an accurate resemblance between the content of the current page and the page they link to.

5.3.9   Results (HTML)

HTML (see Appendix C, Section 5) from the criteria list was used to determine to what extent top webpages make use of keyword rich sentences, which relate to products, services and company details in their headings. Results from Section 5 in the study are depicted in Table 5.6.

TABLE 5.6: Results from HTML (Section 5) on criteria list

5. HTML	%
Headings accurately describing the products and services	79.9%

5.3.10   Analysis (HTML)

Search engines often rank keywords listed in headers higher than the keywords listed in the body text (Kritzinger & Weideman 2005; Nobles & O'Neil 2000: 43). This can be attributed to search engines often seeing bold text as the beginning of a paragraph. The bold text is also more visible to the user and therefore could minimise the risk of indexing false keywords.

The results returned (see Table 5.6) strongly supported the use of information-rich headings. A total 79.9% of the websites analysed, contained headings with keyword-rich sentences, which pertained to the products, services and company details of the website.

5.3.11   Results (Meta-tags)

Meta-tags (see Appendix C, Section 6) required the learners to analyse the source code of each website. Full instructions to conduct the analysis on the source code were provided in an "Instructions to Analyst" page, (see Appendix F). The main focus of meta-tags is to provide optional or additional information about a webpage and its content, quality, condition and other characteristics. Therefore, the majority of meta-tags would adhere to similar rules. Some of these rules which are recommended by Craven (2003), Nobles and O'Neil (2000: 63), Gorman (2000: 235) and Galon (1999: 57) are listed in Table 5.7.

TABLE 5.7: Recommendations for Meta-tags

Meta-tag Recommendations	
1	Keywords should solve a problem
2	Keywords closer to the front will carry more weight
3	Keywords should be appealing, captivating, compelling or eye-catching
4	Keywords should be easy to read and understand
5	Refrain from using stop words
6	Keywords should be meaningful
7	Keywords should be in lower case

5.3.11.1 Results (Dublin Core)

Dublin Core (see Appendix C, Section 6.1) from the criteria list was used to determine to what extent top webpages make use of Dublin Core to describe several aspects of their site. Results from Section 6.1 in the study are depicted in Table 5.8.

TABLE 5.8: Results from Dublin Core (Section 6.1) on criteria list

6.1 Dublin Core	%
Webpages making use of Dublin Core	5.60%

5.3.11.2 Analysis (Dublin Core)

Even though Dublin core was originally developed to provide additional information about a website to the user, it never really intended to become part of the visibility elements required to provide a visible website (Weibel *et al*., 1995).

The results (see Table 5.8) indicated that only a small portion (5.6%) of websites ranking in the top positions still make use of this type of meta-tag. The Dublin Core tags could provide content rich information about the site, but according to the conducted literature survey, they carry very little value in the indexing process.

5.3.11.3 Results (Title tag)

The title tag (see Appendix C, Section 6.2) from the criteria list was used to determine to what extent top webpages make use of title tags, how well these tags relate to the content of the site, the structure of the title, as well as

the length of the title. Results from Section 6.2 in the study are depicted in Table 5.9 and Graph 5.6.

TABLE 5.9: Results from Title tag (Section 6.2) on criteria list

6.2 Meta-title tag	%
Webpages making use of the Title tag	99.3%
Keywords in title which relate to products, services or company name	94.4%
Titles containing unnecessary punctuation marks or filter words	7.7%
Titles written in capital letters	3.5%
Titles which are commonly understood	85.3%
Titles longer than ten words	26.6%

GRAPH 5.6: Results from Title tag (Section 6.2) on criteria list

5.3.11.4 Analysis (Title tag)

The tile tag has been proven by Craven (2003) to carry the highest weight of all meta-tags when it comes to webpage indexing. According to Galon (1999: 57), search engines initially look at the title to find important keywords. The title tag appears in several areas for the user to see and can therefore improve both the visibility of a website, and its usability. Galon (1999: 57) also advises developers not to exceed the use of ten keywords in the title tag.

Results returned from the statistical analysis (see Table 5.9 and Graph 5.6) strongly proved the value of the recommendations listed in Table 5.7. A total of 99.3% of top webpages analysed made use of a title tag. Of these webpages, 94.4% provided keywords related to products, services or

company name, while 85.3% of keywords were meaningful and understood by the evaluators. Furthermore, the majority (92.3%) of webpages made accurate use of the title tag, by providing a title without stop words, fewer than ten words (73.4%) and keywords that were written in lowercase (96.5%). These results prove the importance of having a well structured and carefully chosen title tag.

5.3.11.5 Results (Meta-description tag)

The Meta-description tag (see Appendix C, Section 6.3) from the criteria list was used to determine to what extent top webpages make use of description tags, how well these tags relate to the content of the site, as well as the structure of the tag. Results from Section 6.3 in the study are depicted in Table 5.10 and Graph 5.7.

TABLE 5.10: Results from Meta-description tag (Section 6.3) on criteria list

6.3 Meta-description tag	%
Webpages making use of the Meta-description tag	63.9%
Webpages where title is repeated in the Meta-description tag	35.9%
Webpages where words in the description tag are commonly understood	79.4%
Webpages where description tag is used correctly	85.9%
Webpages where description is related to products, services or company name	95.7%

GRAPH 5.7: Results from Meta-description tag (Section 6.3) on criteria list

5.3.11.6 Analysis (Meta-description tag)

The description tag was considered an effective way for search engines to extract information-rich keywords. However, due to the excessive use of spamming techniques to manipulate search engine algorithms, very little weight is currently assigned to these descriptions (Henzinger *et al.*, 2002: 3). Nevertheless, authors in the likes of Nobles and O'Neil (2000: 74) and Craven (2003) still recommend the inclusion of this tag to provide content-rich information to the minority of search engines which still support it.

Statistical results (see Table 5.10 and Graph 5.7) from the analysis process reflected a strong presence (63.9%) of the description tag. Recommendations listed in Table 5.7 were also strongly supported. Of the analysed websites containing the description tag, 95.7% provided keywords related to products, services or company name, while 79.4% were meaningful and understood by the evaluators. There was also a small percentage (14.1%) where description tags were used incorrectly. Examples include cases where several keywords were repeated, or where the description was left blank. There were also a number of instances (35.9%) where the title tag was repeated in the description tag.

5.3.11.7 Results (Meta-keyword tag)

The Meta-keyword tag (see Appendix C, Section 6.4) from the criteria list is very similar to the Meta-description tag (see Paragraph 5.3.11.5 and Paragraph 5.3.11.6). Both the description tag and the keyword tag provide keywords related to the content of the webpage. Section 6.4 (see Appendix C) was used to determine to what extent top webpages make use of keyword tags and how well these tags relate to the content of the site. Results from Section 6.4 in the study are depicted in Table 5.11 and Graph 5.8.

TABLE 5.11: Results from Meta-keyword tag (Section 6.4) on criteria list

6.4 Meta-keyword tag	%
Webpages making use of the Meta-keyword tag	63.2%
Webpages where keyword tag description is commonly understood	78.0%
Webpages where keyword tag has words related to products, services or company name	96.7%

GRAPH 5.8: Results from Meta-keyword tag (Section 6.4) on criteria list

5.3.11.8 Analysis (Meta-keyword tag)

The keyword meta-tag is similar to the description tag. The keyword tag contains keywords related to the content of the webpage, whereas the description tag contains a small description of the contents of the webpage. Historically, it was an effective source for search engines to extract information rich keywords. The keyword tag has dropped in value due to excessive use of spamming techniques (Henzinger *et al.*, 2002: 3). Craven (2003) however, still recommends the inclusion of this tag to provide content rich information to the minority of search engines which still support it.

Statistical results (see Table 5.11 and Graph 5.8) from the analysis process returned a strong presence (63.2%) of the keyword tag. Keyword recommendations made in Table 5.7 were also strongly supported. Of the analysed websites containing the keyword tag, 96.7% included keywords related to products, services or company name, while 78% were meaningful and understood by the evaluators.

5.3.11.9 Results (Meta-header tag)

The Meta-header tag (see Appendix C, Section 6.5) from the criteria list was used to determine to what extent top webpages make use of header tags and how well these tags relate to the content of the site. Results from Section 6.5 in the study are depicted in Table 5.12 and Graph 5.9.

Results from Meta-header tag (Section 6.5) on criteria list

6.5 Meta-header tag	%
Webpages making use of any one of the six header tags (H1- H6)	23.6%
Webpages where words in the header tag description are commonly understood	91.2%
Webpages where header tag contains keywords related to products, services or company name	88.2%

GRAPH 5.9: Results from Meta-header tag (Section 6.5) on criteria list

5.3.11.10 Analysis (Meta-header tag)

The header tag is used to emphasise the importance of a phrase, by setting the size of the font. Research by Craven (2003) showed that the H1 (heading 1) and H2 (heading 2) tags, are the second and third most highly weighted tags after the title tag. Nobles and O'Neil (2000: 112) also confirmed that header tags were considered a very effective method to implement and to emphasise website relevancy. As with all the other meta-tags, this tag should also adhere to the rules recommended in Table 5.7.

Results (see Table 5.12 and Graph 5.9) returned that only a few webpages (23.6%) made use of the header tag. Of the pages which made use of the header tag, 88.2% had header descriptions which included keywords related to products, services or company name, while 91.2% were meaningful and understood by the evaluators.

5.3.11.11 Results (Meta-robot tag)

Meta-robot tag (see Appendix C, Section 6.6) from the criteria list was used to determine to what extent top webpages made use of robot tags. Results from Section 6.6 in the study are depicted in Table 5.13.

TABLE 5.13: Results from Meta-robot tag (Section 6.6) on criteria list

6.6 Meta-robot tag	%
Webpages containing any Meta-robot tags	43.8%

5.3.11.12 Analysis (Meta-robot tag)

Robot tags are used to indicate to crawlers, which areas within a site to scan and which not to scan. These tags carry very little weight in the indexing process of websites, but they could however dramatically reduce the time a crawler spends analysing a site. By including this tag within the code of a website, the developer ensures that only the most relevant and content rich information is analysed by the search engine crawlers (Nobles & O'Neil, 2000: 212-214; Craven, 2003).

Robot tags usually manifest in one of two ways. One being the inclusion of a robot meta-tag within the meta-tag section, and the other the use of robot tags within a separate file (called robots.txt). Both these methods were tested. The results (see Table 5.13) returned that a total of 43.8% of websites analysed, make use of robot tags to redirect search engine crawlers.

5.3.11.13 Results (Alt tag)

Alt tags (see Appendix C, Section 6.7) from the criteria list were used to determine to what extent top webpages make use of alt tags and how well these tags relate to the content of the site. Results from Section 6.7 in the study are depicted in Table 5.14 and Graph 5.10.

TABLE 5.14: Results from Alt tags (Section 6.7) on criteria list

6.7 Alt tag	%
Websites making use of Alt tags	69.4%
Webpages where words in the alt tag description tags are commonly understood	77.0%
Webpages where alt tags contain keywords related to products, services or company name	68.0%

108

GRAPH 5.10: Results from Alt tags (Section 6.7) on criteria list

5.3.11.14 Analysis (Alt tag)

The main purpose of the alt tag is to provide information about an image or graphic when it is invisible to the user. Such information is aimed at, for example, assisting visual disabled users (screen output readers), users with graphics turned off, etc. The use of well structured alt tags to increase the visibility of a site is well supported in literature.

Konia (2002: 191) stated:

"…certain engines look for keywords in the alt tags…"

Nobles and O'Neil (2000: 45) stated:

"…this can give this keyword a boost in relevancy with many of the search engines…"

Thurow (2003: 82) stated:

"Some search engines index alternative text, thus making your graphic images another place to strategically place keywords."

Statements made by the above authors were supported by the research results (see Table 5.14 and Graph 5.10), which showed that 69.4% of all

sites analysed, made use of alt tags. Of these pages, 68.0% contained descriptions which included keywords related to products, services or company name, while 77.0% of webpages contained keywords that were meaningful and which were understood by the evaluators.

5.3.12   Results (Spamming)

Spamming (See Appendix C, Section 7) from the criteria list was used to determine to what extent top webpages make use of spamming techniques, to manipulate search engines to assign them a higher rating. Results from Section 7 in the study are depicted in Table 5.15.

TABLE 5.15: Results from Spamming (Section 7) on criteria list

7. Spamming	%
Webpages containing visible form of link spamming	3.5%

5.3.13   Analysis (Spamming)

Spamming has long been viewed as an effective way to manipulate search engines to achieve higher rankings (Henzinger *et al.*, 2002: 2). Several spamming techniques exist, but the majority are difficult to detect. As a result, only one spamming technique namely "link spamming" was searched for within the websites analysed. Link spamming is relatively easy to identify on a website, as it usually comprises a collection of different links pointing to other sites with the sole purpose of raising its link popularity. Several authors (Notess, 1999: 86; Thurow, 2003: 225; Henzinger *et al.*, 2002: 4; Van Steenderen, 2001) advise against the use of spam, since the majority of search engines have begun to implement safety measures which avoid, and in some cases blacklist sites containing spam.

Results from statistical analysis (see Table 5.15) returned that only 3.5% of websites analysed, which listed in top results, contained possible signs of link spamming features. This result strengthens the notion that spamming should be avoided at all times.

5.3.14   Results (Flash)

Flash (see Appendix C, Section 8) from the criteria list was used to determine to what extent top webpages make use of Flash, as an alternative to graphics and/or JavaScript. Websites containing Flash were also analysed to determine if the pages that contained large sections of Flash, provided the user with an option to load the site without the use of Flash. Furthermore, sites were also analysed to determine if they contained Flash within their navigational structure. Results from Section 8 in the study are depicted in Table 5.16.

TABLE 5.16: Results from Flash (Section 8) on criteria list

8. Flash	%
Webpages consisting of more than 50% Flash images	0.0%
Webpages with an option to load a duplicate of the page without flash	0.0%
Webpages containing Flash in their navigational structure	0.0%

5.3.15   Analysis (Flash)

Currently Flash pages cannot be extensively crawled or indexed by the majority of search engines as discussed in Paragraph 3.3.8.2. Some search engines do however follow embedded links within the Flash file format and some techniques also exist to build a visible website containing Flash. Several authors however recommend excluding Flash if possible and rather improve the use of hypertext and links as these elements are currently weighted higher (Goh & Wang 2004: 144; Thurow, 2003: 145-150; Gerhart 2001; Vorster, 2005).

Survey results (see Table 5.16) strongly supported the fact that Flash should currently be kept to a minimum. Results indicate that all of the sites analysed do not make use of Flash which take up more than 50% of the webpage's content. Of those webpages which had a small amount of Flash on their page, none made use of Flash in its navigational structure.

5.3.16   Results (Banner Advertising)

Banner Advertising (see Appendix C, Section 9) from the criteria list was used to determine to what extent top webpages make use of banner advertising methods and how well these advertising methods relate to the content of the site. Results from Section 9 in the study are depicted in Table 5.17 and Graph 5.11.

TABLE 5.17: Results from Banner Advertising (Section 9) on criteria list

9. Banner advertising	%
Webpages containing banner advertising	26.4%
Webpages with banners reflecting content of the page	76.3%

GRAPH 5.11: Results from Banner Advertising (Section 9) on criteria list

5.3.17   Analysis (Banner Advertising)

Banner advertising was believed to be the most popular method of online advertising (Gorman, 2000: 73). With this advertising medium, a developer could exchange banner ads with other sites, with the purpose of linking them to the developer's site. This whole process could be facilitated within minutes, while in the instance of other mediums advertisement could extend over a number of days. With banner ads, developers have the power to edit or replace an advertisement should it be required. Banner advertising could improve the link popularity of a site and therefore also the visibility (Henzinger *et al.*, 2002: 6). A developer should take care not to abuse this

advertising medium, as the excessive use of banner ads could be seen by a search engine as a potential spamming technique (Thurow, 2003: 225).

Results (see Table 5.17 and Graph 5.11) from the statistical analysis process indicated that developers were divided on the use of banner ads. A total of 26.4% of the websites listed in top results, make use of the banner advertising medium. Of those websites which made use of banner ads, 76.3% contained banner ads which reflected the content of the website.

5.3.18   Results (Link Popularity)

Link Popularity (see Appendix C, Section 10) from the criteria list was used to determine to what extent top webpages make use of links pointing to webpages within a website, or links pointing to other pages. The links were then analysed to determine how well these links relate to the content of the site. Results from Section 10 in the study are depicted in Table 5.18 and Graph 5.12.

TABLE 5.18: Results from Link Popularity (Section 10) on criteria list

10. Link popularity	%
Webpages containing links to other areas of the site	100.0%
Webpages where links are in context with the content of the current page	84.0%

GRAPH 5.12: Results from Link Popularity (Section 10) on criteria list

113

5.3.19 Analysis (Link Popularity)

Search engines see links to other sites or pages as valuable elements in their algorithm techniques, as discussed in Paragraph 3.3.11. Search engines typically assume that when a site provides a link to another site or page, it is because the author believes that the other site or page has relevant and information-rich content (Henzinger *et al.*, 2002: 6). Some search engines would follow a link and then compare the content of both sites to determine if the sites reflect similar data. Should these sites contain thematic content, the site being linked to would be assigned a greater weight than a link from an unrelated site (Konia, 2002: 295).

Due to the complexity of testing a site's link popularity, the sites were only checked for both links leading to other areas (webpages or websites), and if the descriptions of the links reflected the content of the link site accurately. Survey results (see Table 5.18 and Graph 5.12) returned that all of the top websites analysed contained links pointing to webpages within the website, or links pointing to other pages. Furthermore, 84.0% of the websites which contained links, had links which related to the content of the site analysed.

5.3.20 Results (JavaScript)

JavaScript (see Appendix C, Section 11) from the criteria list was used to determine to what extent top webpages make use of JavaScript to automate certain application tasks. The webpages were also scanned for the use of JavaScript files and the <noscript > tag. Results from Section 11 in the study are depicted in Table 5.19 and Graph 5.13.

TABLE 5.19: Results from JavaScript (Section 11) on criteria list

11. JavaScript	%
Web pages making use of the JavaScript	71.5%
Webpages with JavaScript code enclosed in a '.js' file	65.1%
Webpages making use of the <noscript> tag	28.2%

GRAPH 5.13: Results from JavaScript (Section 11) on criteria list

5.3.21   Analysis (JavaScript)

JavaScript is a popular programming language used to add animation and other functionality to webpages (Thurow, 2003: 123). It holds multiple benefits as an interactive tool with users. The problem with this technology is that it makes it difficult for the search engines to analyze a webpage containing JavaScript programming code. Nobles and O'Neil (2000: 184) recommend that script be placed in a separate file, as opposed to in a file where the webpage code resides. In so doing, the developer could provide a keyword-rich page with added advanced functionality. A further method to increase a webpage's visibility when using JavaScript, is to make use of the <noscript> and </noscript> tags, which are placed between the body tags. These tags allow the developer to include alternative keyword-rich content. This tag in addition provides for alternative content for the user, who prefers to search the web with JavaScript disabled (Thurow, 2003: 131).

Results from the statistical analysis process (see Table 5.19 and Graph 5.13) indicated that 71.5% of websites ranking in the top positions made use of JavaScript. Of these websites, the majority (65.1%) enclose part or all of their JavaScript code in a separate file. A further 28.2% of the analysed websites listed in to top positions make use of the <noscript> tag. These results proved that a website could achieve high listings with the use of JavaScript.

115

5.4 PHASE 6 – PERSONAL STANDARDIZED INTERVIEW WITH A SOUTH AFRICAN SEARCH ENGINE REPRESENTATIVE

5.4.1 Introduction

Phase 6 involved an interview with Vorster (2005), a representative of a leading South African search engine (Ananzi). Due to the fact that personal interviews are the most versatile and flexible of the three major questionnaire methods (Struwig & Stead, 2001: 86), a questionnaire was designed which contained several unstructured questions (see Appendix J) based on the headings of the criteria list (see Appendix C).

Both the interviewer and the interviewee were given the opportunity to provide further explanations or clarifications, allowing for multiple responses to each question. Furthermore, to ensure that the interviewee understands the objective and process of the interview, the information being targeted by the investigative questions was explained.

5.4.2 Interview Results

5.4.2.1 Domain Names

According to Vorster (2005), domain names used to be extremely important. This however has changed quite significantly over the last eight - ten months, although Google still uses domain names as important criteria in its ranking algorithm.

Vorster (2005) (verbatim as per the interview script) provided the following example to explain the reason for the drop in domain importance by stating:

> "So you'll get sites that call themselves www.free-mp3-online.com. So if you type in free online mp3's and hit search, the site would appear in the top results. It used to be extremely powerful, but it changed quite significantly. The best way to check it, when I was in the gambling industry, with the domain name gambling.com, you type in gambling

and it always came on tops. The casino had this guy who used to make millions, tens of millions of dollars out of mining. He drove a lot of gambling traffic through to his site and gambling traffic is very valuable traffic in terms of online traffic. It always used to be no1 and it started dropping 20, 30, 40. I think it is still somewhere in the top ten of most search engines. Other guys start outwitting him. The domain used to be very important, but is not as important any more, purely because - If you own a company that sells wine, the domain name wine.com is not going to work for you. Don't have it, have it like privatecellar.com or something. All the companies like shopping.com, money.com banged out. It should be a standalone name that associates you with something. Wine.com you think they sell wine, but when you go there they sell beer and wine. When you want to buy beer you would go to beer.com. The common name used to be very powerful but now hardly any more. In terms of that more and more search engines index amazon.com for books. They are not going to books.com. They are going to Amazon, Kalahari and Barnes & Noble. Because they expect them, that companies without the name in the URL are often more specific in terms of products so they are not looking at the URL as much. "

Vorster (2005) was further of the opinion that domain names are still used by a number of search engines. Search engines are starting to notice that the biggest bookstore in the world is not called books.com, but Amazon.com. Vorster (2005) then ended the topic by stating that search engines are in the process of changing the way they look at domain names.

5.4.2.2   HTML Naming Conventions or File Name

According to Vorster (2005), the majority of search engines do not even consider HTML naming conventions. This is due to the fact that developers use their own preferred naming convention and also that there is no universal standard with regard to naming conventions.

5.4.2.3 Frames

Vorster (2005) was of the opinion that frames are becoming more and more important as indexing technology improves. Furthermore, frames are used by web developers because the use of frames improves the ease-of-use of websites and can more often than not be built within a day. Without frames, the same project can take up to a week. Building a site with frames is cheaper and better for maintenance. Search engines are starting to notice that more and more sites make use of frames, and therefore are developing techniques to index these pages.

According to Vorster (2005):

> "…everyone is going to be looking more and more at frames as more and more sites are developed with frames. Search engines could even start weighting frame sites more. They would look at these sites and see it as a fresher, newer and better site. If you are using frames now you might receive a lower ranking, but over time this could change."

5.4.2.4 Hypertext / Anchor Text

Vorster (2005) stated that Hypertext is currently one of the most important elements to achieve a more visible website. A link on a site linking to a developer's site, for example ABCwholesalers.com which is ranked by google.com as a zero, would not increase the website's visibility. If however the developer uses a link from another site linking to the developer's site, and that site were ranked highly by Google, it would make a difference. Vorster (2005) (verbatim as per the interview script) provided the following example in support of the above:

> "A guy who interlink to Ananzi with some odd link on Ananzi's homepage like weddings or something, something really diverse. He paid R10 000 for a link on Ananzi's homepage and he said I'm paying to much money for it and am not getting any traffic from it. He took it and said I don't want the contract any more. Two days later he came back and said; please put it back on for me. I'll pay R10 000 a month.

Now we told him that it would be R15 000 a month. He said: I'll pay it. I said: But you said you aren't getting any traffic from it. He said: I went from my Google ranking from the top three to below the 15th position, purely due to that link on your site. That link on your site is so valuable in terms of my ranking that I will pay R15 000 per month for it just to rank on Google. "

5.4.2.5   HTML

Referring to HTML code, Vorster (2005) stated that the developer should always try to keep the code used to develop the layout of the site as clean and basic as possible. Furthermore, Vorster (2005) expressed the opinion that:

"If you have a rubbish site, they would definitely pick it up."

The effect of spelling mistakes on webpages (see Paragraph 3.3.4) and its effect on website visibility however, vary and are still debated according to Vorster (2005).

5.4.2.6   Meta-tags

Vorster (2005) was of the opinion that all meta-tags are of low importance but should still be included even though they receive very little weight in the majority of search engines. Furthermore, should none of these tags be included, a site would not get ranked.

Vorster (2005) (verbatim as per the interview script) provided the following example in support of the above:

"…my car has to have four tyres to drive. In order for a car to be a car it has to function. All the tags are "have to haves" in order to function. But if you got them it wouldn't mean that you would get a high ranking, but if you haven't got them you are not going to get a ranking."

119

5.4.2.7   Spamming

Vorster (2005) was of the opinion that, should a developer get away with spamming, it would have been worth using it, as spamming was not illegal, it merely tricked the search engine. Vorster (2005) however, would not recommend the use of spamming to manipulate search engine algorithms. If a site gets blacklisted, search engines can remove the website's URL from the listings for six to 12 months. This does not mean that search engines move the sites from position one to position 100. It does mean that search engines will remove the website from their index.

Vorster (2005) (verbatim as per the interview script) provided the following sample in support of the above:

> "If for example company ABC is blacklisted, people would not be able to find this company on the web no matter what keywords they use to search for information. Don't take the risk."

5.4.2.8   Flash

Vorster (2005) stated that Flash may look "nice", but it is detrimental to ranking. He also stated that if a developer can get away without using Flash, the developer should rather do so. Flash though, is becoming more popular due to its usability factor.

Vorster (2005) (verbatim as per the interview script) provided the following sample in support of the above:

> "Developers are getting better and better. They are beginning to place more and more hidden links within Flash, but the general rule of thumb now is the less Flash the better. The more text the better. You'll probably find in the nearer times that if you've got Flash in your site that they would still index it the right way. No Flash means more hypertext, therefore at this point of time developers should rather make use of hypertext than Flash."

5.4.2.9 Banner advertising

Banner ads according to Vorster (2005) often provide links, but seldom in the form of a hyperlink (URL). Furthermore, banner ads usually contain a script or text link which communicates to a banner server when clicked. When a user clicks a banner, it will track the user with the assistance of a "cookie" and then push the user to the site advertised on the banner. As a result, it depends on the type of banner ad used, and the link behind the banner.

5.4.2.10 Link popularity

Vorster (2005) stated that link popularity is currently one of the most important elements necessary to achieve a more visible website. Furthermore, should a press release be written and displayed on a site, and CNN publishes a link on its site containing the article, it is worth thousands of dollars. It will dramatically increase the visibility of a webpage containing the article, due to the relevance and importance level of CNN's website.

Vorster (2005) (verbatim as per the interview script) provided the following sample in support of the above:

> "Search Engines would give CNN a nine out of ten ranking because they believe CNN won't put text links on their site that are rubbish. If they link to your site and the keywords around that link state the article contains for example information about apple trees and your site does contain information about apple trees the search engine would go; these site are really relevant. Let's give them a high ranking. "

As a result, a link on a more popular site increases the ranking as in Hypertext / Anchor Text (see Paragraph 5.4.2.4). In addition, a link to a good site would increase a site's visibility significantly.

5.4.2.11 JavaScript

Vorster (2005) had limited knowledge of JavaScript, and could not provide any recommendations on the advantages and disadvantages of this element.

5.4.3   Conclusion

Vorster (2005) was asked to rank the elements discussed according to importance on a scale from one to five, where one carried the highest importance and five carried the lowest importance.

Of the elements discussed, Vorster (2005) rated the inclusion of all meta-tags as the most important, due to the fact that the inclusion of these tags is an absolute requirement, even though they do not really increase the visibility of the site. Secondly, Vorster (2005) rated hypertext as a two, due to the fact that the majority of search engines look at back links. Link popularity was rated third. Limited use of Flash was rated fourth and the importance of domain names was rated last on the importance scale, due to the resemblance between the site and the domain name, which still carries a relatively high weight in some of the major search engines.

These top five elements identified through the interview are summarised in Table 5.20.

TABLE 5.20: Most important visibility elements identified in expert interview

Ranking	Element
1	Inclusion of all meta-tags
2	Prominent Hypertext
3	Link Popularity
4	No Excessive use of Flash
5	Prominent Domain Names

5.5   CLOSURE

The results gleaned from the survey proved that SMMEs are in need of improved website visibility. Only 6.3% of the 300 SMME websites used to conduct this study, ranked in the top 30 results of six leading search engines. The research further identified 12 possible visibility elements present in the websites appearing in the top results of four search engines (Google, MSN, Yahoo and Ananzi). The results gathered from academic literature and

quantitative research were triangulated against results gathered from an interview with a representative of a popular South African search engine (Ananzi).

It should be noted that even though the elements are rated according to importance, it does not guarantee a higher visibility should it be applied in that order. These elements provide a framework towards a more visible website and could vary as search engine algorithms adapt to new indexing techniques.

CHAPTER 6
CONCLUSION

6.1 INTRODUCTION

The purpose of this chapter is to draw the final conclusions from this research in this thesis. Both quantitative research and qualitative research will be combined and ranked to produce a potential model that could be used by developers as a basis to improve the visibility of a website. A summary of the results is depicted in the paragraphs below.

6.2 SUMMARY

6.2.1 Literature

Through the conducted literature review it has become evident that SMMEs with websites need assistance to improve the visibility of their websites, allowing them to attain the full potential of e-commerce. A visible website could provide SMMEs with the ability to successfully access a global information infrastructure. It has also been highlighted that, to achieve a visible website, the developer should have at least general knowledge about the different search engines and also how these search engines operate.

Research was conducted into the evolution of indexing techniques to better understand some fundamental processes used by information retrieval systems, and to identify the strength and weaknesses in their design. Two search service types were also identified, namely directories and search engines. These search engines were then further analysed by identifying different strategies used by the different types of search engines and directories.

The research has shown that search engines make use of different search algorithms to undertake the difficult task of effectively ranking increasing

website submissions. The algorithms used by these search engines also constantly vary from year to year, due to the development of search engine spam produced by developers to manipulate search engines' algorithms. The responsibility therefore lies with the developer to stay up to date with the latest trends regarding website visibility. This author has however identified several visibility elements which have carried substantial weight to enhance website visibility. While some of these elements currently carry very little weight, they should still be implemented for those search engines which still support them. Elements identified include the following:

- Frames.
- Hypertext / Anchor Text.
- HTML.
- Keywords.
- HTML Naming Conventions.
- Meta-tags.
  - Dublin Core.
  - Meta-title tag.
  - Meta-description tag.
  - Meta-keyword tag.
  - Meta-header tag.
  - Meta-robot tag.
  - Alt tags.
- Spamming.
  - Text Spam.
  - Link Spam.
  - Cloaking Spam.
  - Domain Spam.
- Graphics, Sounds, Video, PDF, Flash.
  - Graphics.
  - Flash.
  - PDF.
- Banner advertising.

- Awards.
- Link popularity.
- JavaScript.

It has been identified that even though some of these elements are not recommended by some authors in the literature review (e.g. JavaScript), a developer could still make use of them and still be visible by implementing them. The analogy can be drawn that a developer should always strive for simplicity when designing a webpage to ensure that crawlers cover every important part of the webpage.

6.2.2   Quantitative Research

A total of 144 websites were analysed to identify the elements used and avoided by website owners to achieve top positions. This study was conducted by making use of nine keywords and four search engines. Visibility elements gathered through academic literature were then ranked according to the usage of these visibility elements within the websites ranking in the top positions when searching for predetermined keywords. The elements present in these websites are ranked in Tables 6.1 and 6.2.

TABLE 6.1: Visibility elements present in top results

No	Visibility Elements Present in top results	Occurrence
1	Inclusion of some meta-tags	100%
2	Hypertext / Anchor text	100%
3	No Flash or fewer than 50% of content	100%
4	No Visible Link Spamming	96.50%
5	No Frames	94.40%
6	Prominent Link Popularity	84.00%
7	Prominent Headings	79.90%
8	No Banner Advertising	73.60%
9	Prominent Domain Names	68.80%
10	Prominent HTML Naming conventions	63.90%

TABLE 6.2: Meta-tags present in top results

No	Meta-tags present in top results	Occurrence
1	Prominent Title tag	94.40%
2	Prominent Description tag	50.70%
3	Prominent Keyword tag	49.30%
4	Prominent Alt tags	47.20%
5	Robot tag	43.80%
6	Prominent Header tag	20.80%
7	Dublin Core	5.60%

6.2.3   Qualitative Research

In this research process, an expert interview was conducted with a representative of a leading South African search engine. The goal of this interview was to confirm or to provide clarification on the importance of elements identified through the academic literature and quantitative research. The top five elements gathered from the expert interview are reflected in Table 6.3.

TABLE 6.3: Visibility elements identified in expert Interview

Rank	Visibility elements identified in expert Interview
1	Inclusion of all meta-tags
2	Prominent Hypertext
3	Link Popularity
4	No Excessive use of Flash
5	Prominent Domain Names

The element rated first (Inclusion of all meta-tags) is further summarised in Table 6.2.

6.2.4   Research Summary

To draw up the final model, the author assigned a ranking to the top elements identified through the different research approaches. Table 6.4 summarizes the ranking of the top ten elements identified through the quantitative approaches. The ten elements were ranked (see "Rank" column) to add up to "55" (1+2+3...+10). Since the first three elements occurred the

same number of times, a ranking of "2" was assigned to all three elements. In order that the sum of the ten rankings remains "55" (1+2+3 = 6 → 6/3 = 2), a ranking of 2 is assigned to the first three items.

TABLE 6.4: Ranking of visibility elements present in top results

Rank	Visibility Elements Present in top results	Occurrence
2	Inclusion of some meta-tags	100%
2	Hypertext / Anchor text	100%
2	No Flash or fewer than 50% of content	100%
4	No Visible Link Spamming	96.50%
5	No Frames	94.40%
6	Prominent Link Popularity	84.00%
7	Prominent Headings	79.90%
8	No Banner Advertising	73.60%
9	Prominent Domain Names	68.80%
10	Prominent HTML Naming conventions	63.90%

The ranking of Table 6.3 and Table 6.4 was combined in Table 6.5 to provide a final ranking for the compilation of a top ten list of elements which could serve as a potential model to improve the visibility of a website. Rankings of elements appearing in both research approaches were added and divided by two to produce an average ranking. Only one expert interview was used in the qualitative part of the research, while a high number of websites were included in the quantitative part of the research. The ranking system however, stabilizes the effect of this apparent imbalance. Quantitative results range from one to ten, while the qualitative results cover only one to five.

TABLE 6.5: Leading visibility elements – MODEL

Number	Leading Visibility Elements	Rank
1	Inclusion of meta-tags	1.5
2	Hypertext / Anchor text	2
3	No Flash or fewer than 50% of content	3
4	No Visible Link Spamming	4
5	Prominent Link Popularity	4.5
6	No Frames	5
7	Prominent Domain Names	7
8	Prominent Headings	7
9	No Banner Advertising	8
10	Prominent HTML Naming conventions	10

Table 6.5 represents the model which could be used by SMMEs as a basis to improve the visibility of their websites. A lower figure in the Rank column (Table 6.5) indicates a higher ranking, i.e. a more important element. It should be noted that while these potential elements may provide better visibility at the time of this study, improved ways to enhance website visibility evolve on a sustained basis. As a result, these elements should be seen as a foundation to build upon as opposed to a definite solution to a poorly visible website.

6.3   SIGNIFICANCE OF STUDY

By comparing the data gathered from the academic literature with data gathered from an experimental study, the author was able to identify to what extent these elements are used in websites ranking in the top positions. Data gathered from an expert interview was then used to triangulate data gathered. This resulted in a model being created (see Table 6.5) that is not only supported by research, but also by current website designers. This model furthermore provides possible techniques for increasing website visibility to small business organizations that lack the necessary funds, knowledge and skills to outsource their visibility needs.

It is furthermore anticipated that, should this model be implemented, the following outcomes can be expected by SMMEs:

- Higher ranking in search engine results due to higher website visibility.
- Increased crawler friendliness without sacrificing human friendliness.
- Increased content rich webpages benefiting the search engine and the user.
- Webpages which load faster, thus providing a huge benefit to users operating on slower connections.
- Increased usability of websites to disabled users by implementing text that describes the graphics.

- Increased usability made possible by the site map, allowing users to shift to any location on the website.

6.4   FUTURE RESEARCH

Future research could include the application of elements identified in the model into a website, submitting the modified website to search engines, and monitoring its ranking. The results can then be used to produce further research into website visibility and also validating the impact of certain top (e.g. top five) elements identified in the model.

6.5   FINAL CONCLUSION

While an accurate index is at the centre of a search engine system, the user's ability to access and retrieve information is of equal importance. To this end, commercial Internet search engines and directories all try to ensure that the searched results presented are beneficial to users.

To design and maintain a visible website is a time consuming process and should be pursued on a sustained basis. By achieving visibility today does not mean that an optimally visible website is evident the next day. It is a process that could take months to master successfully. As Internet technology evolves, new search engine algorithms will evolve in parallel. For search engines to maintain top quality results, they need to develop new indexing techniques, partially due to new spamming techniques and to attain high positions.

At the time of writing, search engines are taking decisions on dynamic pages and Flash, but in the near future ways to overcome this problem may be found. As a result, elements which received lower importance ratings should not necessarily be ignored. Some of the elements identified could certainly improve the ranking of a website while others would hardly be noticed by a search engine. Developers should perform constant research into website visibility to ensure that they are up to date with the latest trends and technologies. Alexander attests accordingly in Quotegarden.com (2005):

"We learn more by looking for the answer to a question and not finding it than we do from learning the answer itself."

This author is of the opinion that the research question - *Can an effective visibility strategy be developed to ensure that SMMEs draw website traffic which could potentially lead to increased sales?* - has been successfully answered through the research in this thesis. An effective visibility strategy has been developed by means of a model (see Table 6.5), which provides a solid basis to ensure website visibility. Even though no increase in visibility was guaranteed, it was proven that these elements are currently the elements implemented by a large number of visible e-commerce companies.

REFERENCES

Adams, I. 2003. Overcoming webpage printing problems. *Online Magazine*, Issue September/October, 36-41.

Alimohammadi, D. 2003. Meta-tag: a means to control the process of web indexing. *Online Information Review,* 27(4):238-242.

Baard, V.C. 2004. Interactive information consulting system for South African small businesses – Part 1. *South African Journal of Information Management,* 6(2), June. http://www.sajim.co.za [20 February 2005].

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

Bearman, D., Miller, E., Rust, G., Trant, J. & Weibel, S. 1999. A common model to support Interoperable metadata: Progress report on reconciling metadata requirements from the Dublin core and INDECS/DOI Communities. *D-Lib Magazine,* 5:(1), January.
http://www.dlib.org/dlib/january99/bearman/01bearman.html [13 May 2005].

Bennett, R. 1997. Export marketing and the Internet: Experiences of website barriers among UK businesses. *International Marketing Review,* 14(5):324-344.

Bester, F. 2003. Challenges and risks for South Africa in the information economy. *South African Journal of Information Management,* 5(2), June. http://www.sajim.co.za [01 February 2005].

Blair, D.C. 1990. *Language and representation in Information retrieval*. Elsevier Science Publishers. AE: Amsterdam.

Boter, H. 2005. SME perspectives on business support services: The role of company size, industry and location. *Journal of Small Business and Enterprise Development,* 12(2):244-258.

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

Bridges. 2003 . *Minnesota's Gateway to Environmental Information*. http://www.bridges.state.mn.us/ [12 April 2005].

Callan, D. 2004. *Your Domain Name*. http://www.akamarketing.com/your-domain-name.html [23August 2005].

Cape Chamber of Commerce. 2005a. *Mission Statement*. http://www.capechamber.co.za/mission.asp [30 July 2005].

Cape Chamber of Commerce. 2005b. *Members*. http://www.capechamber.co.za/members.asp [15 July 2005].

Clark, M.A, Chou, D.C. & Yen, D.C. 2001. Domain names management: a strategy for electronic commerce security. *Information Management and Computer Security,* 9(5):225-232.

Coca-Cola. 2003. *The Coca-Cola company.* http://www.coca-cola.com/FlashIndex1.html [10 August 2005].

Constantinides, E. 2002. From physical marketing to web marketing: the web-marketing mix. *Proceedings of the 35th Hawaii International Conference on System Science – 2002*, Big Island, Hawaii, 7-10 January 2002. New York: IEEE: 7:2628- 2638.

Cool, A. 2003. Solving e-commerce issues: some web strategies. *South African Journal of Information Management,* 5(1), March. http://www.sajim.co.za [28 February 2005].

Cooper, D.R. & Schindler, P.S. 2003. *Business research methods.* 8th ed. New York: McGraw-Hill Education Publishing.

Craven, T.C. 2001. Changes in meta-tag descriptions over time. *First Monday,* 6(10), October. http://firstmonday.org/issues6_10/craven/index.html [28 February 2005].

Craven, T.C. 2002. What is the title of a webpage? A study of Webography practice. *Information Research Journal,* 7(3), April. http://InformationR.net/ir/7-3/paper130.html [28 February 2005].

Craven, T.C. 2003. HTML tags as extraction cues for webpage description construction. *Information Science Journal,* 6():1-12.

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

Dogpile.com. 2005. *What makes Dogpile special?* http://www.dogpile.com/info.dogpl/search/hep/about.htm [12 April 2005].

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

Durkin, M. & McGowan, P. 2001. 'Net effect' – views from the periphery: exploring the role and importance of the Internet on marketing activity in entrepreneurial firms. *Irish Marketing Review*, 14(1):15-25.

Eldabi, T., Irani, Z., Paul, R.J. & Love P.E.D. 2002. Quantitative and qualitative decision-making methods in simulation modelling. *Management Decision*, 40(1):64-73.

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

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

Gandre, A. & Lebbe, R.V. 2005. WP5 – Indexing: Final indexing report: A biodiversity collection access service for Europe (BioCASE). University Pierre Et Marie Curie, 1-14. http://www.biocase.org/Doc/Results/WP5/D34_Final %20Indexing%20Report.pdf [12 May 2005].

Gerhart, A. 2001. The roadblock between your website and customers is Flash. *SEO strategies.* http://seotoday.com/browse.php/category/articles/ id/127/index.php [30 May2003].

Goh, D.H. & Wang, J.C. 2004. Designing a virtual exhibition using Scalable Vector Graphics. *Proceedings of New Information Perspectives,* 56(3):144-155, March.

Google.com. 2004. *Frequently Asked Questions.* http://www.google.com/help/faq_filetypes.html [30 May 2005].

Google.com. 2005. *Google Information for Webmasters.* http://www.google.co.za/webmasters/guidelines.html [12 September 2005]

Gorman. R.T. 2000. *Get your business online: How to conceptualize, design, and build an effective business website in less than 30 days.* USA:Streetwise Publishing.

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

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

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

Guenther, K. 2004c. What's your style: Organizing for the web. *Online Magazine*, Issue July/August, 46-47.

Henzinger, M.R., Motwani, R. & Silverstein, C. 2002. *Challenges in web search engines.* http://www.acm.org/sigs/sigir/forum/F2002/henzinger.pdf [2 April 2005].

Hoffman, E.C. 2002. Information and communications technology, virtual offices and telework. *South Africa Journal of Information Management,* 4(2), June. http://www.sajim.co.za [20 May 2005].

Hubbard, J. 2004. *Indexing the Internet*. Essay for Drexel University College of Information Science and Technology. http://www.tk421.net/essays/babel.html [25 May 2005].

Internet Systems Consortium. 2005. *ISC domain survey: Number of Internet hosts.* http://www.isc.org/index.pl?/ops/ds/host-count-history.php [12 April 2005].

Isaacs, Z. 2004. *Good IT strategy helps SMMEs get it right.* http://www.CITInews.com [10 March 2005].

Jansen, B.J. 2000. The effect of query complexity on Web searching results. *Information Research*, (6)1, October. http://InformationR.net/ir/6-1/paper87.html [15 August 2005].

Kiani, G.R. 1998. New game, new rules: Will traditional mentality work in the marketspace? *Management Research News,* 21(6):1–9.

Konia B.S. 2002. *Search Engine optimization with WebPosition GOLD^{TM} 2.* Texas: Wordware Publishing.

Kritzinger, W.T. & Weideman, M. 2005. A study on the correct usage of webpage keywords to improve search engine ranking. *Proceedings of the 7th Annual Conference on WWW Applications*. 31 August - 2 September. South Africa, Cape Town. http://www.uj.ac.za/www2005/ [12 August 2005].

Leacock, T. 2005. Building a sustainable e-learning development culture. *The Learning Organization,* 12(4):355-367.

Lubbe, S. 2003. Development of a case study methodology in the information technology (IT) field in South Africa: a step-by-step approach. *South Africa Journal of Information Management*, 5(4). http://www.sajim.co.za [13 July 2005].

Mamma.com. 2005. *About Mamma*. http://www.mamma.com/info/about.html [12 April 2005].

Marketleap. 2005. *Marketleap Search Engine Marketing Tools.* http://tools.marketleap.com/verify/ [10 September 2005].

MINIUSA.COM. 2005. *Build your own*. http://www.miniusa.com/link/buildyourown/coopers/ [10 September 2005].

Nielsen/NetRatings. 2005. *Nielsen/Netratings releases top 10 search engine share rankings from July 2005*. http://www.netratings.com/pr/pr_050824.pdf [05 September 2005].

Nobles, R & O'Neil S. 2000. *Maximize website traffic*: build website traffic fast and free by optimizing search engine placement. USA: Avon, MA: Corporation.

Notess, G.R. 1999. On the Net: Rising relevance in search engines. *Online Magazine,* Issue May. 84-87.

Notess, G.R. 2003. A decade on the Net. *Online Magazine,* Issue March/April, 53-55.

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

Poulter, A. 1997. The design of World Wide Web search engines: a critical review. *Program,* 31(2):131-145.

Quotegarden.com. 2005. Quotations about learning. http://www.quotegarden.com/learning.html [15 September 2005].

Rajgopal, S., Kotha, S. & Venkatachalam, M. 2000. The relevance of web traffic for Internet stock prices. *Research Paper No. 1616,* 1-33. http://www.ssrn.com/abstract=207989 [12 May 2005].

Rogerson, C.M. 2001. Towards a framework for rural SMME development in South Africa, in M.Khosa (ed), Empowerment through economic transformation. Durban: *African Millennium Press.* 271-298.

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

Rowley. J. 2004. Online Branding. *Online Information Review,* 28(2):131-138.

Ru, Y. & Horowitz, E. 2005. Indexing the invisible web: a survey. *Online Information Review*, 29(3):249-265.

Searchengines.com. 2003. *Don't miss DirectHit -- The ins and outs of click popularity and stickiness.* http://www.searchengines.com/directhit.html [2 April 2005].

Silverstein, C., Henzinger, M.R., Marais, H. & Moricz, M. 1999. Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum,* 33(1):6-12.

Simpande, J. & Jakovljevic, M. 2003. Integrating the Internet and enterprise resource planning (ERP) systems in South Africa electricity utility companies. *South African Journal of Information Management,* 5(1). http://www.sajim.co.za [20 April 2005].

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

Smith, A.G. 2002. Does metadata count? A Webometric investigation. *Proceedings of DC-2002,* Florence, 14-17 October 2002. 133-138.

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

South Africa Business Guidebook. 2003. *Boosting small businesses*. http://www.southafrica.info/doing_business/economy/development/smallbusiness.htm [13 May 2005].

Sparkes, A. & Thomas, B. 2001. The use of the Internet as a marketing of Welsh agri-food SMEs in the twenty-first century. *British Food Journal,* 103(5):331-347.

Spink, A. & Xu, J.L. 2000. Selected results from a large study of Web searching: the Excite study. *Information Research*, 6(1). http://InformationR.net/ir/6-1/paper90.html [15 February 2005].

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

Struwig F.W. & Stead G.B. 2001. *Planning, designing and reporting research*. Pearson Education Publishing. South Africa.

Subia Creative. 2002. Justifying an investment in search engine Marketing. http://www.subiainteractive.com/art-pref-leads.html [20 May 2004].

Sullivan, D. 2002. *Using Direct Hit Popularity Results*. http://searchenginewatch.com/sereport/article.php/2165041 [13 June 2005].

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

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

Van Steenderen, M. 2001. Web site management: making a Web site more visible. *South African Journal of Information Management,* 2(4). http://www.sajim.co.za [13 May 2005].

Van der Westhuizen, M. 2001. The invisible Web. *South African Journal of Information Management,* 3(4). http://www.sajim.co.za [13 May 2005].

Vorster, F. 2005. Interview with the representative of Ananzi on 30 August 2005. [Recorded on audio cassette].

Weibel, S., Godby, J., Miller, E. & Daniel, R. 1995. *OCLC/NCSA Metadata Workshop Report.* http://www.oasis-open.org/cover/metadata.html [13 May 2005].

Weideman, M. 2001. Internet Searching as a study aid for Information Technology and Information Systems learners at a tertiary level. Doctor of Philosophy thesis, University of Cape Town, Cape Town.

Weideman, M. 2002. Effective application of metadata in South African HEI websites to enhance visibility to Search Engines. *Proceedings of the 4th Annual Conference on WWW Applications (WWW2002)*, 4 - 6 September 2002, Bellville, South Africa, (CD-ROM).

Weideman, M. 2004. Ethical issues on content distribution to digital consumers via paid placements, as opposed to website visibility in search engine results. *Proceedings of ETHICOMP 2004*, 14-16 April 2004, University of the Aegean, Syros. (2):904-915.

Weideman, M. & Haig-Smith, T. 2002. An investigation into search engines as a form of targeted advert delivery. *Proceedings of South African Institute of Computer Scientists and Information Technologists (SAICSIT): Enablement Through Technology*, 16 - 18 September 2002, Port Elizabeth, South Africa, 258.

Weideman, M. & Kritzinger, W. 2003. Search engine information retrieval: Empirical research on the usage of meta-tags to enhance website visibility and ranking of e-commerce websites. *Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI2003), Information Systems, Technologies and Applications*, 28 - 30 July 2003, Orlando, USA. 231 - 236.

Wikipedia.com. 2005. *Statistics*. http://en.wikipedia.org/wiki/Statistical [21July 2005].

Wilson, K.C. 2002. *Automatic Indexing: Problems and Solutions. Strathclyde University.* http://www.humbul.ac.uk/ltsn-humbul/survey/survey _appendix8.doc [12 May 2005].

Wood, L.E. 1998. *Ethnographic interviewing for client-centered design*. Provo, UT: Brigham Young University, Department of Psychology. http://www.byu.edu/~woodl/paper.htm [30 July 2005].

Yahoo.com. 2004. *What is advanced search?.* http://help.yahoo.com/help/us/ysearch/basics/basics-08.html [30 May 2005]

Xavier, S. 2005. Are you at the top of your game? Checklist for effective leaders. *Journal of Business Strategy,* 26(3):32-42.

Zhang, J. & Dimitroff, A. 2004. The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing and Management*, 41(2005):665-690, January.

APPENDIX A
CRITERIA 1

Criteria 1

1. Domain Names

1.1 If the site is a business site, does it have a virtual domain name? (www.companyname.com)? — Yes ☐ No ☐

1.1 If the site is not a business, does the domain name contain keywords related to the site's products or services offered? — Yes ☐ No ☐

1.1 Please indicate how you would describe domain name of the site:
short — Yes ☐ No ☐
easy to pronounce and spell — Yes ☐ No ☐
descriptive — Yes ☐ No ☐
memorable. — Yes ☐ No ☐

2. HTML Naming Conventions

2.1 Please indicate how you would describe the file naming conventions of the homepage:
short — Yes ☐ No ☐
easy to say and spell — Yes ☐ No ☐
descriptive — Yes ☐ No ☐
memorable — Yes ☐ No ☐

2.2 If the filename represents 2 keywords is the developer making use of a hyphen or an underscore e.g. cycling-news.html or cycling_news.html? — Yes ☐ No ☐

3. Frames

3.1 Does the main page contain frames? — Yes ☐ No ☐

3.2 If the website does contain frames, did the developer made use of the <noframes> tag? — Yes ☐ No ☐

4. Hypertext / Anchor Text

4.1 Does the main page contain any hypertext? — Yes ☐ No ☐

4.2 Did the developer made use of keywords in the hypertext describing the most important products and services offered by the website? — Yes ☐ No ☐

5. HTML

5.1 Did the developer made use of keywords describing the products and services offered by the website? — Yes ☐ No ☐

6. Meta-tags

6.1 Dublin Core

6.1.1 Does the site make use of Dublin Core? — Yes ☐ No ☐

6.2 Meta-title tag

6.2.1 Does the title of the main site contain the company name or part of it? — Yes ☐ No ☐

6.2.2 Are there any unnecessary punctuation marks, sales hype, and filter words (a, an, but, or, nor, for, the, etc) in the title? — Yes ☐ No ☐

6.2.3 Is the title written in capital letters? — Yes ☐ No ☐

6.2.4 Are there any abbreviations or acronyms in the description that are commonly understood? — Yes ☐ No ☐

6.2.5 Did the developer made use of keywords in the hypertext describing the most important products and services offered by the website? — Yes ☐ No ☐

6.2.6 Is the title longer than 10 words? — Yes ☐ No ☐

6.3 Meta-description tag

6.3.1 Does the main page make use of the Meta-description tag? (if so, please continue else proceed to next heading) — Yes ☐ No ☐

6.3.2 Is the title repeated in the meta-description tag? — Yes ☐ No ☐

6.3.3 Are there any abbreviations or acronyms in the description that are commonly understood? — Yes ☐ No ☐

6.3.4 Is the description tag used correctly? (e.g. not just a collection of keywords) — Yes ☐ No ☐

6.3.5 Did the developer made use of keywords in the hypertext describing the most important products and services offered by the website? — Yes ☐ No ☐

6.4 Meta-keyword tag

6.4.1 Does the main page make use of the Meta-keyword tag? (if so, please continue else proceed to next heading) — ☐ Yes ☐ No

6.4.2 Are there any abbreviations or acronyms in the description that are commonly understood? — ☐ Yes ☐ No

6.4.3 Did the developer made use of keywords in the hypertext describing the most important products and services offered by the website? — ☐ Yes ☐ No

6.5 Meta-header tag

6.5.1 Does the main page make use of the Meta-header tags? (if so, please continue else proceed to next heading) — ☐ Yes ☐ No

6.5.2 Does the site make use of header tags (H1- H6)? — ☐ Yes ☐ No

6.5.3 Are there any abbreviations or acronyms in the description that are commonly understood? — ☐ Yes ☐ No

6.5.4 Did the developer made use of keywords in the hypertext describing the most important products and services offered by the website? — ☐ Yes ☐ No

6.6 Meta-robot Tag

6.6.1 Does the site contain any Meta-robot tags? — ☐ Yes ☐ No

6.7 Alt Tags

6.7.1 Does the main page make use of the Alt tags? (if so, please continue else proceed to next heading) — ☐ Yes ☐ No

6.7.2 Are there any abbreviations or acronyms in the description that are commonly understood? — ☐ Yes ☐ No

6.7.3 Did the developer made use of keywords in the hypertext describing the most important products and services offered by the website? — ☐ Yes ☐ No

7. Spamming

7.1 Does the main page contain any visible forms of spamming eg. text spam, link spam, cloaking spam? — ☐ Yes ☐ No

8. Flash

8.1 Does the main page make use of Flash images? (if so, please continue else proceed to next heading) — ☐ Yes ☐ No

8.2 Does the developer provide the user with an option to load a duplicate of the site without flash? — ☐ Yes ☐ No

8.3 Does the site contain Flash in its navigational structure? — ☐ Yes ☐ No

9. PDF

9.1 Is the site in PDF format? — ☐ Yes ☐ No

10. Banner advertising

10.1 Does the main page contain any banner advertising? — ☐ Yes ☐ No

10.2 If the site does contain banner ads, do the banner ads reflect some of the content of the site? — ☐ Yes ☐ No

11. Link popularity

11.1 Does the site contain any links to other sites other than banners? — ☐ Yes ☐ No

11.2 If the site does contain links to other sites, is the links to the other sites in context with the content of the current site? — ☐ Yes ☐ No

12. JavaScript

12.1 Does the main page make use of the JavaScript? (if so, please continue else proceed to next heading) — ☐ Yes ☐ No

12.2 Is the JavaScript code enclosed in a '.js' file? — ☐ Yes ☐ No

12.3 Did developer make use of the <noscript> tag? — ☐ Yes ☐ No

Criteria 2

1. Domain Names

1.1. Is the website an e-commerce site (business oriented)? □ Yes □ No
If your answer is no, please proceed to question 1.3
1.2 Does the website have a virtual domain name (e.g. **www.companyname.com**)? □ Yes □ No
1.3 Does the domain name contain keywords related to products, services or company name □ Yes □ No
1.4 Please indicate how you would describe domain name of the site:
    short (2 or less keywords) □ Yes □ No
    easy to pronounce and spell □ Yes □ No
    descriptive □ Yes □ No
    memorable. □ Yes □ No

2. HTML Naming Conventions

If you are on the homepage of the site then please follow any link from the homepage to another part of the page and continue. Should you not be on homepage then continue.
2.1 Please indicate how you would describe the file naming conventions of the website:
    short (2 or less keywords) □ Yes □ No
    easy to say and spell □ Yes □ No
    descriptive □ Yes □ No
    memorable □ Yes □ No

If the filename represents 2 or more keywords proceed to 2.2 else advance to section 3
2.2 Is the developer making use of a hyphen and/or an underscore e.g. cycling-news.html or cycling_news.html? □ Yes □ No

Before continuing please return to previous page if you followed a link.

3. Frames
3.1 Does the main page contain frames? □ Yes □ No
If your answer is no please continue to next heading.
3.2 If the website does contain frames, did the developer made use of the <noframes> tag? □ Yes □ No

4. Hypertext / Anchor Text
4.1 Does the main page contain any hypertext? □ Yes □ No
If your answer is no please continue to next heading.
4.2 Does the keywords in the anchor text accurately describe the products and services offered by the website? □ Yes □ No

5. HTML
5.1 Does the keywords in the headings accurately describe the products and services offered by the website? □ Yes □ No

6. Meta-tags
Please right-click on website and select view source

6.1 Dublin Core
6.1.1 Does the site make use of Dublin Core? □ Yes □ No

6.2 Meta-title tag
If the website is not an E-commerce website then proceed to question 6.2.2
If the website contains no title, please proceed with section 6.3
6.2.1 Does the title of the main site contain the company name or part of it? □ Yes □ No
6.2.2 Are there any unnecessary punctuation marks, sales hype, and filter words (a, an, but, or, nor, for, to) in the title? □ Yes □ No
6.2.3 Is the whole title written in capital letters? Eg. SACYCLING – ALL YOUR CYCLING NEEDS □ Yes □ No
6.2.4 Are the any abbreviations or acronyms in the description commonly understood? □ Yes □ No
6.2.5 Does the contain keywords related to products, services or company name? □ Yes □ No
6.2.6 Is the title longer than 10 words (**excluding 'Internet Explorer'**)? □ Yes □ No

6.3 Meta-description tag
6.3.1 Does the main page make use of the Meta-description tag? (if so, please continue else proceed to section 6.4) □ Yes □ No
6.3.2 Is the title repeated in the meta-description tag? □ Yes □ No
6.3.3 Are there any abbreviations or acronyms in the description that are commonly understood? □ Yes □ No
6.3.4 Is the description tag used correctly? (e.g. not just a collection of keywords) □ Yes □ No
6.3.5 Does the description tag contain keywords related to products, services or company name? □ Yes □ No

6.4 Meta-keyword tag
6.4.1 Does the main page make use of the Meta-keyword tag? (if so, please continue else proceed to section 6.5) □ Yes □ No
6.4.2 Are there any abbreviations or acronyms in the description that are commonly understood? □ Yes □ No
6.4.3 Does the keyword tag contain keywords related to products, services or company name? □ Yes □ No

6.5 Meta-header tag
6.5.1 Does the site make use of header tags (H1- H6)? (if so, please continue else proceed to section 6.6) □ Yes □ No
6.5.2 Are there any abbreviations or acronyms in the description that are commonly understood? □ Yes □ No
6.5.3 Does the header tag contain keywords related to products, services or company name? □ Yes □ No

6.6 Meta-robot Tag
6.6.1 Does the site contain any Meta-robot tags? □ Yes □ No

6.7 Alt Tags
6.7.1 Does the main page make use of the Alt tags? (if so, please continue else proceed to next heading) □ Yes □ No
6.7.2 Are there any abbreviations or acronyms in the description that are commonly understood? □ Yes □ No
6.7.3 Does the alt tags contain keywords related to products, services or company name? □ Yes □ No

7. Spamming
7.1 Does the main page contain any visible forms of link spamming? □ Yes □ No

8. Flash
8.1 Does the main page make use of Flash images? (if so, please continue else proceed to section 9) □ Yes □ No
8.2 Does the developer provide the user with an option to load a duplicate of the site without flash? □ Yes □ No
8.3 Does the site contain Flash in its navigational structure? □ Yes □ No

9. Banner advertising
9.1 Does the main page contain any banner advertising? (if so, please continue else proceed to section 11) □ Yes □ No
9.2 Does the banner ads reflect some of the content of the site? □ Yes □ No

10. Link popularity
10.1 Does the site contain any links to other areas parts of the site? (if so, please continue else proceed to section 12) □ Yes □ No
10.2 If the site does contain links to other sites, are the links to the other sites in context with the content of the current site? □ Yes □ No

11. JavaScript
1.1 Does the main page make use of the JavaScript? (if so, please continue else proceed with next evaluation) □ Yes □ No
11.2 Is the JavaScript code enclosed in a '.js' file? □ Yes □ No
11.3 Did developer make use of the <noscript> tag? □ Yes □ No

Criteria 3

Keyword no – Ranking – Search Engine

1. Domain Names (www.sa-cycling.com)

1.1 Does the domain name (www.sa-cycling.com)contain keywords related to products, services or company name? ☐Yes ☐No

1.2 Please indicate how you would describe domain name (www.sa-cycling.com) of the site (answer all four questions):
- short (2 or less keywords) ☐Yes ☐No
- easy to pronounce and spell ☐Yes ☐No
- descriptive ☐Yes ☐No
- memorable. ☐Yes ☐No

2. HTML Naming Conventions or File Name (www.sa-cycling.com/contact_list.html)

If you are on the homepage of the site then please follow any link from the homepage to another part of the site and continue. Should you not be on homepage then continue.

2.1 Please indicate how you would describe the file (www.sa-cycling.com/contact_list.html) naming conventions of the website (answer all four questions):
- short (2 or less keywords) ☐Yes ☐No
- easy to say and spell ☐Yes ☐No
- descriptive ☐Yes ☐No
- memorable ☐Yes ☐No

2.2 Does the filename contain keywords related to products, services or company name? ☐Yes ☐No

If the filename represents 2 or more keywords proceed to 2.2 else advance to section 3

2.3 Is the developer making use of a hyphen and/or an underscore e.g. cycling-news.html or cycling_news.html? ☐Yes ☐No

Before continuing please return to previous page if you followed a link.

3. Frames

3.1 Does the main page contain any visible frames? ☐Yes ☐No

If answer is no please continue to section 4.

3.2 If the website does contain frames, did the developer made use of the <noframes> tag? ☐Yes ☐No

4. Hypertext / Anchor Text

4.1 Does the webpage contain any hyperlinks (text which link to other parts of the website)? ☐Yes ☐No

If your answer is no please continue to section 5.

4.2 Do the keywords in the hypertext accurately describe the products and services offered by the website? ☐Yes ☐No

5. HTML

5.1 Do the keywords in the headings accurately describe the products and services offered by the website? ☐Yes ☐No

6. Meta-tags

6.1 Dublin Core

Please right-click on website and select view source

6.1.1 Does the site make use of Dublin Core? ☐Yes ☐No

6.2 Meta-title tag

6.2.1 Does the page make use of the Meta-Title tag? (if yes, please continue, else proceed to section 6.3) ☐Yes ☐No

6.2.2 Does the title contain keywords related to products, services or company name? ☐Yes ☐No

6.2.3 Are there any unnecessary punctuation marks or filter words (a, an, but, or, nor, for, the) in the title? ☐Yes ☐No

6.2.4 Is the whole title written in capital letters? Eg. SACYCLING – ALL YOUR CYCLING NEEDS ☐Yes ☐No

6.2.5 Are the words in the title tag commonly understood? ☐Yes ☐No

6.2.6 Is the title longer than 10 words (excluding ' Microsoft Internet Explorer')? ☐Yes ☐No

6.3 Meta-description tag

6.3.1 Does the page make use of the Meta-description tag? (If yes, please continue, else proceed to section 6.4) ☐Yes ☐No

6.3.2 Is the title repeated in the meta-description tag? ☐Yes ☐No

6.3.3 Are the words in the description tag commonly understood? ☐Yes ☐No

6.3.4 Is the description tag used correctly? (e.g. not just a collection of keywords) ☐Yes ☐No

6.3.5 Does the description tag contain keywords related to products, services or company name? ☐Yes ☐No

6.4 Meta-keyword tag

6.4.1 Does the page make use of the Meta-keywords tag? (if yes, please continue, else proceed to section 6.5) ☐Yes ☐No

6.4.2 Are the words in the keyword description commonly understood? ☐Yes ☐No

6.4.3 Does the keyword tag contain keywords related to products, services or company name? ☐Yes ☐No

6.5 Meta-header tag

6.5.1 Does the site make use of any one of the 6 header tags (H1- H6)? (if yes, please continue, else proceed to section 6.6) ☐Yes ☐No

6.5.2 Are the words in the header tag description commonly understood? ☐Yes ☐No

6.5.3 Does the header tag contain keywords related to products, services or company name? ☐Yes ☐No

6.6 Meta-robot Tag

6.6.1 Does the site contain any Meta-robot tags? ☐Yes ☐No

6.7 Alt Tags

6.7.1 Does the main page make use of Alt tags? (if yes, please continue, else proceed to section 7) ☐Yes ☐No

6.7.2 Are the words in the alt tag description commonly understood? ☐Yes ☐No

6.7.3 Do the alt tags contain keywords related to products, services or company name? ☐Yes ☐No

7. Spamming

7.1 Does the main page contain any visible form (not just a collection of links) of link spamming? ☐Yes ☐No

8. Flash

8.1 Does more than 50% of the webpage content consist out of Flash images? (if yes, please continue else proceed to section 9) ☐Yes ☐No

8.2 Does the developer provide the user with an option to load a duplicate of the page without flash? ☐Yes ☐No

8.3 Does the page contain Flash in its navigational structure (buttons, options)? ☐Yes ☐No

9. Banner advertising

9.1 Does the main page contain any banner advertising? (if yes, please continue else proceed to section 10) ☐Yes ☐No

9.2 Do the banner ads reflect some of the content of the page? ☐Yes ☐No

10. Link popularity

10.1 Does the page contain any links to other areas of the site? (if yes, please continue else proceed to section 11) ☐Yes ☐No

10.2 Are the links to the other areas in context with the content of the current page? ☐Yes ☐No

11. JavaScript

11.1 Does the main page make use of the JavaScript? (if yes, please continue else proceed with next evaluation) ☐Yes ☐No

11.2 Is the JavaScript code enclosed in a '.js' file? ☐Yes ☐No

11.3 Did the developer make use of the <noscript> tag? ☐Yes ☐No

APPENDIX D
DEFINITIONS

Frames
The use of multiple, independent sections to create a single Web page. Each frame is built as a separate HTML file but with one "master' file to identify each section. When a user requests a page with frames, several pages will be displayed as panes. Sites using frames may report one page request with several panes as multiple page requests. Most audit firms count only the master HTML page request and therefore can accurately report the page requests.

Domain Names
The names and words that companies designate for their registered Internet Web site addresses, such as the "Forbes" magazine name in the URL http://www.forbes.com. Trademark disputes arise when more than one company tries to use the same domain name, or one company appropriates another company's brand or product name for its URL.

HTML Naming conventions
Class names typically are comprised of words that each begin with a Capital Letter. Method names usually begin with a lowercase letter and each subsequent word is capitalized. Both member and local variable names are usually in one of two formats. They could start with a lower case letter and each subsequent word is capitalized. Or they could contain underscores in between words or trailing single word names.

Hyperlinks / Anchor Text
A non-sequential method for reading a document displayed on a computer screen. Instead of reading the document in sequence from beginning to end, the reader can jump to topics by selecting a highlighted word or phrase embedded within the document. This activates a link, connecting the reader to another place in the same document or to another document. The resulting matrix of links is called a web.

HTML
(Hypertext Markup Language) The document format language used on the World Wide Web. Web browsers read HTML and display the page.

Keywords
A word that you use to search for a topic. Keywords are retrieved from words the author used in the article and need not be from an official list. Also known as "natural vocabulary."

Meta Tags
Keywords inserted in the meta-tag portion of the HTML source document by the Web page author. If Web pages don't have much text, meta-tags help them come up in a keyword search.

Dublin Core
New browser window will open for the Dublin Core Metadata Element Set. Dublin Core Metadata Element Set (Dublin Core Metadata Initiative): A standard set of about 15 elements (title, creator, subject, etc.), with optional qualifiers, used to structure descriptive records and facilitate information sharing. Originally intended for use in describing Web-based resources, it is now used also for describing physical collections in museums, libraries, archives, and other repositories.

Title Tag
An HTML tag used to define the text in the top line of a Web browser, also used by many search engines as the title of search listings and for bookmark identification.

Description Tag
HTML tag used to by Web page authors to provide a description for search engine listings.

Keyword Tag
A meta tag listing the relevant keywords a user may enter when trying to find a specific category on a search engine. You should have no more than 18 unique keywords or short phrases separated by commas, with any single word never repeated more than 7 times.

Header Tag
This is a tag that serves as a logical divider of the text into parts and as a text heading. Header tags are used by the program to make a TOC. There are six header tags are defined in HTML, ranged by importance. It is defined as: <h#>....</h#>, where # is [1...6].

Robot.txt
A file written and stored in the root directory of a website that restricts the Search Engine spiders from indexing certain pages of the website. This file is used to disallow certain spiders from seeing files that you not want them to see. You can also prevent a certain spider to look at any of the web pages through this file.

Alt Tags
Refers to alternative text that displays in place of an image on a Web page. Alt text is used by text-only browsers and visually impaired users who rely on screen readers. In some browsers, this text also appears while the image is downloading, if the image link is broken, or when the pointer is placed over the image.

Text Spam
Spammers quickly created techniques to modify the text in such a way that the search engine rated the page higher than it deserved

Link Spam
Link spam (also called blog spam or comment spam) is a form of spamming or spamdexing that recently became publicized most often when targeting weblogs (or blogs), but also affects wikis (where it is often called wikispam), guestbooks, and online discussion boards. Any web application that displays hyperlinks submitted by visitors or the referring URLs of web visitors may be a target.

Cloaking Spam
The process by which your site can display different pages under different circumstances. It is primarily used to show an optimized page to the search engines and a different page to humans. Most search engines will penalize a site if they discover that it is using cloaking.

Banner Advertising
A marketing mechanism that contains strips of advertisements that are sporadically positioned on a web page and are extremely popular on the World Wide Web. These types of ads generally take up a considerable amount of bandwidth and are sometimes disturbing to the Web user.

Awards
Awards which could range from graphic awards to content awards are yet another way of improving a website's visibility. Receiving awards hold several advantages, not just to the user, but also to the developer.

Link popularity
Essentially a measure of how many other sites indexed by the same search engine have links to your site. Link popularity is essentially simply a count of links to a particular site; see also link analysis.

JavaScript
JavaScript is a scripting language developed by Netscape to enable Web authors to design interactive sites. Although it shares many of the features and structures of the full Java language, it was developed independently. JavaScript can interact with HTML source code, enabling Web authors to spice up their sites with dynamic content. JavaScript is endorsed by a number of software companies and is an open language that anyone can use without purchasing a license. It is supported by recent browsers from Netscape and Microsoft, though Internet Explorer supports only a subset, which Microsoft calls Jscript.

APPENDIX E
CATEGORY LIST (KEYWORDS)

	Category List to be used for keyword selection
1	Agriculture and Fishing
2	Airconditioning, Refrigeration and Cold Storage
3	Automotive
4	Banking Sector
5	Business Services and Consultants
6	Chemical Plastic, Rubber & Glass Industries
7	Clothing, Textiles, Footwear, Leather & Accessories
8	Construction, Building and Related Industries
9	Décor, Household Products & Services
10	Education, Training, Research and Development
11	Energy Sector, Petroleum and Coal
12	Engineering and Scientific Related Services and Products
13	Entertainment and Leisure
14	Export/Import Agents
15	Financial and Accounting Services
16	Food, Beverage and Tobacco Products and Related Services
17	Furniture and Wood
18	Gardening
19	Health, Medical and Welfare Services and Products
20	Human Resources
21	Information Systems, Electronics and Communications Technologies
22	Insurance Sector
23	Intelligence and Defence
24	Jewellery, Arts, Crafts and Curios
25	Legal Services
26	Machinery
27	Marine Services and Products
28	Media, Marketing and Publishing
29	Metals and Metal Products
30	Mining and Quarrying
31	Pharmaceuticals, Toiletries and Cosmetics
32	Printing, Packaging and Stationery
33	Property
34	Public Services Sector
35	Safety and Security
36	Tourism, Travel and Hospitality
37	Transport
38	Water and Related Services

APPENDIX F
INSTRUCTIONS TO ANALYST

1. Open four Internet Explorer in Windows XP
2. Proceed to Google, Yahoo, Msn and Ananzi on the four individual screens
3. Copy the keyword from the excel document and paste into each search engine's 'search box'. Then hit 'search'
4. Analyse should be done on the top four returned results
5. Start with the first results and analyse it by making use of the criteria list.
6. If site should be in '.pdf' format, then please return to search service (by pressing backspace) and proceed with next result.
7. When starting to answer section 6 and 11 please right click on website and select view source.
   a. **Dublin Core** - Search code for 'dc.' Examples of the use of Dublin core is listed below:

```
<!-- Start Dublin Core - Do Not Modify Tags in This Block -->
<!-- Dublin Core Meta Tags generated by TagGen - The Meta Tag Management System -->
<meta name="DC.Title" content="Bridges: Minnesota's environmental information search">
<meta name="DC.Subject" scheme="AACR2" content="Foundations Project (Minn.)">
<meta name="DC.Subject" scheme="LIV-MN" content="Environmental information -- Minnesota">
<LINK REL=SCHEMA.DC HREF="http://www.bridges.state.mn.us/servlet/lexico">
<meta name="DC.Description" content="Search interface to Minnesota state agency online envir
<meta name="DC.Creator.PersonalName" scheme="AACR2" content="Olson, Robert">
<meta name="DC.Publisher.CorporateName" scheme="AACR2" content="Minnesota Dept. of Natural R
<meta name="DC.Contributor.Editor" scheme="AACR2" content="Quam, Eileen">
<meta name="DC.Contributor.Editor" scheme="AACR2" content="Pellinen, Brent">
<meta name="DC.Contributor.PageDesigner" scheme="AACR2" content="Wadsworth, Lynn">
<meta name="DC.Date.Creation" scheme="ISO 8601" content="1998-07-31">
<meta name="DC.Date.Modified" scheme="ISO 8601" content="2002-10-15">
<meta name="DC.Type" content="Text">
<meta name="DC.Format" scheme="HTML" content="text/html">
<meta name="DC.Identifier" scheme="URL" content="bridges.state.mn.us/index.html">
<meta name="DC.Language" scheme="ISO 639-1" content="en">
<LINK REL=SCHEMA.dc HREF="http://purl.org/dc/documents/rec-dces-19990702.htm">
<!-- End Dublin Core - Do Not Modify This Block -->
```

   b. **Title Tag** – The title tag usually appears above the <body> tag and is displayed as <title>Website's title comes here </title>. Do search (ctrl-F) for 'title', within code.
   c. **Meta-Description Tag** - The description tag usually appears above the <body> tag and is displayed as <meta name='Description' content='Website's description comes here'>. Do search (ctrl-F) for 'description', within code.
   d. **Meta-Keyword Tag** - The keyword tag usually appears above the <body> tag and is displayed as <meta name='keywords' content='Website's keywords comes here'>. Do search (ctrl-F) for 'keywords', within code.
   e. **Meta-Header Tag** - The header tag usually appears within the <body> tag and is displayed as <h1>Text that should be bold</h1> (note that the '<h1>' specify the size of the text. The tag <h6> would display the largest font). Do search (ctrl-F) for 'h1', 'h2', 'h3', 'h4', 'h5' and 'h6' within code.
   f. **Meta-Robot Tag** – The robot tag is usually displayed in one of two ways. Within the code or within a separate file (robot.txt). Within the code it would usually appear above the <body> tag and is displayed as <meta name='robots' content='specify instruction to robots'>. Do search (ctrl-F) for 'robot', within code. If not in code then type in the homepage address followed by robots.txt. Example includes: http://www.searchengineworld.com/robots.txt
   g. **Alt Tags** - The alt tag usually appears within the <body> tag and is displayed as alt = 'description of graphic'. Do search (ctrl-F) for 'alt', within code.
   h. **JavaScript** – The use of JavaScript is usually indicated with a <script> tag. One can determine if a page is making use of JavaScript by searching for 'JavaScript' or 'Script' within a webpage. Do search (ctrl-F) for 'script', within code.
8. When four results are analysed within the four search services, then please continue with next keyword.

APPENDIX G
RESULTS FROM PHASE 2

Results table (Phase 2) — rotated spreadsheet containing Pilot data with columns for Keyword No, Search Engine (Google, Yahoo, MSN, Ananzi), Ranking, and URL, evaluated across Section 1, Section 2, Section 3, Section 4, Section 5, and Section 6 with binary (0/1) scores, Sum, Percentage (YES) and Percentage (NO) rows.

	6.6.1	6.7.1	6.7.2	6.7.3	Section 7	Section 8			Section 9		Section 10		Section 11		
					7.1	8.1	8.2	8.3	9.1	9.2	10.1	10.2	11.1	11.2	11.3
8	19	18	18	0	0	0	0	8	8	23	23	21	11	4
33.33%	79.17%	75.00%	75.00%	0.00%	0.00%	0.00%	0.00%	33.33%	33.33%	95.83%	95.83%	87.50%	45.83%	16.67%
66.67%	20.83%	25.00%	25.00%	100.00%	100.00%	100.00%	100.00%	66.67%	66.67%	4.17%	4.17%	12.50%	54.17%	83.33%

151

APPENDIX H
RESULTS FROM PHASE 4

No	Company Name	Keyword	Web Address	Empl	Active	AOL	Google	Lycos	MSN	Netscape	Yahoo!	TOT
1	121 Communication Centres	Customer Care Call Centre Managers	www.121group.co.za	11-20	yes	0	0	0	0	0	0	0
2	24 Hour Spill Response	Hazardous Goods Handling	www.24hourspillresponse.co.za	1-10	yes	0	0	0	0	0	0	0
3	4CIT Software Solutions	Software Development and Database	www.4cit.co.za	1-10	yes	0	0	0	0	0	0	0
4	5* Star Car Rentals	Car Rental and Accommodation	www.carsforhire.co.za	1-10	yes	0	0	0	0	0	0	0
5	A C Screens & Shutters	Solar Screens Roll Shutters Manufacturer	www.acscreens.co.za	1-10	yes	0	0	0	0	0	0	0
6	Abrahams and Gross	Attorneys and Conveyancers	www.abgross.co.za	31-40	yes	0	0	0	0	0	0	0
7	Absa Bank Limited T/A Absa Corporate and M	Corporate Banking Services	www.absa.co.za	41-60	yes	0	0	0	0	0	0	0
8	ABSA Commercial Banks	Bank	www.absa.co.za	81-100	yes	0	0	0	0	0	0	0
9	Absolute Ablutions	Mobile Ablution Trailer Manufacturers	www.absoluteablutions.co.za	11-20	yes	0	0	0	0	0	0	0
10	Acutts	Estate Agents	www.acutts.co.za	1-10	yes	0	0	0	0	0	0	0
11	Adams & Adams	Patent and Intellectual Property Attorneys	www.adamsadams.co.za	1-10	yes	0	0	0	0	0	0	0
12	Adelphi Rent A Car	Daily Car Rental	www.adelphi.co.za	1-10	yes	0	0	0	0	0	0	0
13	Admiral Yacht Manufacturers	Yacht Builders	www.admiralyachts.co.za	1-10	yes	0	0	0	0	0	0	0
14	Advanced Source International	Marketing and Export Agent	www.advancedsource.net	1-10	yes	0	0	0	0	0	0	0
15	Aerodyne Advanced Composites	High Tech Composite Automative Components	www.aerodynetech.com	21-30	yes	0	0	0	0	0	0	0
16	Affordable Adventures Cape Town	Team Building Incentives for Conferences	www.aact.co.za	1-10	yes	0	0	0	0	0	0	0
17	Africa Strategic Asset Protection	IT Security Solution Providers	www.virtualtech.co.za	1-10	yes	0	0	0	0	0	0	0
18	African Djembe Property Finance Consultants	Property Finance Education Team Building	www.africandjembe.co.za	1-10	yes	0	0	0	0	0	0	0
19	African Maritime Services	Manufacturers Suppliers Commercial Fishing Trawling	www.africanmaritime.co.za	1-10	yes	0	0	0	0	0	0	0
20	African Ostrich Marketing	Distribution Ostrich Products	www.african-ostrich.co.za	1-10	yes	0	0	0	0	0	0	0
21	African Technology Holdings	I.T. Company	www.ath.co.za	1-10	yes	0	0	0	0	0	0	0
22	Africotta	Production Tiles Cladding Wall Tiles	www.africotta.co.za	1-10	yes	0	0	0	0	0	0	0
23	Afrique-Interlink	International Trading House	www.afrique-interlink.com	1-10	yes	0	0	0	0	0	0	0
24	A-Gas	Blender Repacker Speciality Gasses Chemicals	www.agas.com	11-20	yes	0	0	0	0	0	0	0
25	AGM Group	Buying Share Holdings into Companies	www.agmgroup.co.za	11-20	yes	0	0	0	0	0	0	0
26	Agriworld	Agriculture Project Development Consultancy	www.agriworld.co.za	1-10	yes	0	0	0	0	0	0	0
27	Air Time TV Facilities	Television Outside Broadcast Facilities	www.airtime.co.za	11-20	yes	1	0	0	0	1	1	3
28	Aircontek Panelling	Manufacturers Insulated Structures	www.aircontek.co.za	31-40	yes	0	0	0	0	0	0	0
29	Albatross Marine Security	Transportation Security Services	www.albatrossms.co.za	1-10	yes	0	0	0	0	0	0	0
30	Alex J Nagel	Real Estate Brokers	www.bayprop.com	1-10	yes	0	0	0	0	0	0	0
31	All About SA	Monthly Advertising Publication Durbanville	www.allaboutsa.co.za	1-10	yes	0	0	0	0	0	0	0
32	Allsurvey Marine	Consulting Engineers and Marine Surveyors	www.allsurvey.co.za	1-10	yes	0	0	1	0	0	1	2
33	Allweld Marine & IndustrialT/A Allweld	Welding Specialists	www.allweld.co.za	1-10	yes	0	0	0	0	0	0	0
34	Alpha Design Associates	Interior Design Retail Corporate	www.alphadesign.co.za	1-10	yes	0	0	0	0	0	0	0
35	Alternative Dive Supply	Suppliers Equipment Subsea Engineering	www.alternativdive.			0	0	0	0	0	0	0
36	Aluglass Showerflex	Manufacturing and Installation Shower Doors	www.aluglass.co.za			0	0	0	0	0	0	0
37	AME Training SA	Accounting and Financial Literacy Training	www.ametraining.co.za			0	0	0	0	0	0	0
38	Andrew Bramley Career Consultants	Career Consulting Organisations and Individuals	www.andrewbramley.co.za	1-10	yes	0	0	0	0	0	0	0
39	Andrew Weir Shipping	Travel Sales and Marketing	www.rms-st-helena.com	1-10	yes	0	0	0	0	0	0	0
40	Andy's Man's Shop	Gents Outfitters Big Men Specialists	www.andys.co.za	1-10	yes	0	0	0	0	0	0	0
41	Anna Basson Properties	Property - Sales and Rentals	www.annabasson.co.za	11-20	yes	0	0	0	0	0	0	0
42	Annenberg Real Estate	Industrial Property Consultants	www.annenberg.co.za	1-10	yes	0	0	0	0	0	0	0
43	Antwerplaw BVBA	Legal Services International Law Firm	www.antwerplaw.com	1-10	yes	0	0	0	0	0	0	0
44	Aommodation Shop	Rental Agency	www.aommodationshop.co.za	1-10	yes	0	0	0	0	0	0	0
45	Apartments and Homes	Estate Agency Self Catering Accommodation	www.capeholidays.co.za	1-10	yes	0	0	0	0	0	0	0
46	Arabella South Africa	Hotels Restaurants Tourist Related	www.arabella.co.za	11-20	yes	0	0	0	0	0	0	0
47	Aries Packaging	Printing Packaging and Display	www.ariespackaging.com	1-10	yes	0	0	0	0	0	0	0
48	arivia.kom	Information Technology and Services	www.ariel.co.za	1-10	yes	0	0	0	0	0	0	0
49	Ark Inflatables	Manufacturers Inflatable Boats	www.arkinflatables.com	1-10	yes	0	0	0	0	0	0	0
50	Armageddon Security Guarding Co.	Provider Security Guarding Services	www.arms-geddon.co.za	1-10	yes	0	0	0	0	0	0	0
51	Armilla	Manufacture and Supply Faux Jewellery	www.armilla.co.za	21-50	yes	0	0	0	0	0	0	0
52	Aromakids	Manufacture Distributor Childrens Toiletries	www.aromakids.com	1-10	yes	0	0	0	0	0	0	0
53	Aromatic Apothecary	Aromatherapy Products	www.aromatic.co.za	1-10	yes	0	0	0	0	0	0	0
54	Arthur Kaplan Jewellers	Retail Jewellers	www.arthurkaplan.co.za	1-10	yes	0	0	0	0	0	0	0
55	Artificial Studio	Web Development Studio	www.artificialstudio.com	1-10	yes	0	0	0	0	0	0	0
56	Artscape Theatre Centre	Performing Arts	www.artscape.co.za	11-20	yes	0	0	0	0	0	0	0
57	Aspen Promotions	IT Based South African Exports	www.aspenpromotions.com	1-10	yes	0	0	0	0	0	0	0
58	Associated Insurance Brokers	Insurance Brokers	www.aib.co.za	11-20	yes	0	0	0	0	0	0	0
59	Associated Octel Company	Petroleum Fuel Additives Manufacturer Supplier	www.octel-corp.com	1-10	yes	0	0	0	0	0	0	0
60	Associated Printing	Printing	www.associated-printing.co.za	41-60	yes	0	0	0	0	0	0	0
61	Astore	Distributors Plastic Pipeline Systems	www.astore.co.za	11-20	yes	0	0	0	0	0	0	0
62	Atlantic Internet Services	Internet Service Provider	www.iantic.net	1-10	yes	0	0	0	0	0	0	0
63	Auction Alliance	Auctioneers Valuers Appraisers Asset Disposal	www.auctionalliance.com	1-10	yes	0	0	0	0	0	0	0
64	Automated Office Technology	Office Equipment Retail	www.canon.co.za	11-20	yes	0	0	0	0	0	0	0
65	B E Systems	Security Systems CCTV Access Control	www.besecurity.co.za	1-10	yes	0	0	0	0	0	0	0
66	B.L. Taylor & Associates	Business Accounting and Taxation Services	www.bltglobal.co.za	1-10	yes	0	0	0	0	0	0	0
67	Badentec	Electrical Reticulation Contractor	www.badentec.com	1-10	yes	0	0	0	0	0	0	0
68	Baldwin-Unifil Filters	Manufacture and Marketing Filtration Products	www.baldwinfilter.com	61-80	yes	0	0	0	0	0	0	0
69	Barksole	Shoe Bag Repairs Key Bars	www.barksole.co.za	1-10	yes	0	0	0	0	0	0	0
70	Barnard Jacobs Mellet PCS	Stockbroking Asset Management	www.bjmdirect.com	1-10	yes	0	0	0	0	0	0	0
71	Bayview Commodities	Import and Export Commodities	www.bayviewcommodities.co.za	1-10	yes	0	0	0	0	0	0	0
72	BCC Pharmaceuticals	Pharmaceutical and General Exporter	www.bsa.com	1-10	yes	0	0	0	0	0	0	0
73	BDO Spencer Steward Chartered Aountants &	Business Advisors Accounting Auditing Taxation	www.bdo.co.za	101-150	yes	0	0	0	0	0	0	0
74	Be Safe Paramedical	Paramedical Manufacture Wholesale and Distributor	www.be-safe.co.za	1-10	yes	0	0	0	0	0	0	0
75	Berco Express	Courier Services	www.bercoexpress.co.za	101-150	yes	0	0	0	0	0	0	0
76	Bermultinational	Management Consultation Organization Development	www.Bermultinational.com	1-10	yes	0	0	0	1	0	1	2
77	Betts Llewellyn	Project and Construction Management Consultants	www.bettsprojects.co.za	1-10	yes	0	0	0	0	0	0	0
78	Big Market	Wholesalers and Distributors Fresh Produce	www.bignet.co.za	1-10	yes	0	0	0	0	0	0	0
79	Blastrite	Financial Consultants Administrators	www.blastrite.co.za	41-60	yes	0	0	0	0	0	0	0
80	Blick	Maintain Business Time Telecommunication Systems.	www.blick.com	21-30	yes	0	0	0	0	0	0	0
81	Blue Chip Institute of Technology In Associatic	I.T. Solutions in all aspects	www.jct.co.za	1-10	yes	0	0	0	0	0	0	0
82	Blue Planet Travel	Travel Agency	www.louisgroup.co.za	1-10	yes	0	0	0	0	0	0	0
83	Body Craft Products	Portable Therapy Couches Salon Equipment	www.bodycraft.co.za	1-10	yes	0	0	0	0	0	0	0
84	Boland Landmark	Estate Agents	www.bolandlandmark.co.za	1-10	yes	0	0	0	0	0	0	0
85	Bona Our Pride Tours	Tour Operator Accommodation Car Rentals	www.bonanitours.co.za	1-10	yes	0	0	0	0	0	0	0

#	Company	Description	Website	Size									Total
86	Bosman's Restaurant	Restaurant and Conference Facilities	www.granderoche.com	61-80	yes	0	0	0	0	0	0	0	0
87	Bouckaert-Soenen	Textile Printing and Badge Manufacturers	www.bs.co.za	41-60	yes	0	0	0	0	0	0	0	0
88	Bowman Gilfillan	Attorneys Notaries and Conveyancers	www.bowman.co.za	101-150	yes	0	0	0	0	0	0	0	0
89	Brainwaves Technology	Internet Service Provider	www.brainwaves.co.za	1-10	yes	0	0	0	0	0	0	0	0
90	Breakout Services	Computer Industry	www.breakout.to	1-10	yes	0	0	0	0	0	0	0	0
91	Breuer Shower Doors	Manufacturers and Distributors Shower Doors	www.duschkabine.de	1-10	yes	0	0	0	0	0	0	0	0
92	Britannia: Aidan K Movers	Movers Storage Shipping	www.britannia.co.za	1-10	yes	0	0	0	0	0	0	0	0
93	Broadspeed Logistics	Clearing and Forwarding	www.broadspeed.co.za	1-10	yes	0	0	0	0	0	0	0	0
94	Brushstrokes	Makeup School and Hair Artistry	www.brushstrokes.co.za	1-10	yes	0	0	0	0	0	0	0	0
95	BTI Branding Solutions	Manufacture of Banner Systems and Flags	www.btibranding.co.za	41-60	yes	0	0	0	0	0	0	0	0
96	Bureau Veritas	Marine Surveys ISO Accreditation	www.bureauveritas.com	1-10	yes	0	0	0	0	0	0	0	0
97	Business Junction	Essential Services Business Owners	www.bizjunction.co.za	1-10	yes	0	0	0	0	0	0	0	0
98	Business Opportunities Network (BON)	Support Service facilitating Business Linkages	www.bon.co.za	1-10	yes	0	0	0	0	0	0	0	0
99	Business Partners	Business Investment Equity and Venture	www.sbdc.co.za	41-60	yes	0	0	0	0	0	0	0	0
100	Business Systems Group	Business Solutions through Technology	www.bsg.co.uk	21-30	yes	0	0	0	0	0	0	0	0
101	Businesscapital	Internet Business Funding Directory	www.businesscapital.co.za	1-10	yes	0	0	0	0	0	0	0	0
102	Byotrol	Chemical Supplier Industry	www.byotrol.co.za	1-10	yes	0	0	0	0	0	0	0	0
103	C C S Distributors	Wholesalers Distributors Cookware	www.sdistributors.co.za	1-10	yes	0	0	0	0	0	0	0	0
104	C V M Fleet Services	Vehicle Maintenance and Servicing	www.cumfleetservices.co.za	11-20	yes	0	0	0	0	0	0	0	0
105	C.A.S.S.	IT and HR Outsourcing Solutions	www.cass.co.za	1-10	yes	0	0	0	0	0	0	0	0
106	C2M Chartered Aountants	Accounting and Bookkeeping	www.c2m.co.za	1-10	yes	0	0	0	0	0	0	0	0
107	CAE College of Audio Engineering	College of Audio Engineers	www.thestudiocae.com	1-10	yes	0	0	0	0	0	0	0	0
108	Cailcom	Organizers Cape Financial Planners Conference	www.cailcom.co.za	1-10	yes	0	0	0	0	0	0	1	1
109	Cambridge University Press	Publishing	www.cup.cam.ac.uk/	11-20	yes	0	0	0	0	0	0	0	0
110	Camdell Publishing & Software	Magazine Publishing and Computer Software	www.oti.co.za	1-10	yes	0	0	0	0	0	0	0	0
111	Campus Camera	Photographic Store and Digital Lab	www.campuscamera.co.za	1-10	yes	0	0	0	0	0	0	0	0
112	Candor Technologies	Application Provider and Web Design	www.candor.com	1-10	yes	0	0	0	0	0	0	0	0
113	Candytron	Agents Wholesalers Importers	www.candytron.co.za	1-10	yes	0	0	0	0	0	0	0	0
114	Capacity Outsourcing	Temporary Employment Service	www.cap.co.za	21-30	yes	0	0	0	0	0	0	0	0
115	Cape Ads	Publishing Classified Free Ads Paper	www.capeads.co.za	21-30	yes	0	0	0	0	0	0	0	0
116	Cape Civils T/A Cape Tennis Courts	Domestic Industrial Indoor Cleaning Machines	www.thermaxvac.com	1-10	yes	0	0	0	0	0	0	0	0
117	Cape Clean-Wall	Cleaning Buildings and Grafetti Removal	www.cleanwall.co.za	1-10	yes	0	0	0	0	0	0	0	0
118	Cape Cobra	Manufacturers Leather Handbags and Briefcases	www.capecobra.co.za	81-100	yes	0	0	0	0	0	0	0	0
119	Cape Consumers	Combined Purchasing Power Negotiate Discounts	www.capeconsumers.co.za	31-40	yes	0	0	0	0	0	0	0	0
120	Cape Heat Exchange	Manufacture Heat Exchangers	www.che.co.za	41-60	yes	0	0	0	0	0	0	0	0
121	Cape Information Technology Initiative	Initiative developing Communications Technology	www.citi.org.za	1-10	yes	0	0	0	0	0	0	0	0
122	Cape Linen	Manufacturers Hospital Hotel Restaurant Linen	www.capelinen.com	1-10	yes	0	0	1	0	0	1		2
123	Cape Mohair	Manufacturers Distributors Mohair Blankets	www.capemohair.co.za	1-10	yes	0	0	0	0	0	0	0	0
124	Cape Multimedia Systems	Software Development	www.cmms.co.za	1-10	yes	0	0	0	0	0	0	0	0
125	Cape Natural Tea Products	Processing Packing Export Rooibos Honeybush	www.rooibostea.co.za	1-10	yes	1	1	0	0	1			3
126	Cape Oil and Gas Supply Initiative	Oil and Gas Service	www.offshoreafrica.co.za	1-10	yes	0	0	0	0	0	0	0	0
127	Cape Peninsula University of Technology	Educational & Business Training	www.ctech.ac.za	1-10	yes	0	0	0	0	0	0	0	0
128	Cape Shade	Shadeports and Tensile Structures	www.CapeShade.co.za	11-20	yes	1	1	1	1	1	1		6
129	Cape Talk 567mw	Radio Station	www.capetalk.co.za	41-60	yes	0	0	0	0	0	0	0	0
130	Cape Town Flight Training Centre	Flying Training	www.cape-town-flying.co.za	1-10	yes	0	0	0	0	0	0	0	0
131	Capetel	Wireless Telecommunications	www.capetel.co.za	1-10	yes	0	0	0	0	0	0	0	0
132	Capital Alliance Life	Life Assurance	www.capitalalliance.co.za	1-10	yes	0	0	0	0	0	0	0	0
133	Capricorn	Development Technology Light Industrial Park	www.capricorn.co.za	1-10	yes	0	0	0	0	0	0	0	0
134	Care Cross Health	Primary Healthcare Development	www.carecross.co.za	1-10	yes	0	0	0	0	0	0	0	0
135	Career Vision	Graduate Recruitment and HR Consultancy	www.careervision.org.za	1-10	yes	0	0	0	0	0	0	0	0
136	Catalyst Innovation Incubator	Financial Services Venture Capital	www.catalystii.net	1-10	yes	0	0	0	0	0	0	0	0
137	CBD Freight Forwarding	Clearing Forwarding Agent Tea Producers	www.cbd.co.za	1-10	yes	0	0	0	0	0	0	0	0
138	CCI (Competitive Capabilities International)	Operations Management Consultancy	www.int.net	31-40	yes	0	0	0	0	0	0	0	0
139	CCII Systems	Specialise Defence Information Technology	www.ii.co.za	11-20	yes	0	0	0	0	0	0	0	0
140	CCW IT Consultants	IT Consultants Linux Unix Windows	www.silver.co.za	1-10	yes	0	0	0	0	0	0	0	0
141	Celerity Systems	Internet Based Applications SMS Messaging	www.bulksms.com	1-10	yes	0	0	1	0	0	0	0	1
142	Cell C Services Provider	Cellular	www.cellc.co.za	81-100	yes	0	0	0	0	0	0	0	0
143	Celsa Properties	Selling Marketing and Letting Property	www.celsa.co.za	1-10	yes	0	0	0	0	0	0	0	0
144	Centa Stop Fire	Sales Service Fire Fighting Equipment	www.stopf.centa.co.za	1-10	yes	0	0	0	0	0	0	0	0
145	Centeq	Computer Networking Internet Service Provider	www.centeq.co.za	1-10	yes	0	0	0	0	0	0	0	0
146	Century Casinos Caledon	Casino and Hotel Operator	www.caledoncasino.co.za	151-200	yes	0	0	0	0	0	0	0	0
147	Ceratex Engineering	Manufacturers Vibratory Abrasives Metal Finishing	www.ceratex.co.za	1-10	yes	0	0	0	0	0	0	0	0
148	Ceres Fruit Juices	Manufacturer 100% Pure Fruit Juice	www.ceres.com	31-40	yes	0	0	0	0	0	0	0	0
149	Cerfs Adventure Playgrounds	Manufacturers of Timber Jungle Jims	www.junglejim.co.za	1-10	yes	0	0	0	0	0	0	0	0
150	CGS Financial Technologies	Value Added Services Retailers Point Sale	www.cgs.co.za	1-10	yes	0	0	0	0	0	0	0	0
151	Chemical Investigation Services	Independent Chemical and Technical Investigators	www.chemtalk.co.za	1-10	yes	0	0	0	0	0	0	0	0
152	China South Africa Travel & Trade Solutions	Cultural Study Exchanges China SouthAfrica	www.chinasouthafrica.co.za	1-10	yes	0	0	0	0	0	0	0	0
153	Chiron Lubricants	Automotive and Industrial Lubricants	www.chironlubricants.com	1-10	yes	0	0	0	0	0	0	0	0
154	Clotex	Cape Clothing Textile Support Centre	www.clotex.co.za	1-10	yes	0	0	0	0	0	0	0	0
155	Coalition Cabling & Electrical	Computer network and Cabling Solutions	www.coalitioncabling.co.za	1-10	yes	0	0	0	0	0	0	0	0
156	Coessa Holdings	Property Developers	www.coessa.com	1-10	yes	0	0	0	0	0	0	0	0
157	Collaborative Process Integrations	Collaborative Process Integration Services	www.cpi.co.za	1-10	yes	0	0	0	0	0	0	0	0
158	Color Me Beautiful	Image Consultants	www.colormebeautiful.co.za	1-10	yes	0	0	0	0	0	0	0	0
159	Color Tone Paints	Manufacturing Paint and Related Products	www.colortone.co.za	1-10	yes	0	0	0	0	0	0	0	0
160	Computer Cabling Technologies	Distributor Networking Components	www.t.co.za	1-10	yes	0	0	0	0	0	0	0	0
161	Concretex	Manufacturers Contractors Concrete Products	www.concretex.co.za	11-20	yes	0	0	0	0	0	0	0	0
162	Cool Gels Sales & Marketing	Manufacturing Distribution of Gel Based Products	www.coolgels.co.za	1-10	yes	0	0	0	0	0	0	0	0
163	Copycat Communications	Advertising and Promotional Services	www.copycat.co.za	1-10	yes	0	0	0	0	0	0	0	0
164	Coronation Asset Management	Financial Services Unit Trusts	www.coronation.co.za	21-30	yes	0	0	0	0	0	0	0	0
165	Corporate Event Staging	Audio Visual Equipment Hire Event Staging	www.corporate-event.co.za	1-10	yes	0	0	0	0	0	0	0	0
166	Corrocoat	Anti-Corrosion Anti-Abrasion Coating Linings	www.corrocoat.com	1-10	yes	0	0	0	0	0	0	0	0
167	Courierit	Tourism Transport Freight	www.courierit.co.za	1-10	yes	0	0	0	0	0	0	0	0
168	Cresco Consultants	Consultants National Skills Development Strategy	www.cresco-con.co.za	1-10	yes	0	0	1	0	0	0	0	1
169	CRS Technologies	Payroll and HR Solutions	www.crs.co.za	11-20	yes	0	0	1	0	0	0	0	1
170	Crystal Logic	Software Solutions Provider	www.openit.co.za	1-10	yes	0	0	0	0	0	0	0	0
171	CT & A Project Management	Management Consulting and Project Management	www.tpo.co.za	1-10	yes	0	0	0	0	0	0	0	0
172	Dalvies Sweet	Sweets Manufacturer	www.dalviessweets.co.za	1-10	yes	0	0	0	0	0	0	0	0
173	Dalys & Associates	Incorporated Insurance Brokers Claims Managers	www.daly.co.za	61-80	yes	0	0	0	0	0	0	0	0
174	Damelin Boland	Training and Education	www.damelin.co.za	1-10	yes	0	0	0	0	0	0	0	0
175	DataX	Information Technology Web Design Hosting	www.datax.co.za	1-10	yes	0	0	0	0	0	0	0	0
176	Dawson Edwards and Associates	Maritime Lawyers	www.oceanlaw.co.za	1-10	yes	0	0	0	0	0	0	0	0
177	De Goede Hoop Development	Property Development and Investment	www.degoede.co.za	11-20	yes	0	0	0	0	0	0	0	0
178	Dei Gratia Training	IT Training Support Computer Supplies	www.deigratia.co.za	1-10	yes	0	0	0	0	0	0	0	0
179	Del Monte Fresh Produce	Fresh Fruit Export Company	www.freshdelmonte.co.za	1-10	yes	0	0	0	0	0	0	0	0
180	Delta Real Estate	Commercial and Industrial Property Brokers	www.factories.co.za	1-10	yes	0	0	0	0	0	0	0	0
181	Delta Vision	IT Education Locals Foreigners	www.deltavision.co.za	1-10	yes	0	0	0	0	0	0	0	0
182	Deneys Reitz	Corporate Commercial Law Firm	www.deneysreitz.co.za	61-80	yes	1	1	1	0	1	1		5
183	Designamite Incorporated	Marketing Design Development Company	www.designamite.com	1-10	yes	0	0	0	0	0	0	0	0
184	Det Norske Veritas	Shipping Services for Managing Risk	www.dnv.com	1-10	yes	1	1	1	0	1			4
185	De-Tect NDT Services	Welding Inspection Technology	www.detect.com	21-30	yes	0	0	0	0	0	0	0	0

#	Company	Description	Website	Size	Listed	C1	C2	C3	C4	C5	C6	C7	Total
186	Devland Holdings	Property Development and Investment Company	www.capeplaces.co.za	1-10	yes	0	0	0	0	0	0	0	0
187	Dial.A.Bed	Retail Bedding	www.dialabed.co.za	1-10	yes	0	0	0	0	0	0	0	0
188	Dial-A-PC Computer Rentals	Computer Rentals	www.dialapc.com	1-10	yes	0	0	0	0	0	0	0	0
189	Die Ou Pastorie	Hotel and Restaurant	www.dieoupastorie.co.za	1-10	yes	0	0	0	0	0	0	0	0
190	Die Suider Afrikaanse Sitrus Ko-operasie Bepe	Citrus Produce Organisation	www.citrussa.co.za	1-10	yes	0	0	0	0	0	0	0	0
191	Digi	Service Bureau Printshop Internet Provider	www.digi.co.za	1-10	yes	0	0	0	0	0	0	0	0
192	Distinctive Choice 358T/A Manvic Distributors	Motor Spares Keys Locksmiths	www.marvic.co.za	1-10	yes	0	0	0	0	0	0	0	0
193	District Mail & Helder Mail	Newspaper	www.helderberg.com	1-10	yes	0	0	0	0	0	0	0	0
194	Diverse Group	Manufacturers of Pediatric Nappy	www.diversegroup.co.za	1-10	yes	0	0	0	0	0	0	0	0
195	Doculam	Indentification Security Solutions	www.doculam.co.za	1-10	yes	0	0	0	0	0	0	0	0
196	Douglas Parker Associates	Market Research Consultants	www.douglasparker.co.za	1-10	yes	0	0	0	0	0	0	0	0
197	Dr Brian Jude & Associates	Training Consultancy	www.drjude.co.za	1-10	yes	0	0	0	0	0	0	0	0
198	Dr Gerntholtz	Intellectual Property Lawyers	www.gpa.co.za	1-10	yes	0	0	0	0	0	0	0	0
199	Drake & Scull Facilities Management	Facilities Management	www.drake-scull.co.za	1-10	yes	0	0	0	0	0	0	0	0
200	Drake International	Human Resources Provider	www.drakeintl.com	11-20	yes	0	0	0	0	0	0	0	0
201	Dreamline	Fund Raising Writing Editing Translations	www.i-biz21.com	1-10	yes	0	0	0	0	0	0	0	0
202	Drop Anchor Beach Lodge	Guest House Bed Breakfast Suites	www.dropanchor.co.za	1-10	yes	0	0	0	0	0	0	0	0
203	DSD Freight Management	Freight Training and Software Development	www.dsd.co.za	1-10	yes	0	0	0	0	0	0	0	0
204	Duckworth	Industrial Strength Food Flavours	www.duckworth.co.za	1-10	yes	0	0	0	0	0	0	0	0
205	Duram	Manufacturer Waterproofing Protective Non-Slip Coatings	www.duram.co.za	41-60	yes	0	0	0	0	0	0	0	0
206	Dynachem	Manufacturers Industrial Cleaning Chemicals	www.dynachem.co.za	1-10	yes	0	0	0	0	0	0	0	0
207	Dynamic Group	Logistical Security Co-Ordinators Foreign Tourists	www.dynamicgroup.co.za	1-10	yes	0	0	0	0	0	0	0	0
208	E2E Business Solutions	Software Consultants specializing CRM Accounting	www.e2e.co.za	1-10	yes	0	0	0	0	0	0	0	0
209	E2-The Exhibition Specialists	Exhibition Organisers, Managers, Consultants	www.cleantex.co.za	1-10	yes	0	0	0	0	0	0	0	0
210	E7 Productions	Web Design DTP Printing Labels	www.e7.co.za	1-10	yes	0	0	0	0	0	0	0	0
211	Eagle Technology	Manufacturers Distributors Electronic Components	www.eagle.co.za	41-60	yes	0	0	0	0	0	0	0	0
212	Early Learning Resource Unit	Research Development Agency Urban Areas	www.elru.co.za	1-10	yes	0	0	0	0	0	0	0	0
213	Easigas	Suppliers Bulk Packaged L.P.G.	www.shell.co.za	1-10	yes	0	0	0	0	0	0	0	0
214	Easy Drive	Low Fare Car Rental Broker	www.easydrive.co.za	1-10	yes	0	0	0	0	0	0	0	0
215	Easylife Kitchens	Manufacturing of Kitchen Cupboards	www.easylifekitchens.co.za	1-10	yes	0	0	0	0	0	0	0	0
216	EasyPay	Computer Transaction Switching Services	www.easypay.co.za	61-80	yes	0	0	0	0	0	0	0	0
217	Eco Blinds and Screens	Blinds and Screens	www.ecoblinds.co.za	1-10	yes	0	0	0	0	0	0	0	0
218	Eczee Products	Manufacturers Portable Shades Shelters	www.eczeeproducts.net	1-10	yes	0	0	0	0	0	0	0	0
219	EGL Eagle Global Logistics	Freight Forwarding and Clearing	www.eaglegl.com	1-10	yes	0	0	0	0	0	0	0	0
220	Electronic Tracking Systems	Importing Distributing Wireless Tracking Devices	www.etstrack.co.za	1-10	yes	0	0	0	0	0	0	0	0
221	Electrowave Cape	Specialists Offshore Marine Electronic Engineering	www.electrowavesa.com	11-20	yes	0	0	0	0	0	0	0	0
222	Emplast	Specialised Tank Lining Corrosion Protection	www.emplast.co.za	11-20	yes	0	0	0	0	0	0	0	0
223	Energy Communications	Marketing Communications Consultancy Advertising	www.mbendi.co.za/ener-g/	1-10	yes	0	0	0	0	0	0	0	0
224	Enhanced Networks	IT Sales and Support	www.enhanced.co.za	1-10	yes	0	0	0	0	0	0	0	0
225	Entech Consultants	Civil Engineering Consultants	www.entech.co.za	21-30	yes	0	0	0	0	0	0	0	0
226	Environmental Resources Management Souther	Environmental Resources Management	www.erm.com	21-30	yes	1	1	1	1	1	1	0	6
227	EOH Consulting Services Western Capet/a EOH	Web Enabled Integrated ERP Solutions	www.ct.eoh.co.za	61-80	yes	0	0	0	0	0	0	0	0
228	Epilepsy South Africa	Protective Workshop	www.epilepsy.org.za	1-10	yes	0	0	0	0	0	0	0	0
229	Equity Solutions	Business Consulting Training	www.equitysolutions.co.za	1-10	yes	0	0	0	0	0	0	0	0
230	Escrow Europe	Specialists Source Code Technical Verification	www.escroweurope.com	1-10	yes	1	1	1	0	1	0	0	4
231	Euro Appliances Cape Town	Appliance Specialist Value Added Services	www.euroappliances.co.za	1-10	yes	0	0	0	0	0	0	0	0
232	Eurocape Holdings	Property Developer	www.eurocapeinvest.com	1-10	yes	0	0	0	0	0	0	0	0
233	Event Makers	Providers Event Management Services	www.eventmakers.co.za	1-10	yes	0	0	0	0	0	0	0	0
234	Eveology Network	Effective Creative Communication Solutions	www.eveology.com	11-20	yes	0	0	0	0	0	0	0	0
235	Exel Supply Chain Services South Africa	Agents Shipping Clearing Forwarding	www.rapidttp.co.za	41-60	yes	0	0	0	0	0	0	0	0
236	Exhibition Freighting	Exhibition Freight Forwarding On-Site Handling	www.ef-gsm.com	1-10	yes	0	0	0	0	0	0	0	0
237	Expediters International	International Freight Forwarding Customs Brokerage	www.expediters.com	1-10	yes	0	0	0	0	0	0	0	0
238	Expertool	Metal Stamping General Precision Engineering	www.expertool.co.za	21-30	yes	0	0	0	0	0	0	0	0
239	Express Training	Management Training	www.expresstraining.com	1-10	yes	0	0	0	0	0	0	0	0
240	Fabrinox	Sheetmetal Component Sub-Assembly Fabricator	fabrinox.co.za	41-60	yes	0	0	0	0	0	0	0	0
241	Factor	General Contractor Electric Power Engineering	www.enf.ru	1-10	yes	0	0	0	0	0	0	0	0
242	Faculty Training Institute	Career Training Computer IT Professionals	www.fti.co.za	1-10	yes	0	0	0	0	0	0	0	0
243	Fair Consultants	Exhibition Organisers	www.fairconsultants.com	1-10	yes	0	0	0	0	0	0	0	0
244	Fastmould Specialists	Manufacture Injection Moulds Plastic Components	www.fastmould.co.za	1-10	yes	0	0	0	0	0	0	0	0
245	FedEx	International Courier	www.fedex.com	1-10	yes	1	1	1	0	1	1	0	5
246	Fedics	Catering	www.fedics.co.za	41-60	yes	0	0	0	0	0	0	0	0
247	First Impression Promotions	Imported Business Gifts Glassware Ceramics	www.yohb.com	1-10	yes	0	0	0	0	0	0	0	0
248	First Storage Concepts	Manufacture Deliver Install Shelving	www.firststorage.co.za	1-10	yes	0	0	0	0	0	0	0	0
249	Flat Rock Communications	Suppliers Communication Cellular Products Services	www.flat-rock.com	1-10	yes	0	0	0	0	0	0	0	0
250	Flieger Law Office	Law Office Domestic International	www.flieger.com	1-10	yes	0	0	0	0	0	0	0	0
251	Flip File	Manufacturers of Plastic Stationery	www.flipfile.co.za	31-40	yes	0	0	0	0	0	0	0	0
252	Folio Translation Consultants	Language Practitioners Translators Interpreters	www.folio-online.co.za	1-10	yes	0	0	0	0	0	0	0	0
253	Foodwear	Manufacturer Distributor Protective Clothing Equipment	www.foodwear.com	1-10	yes	0	0	0	0	0	0	0	0
254	Fortune Freight Services	Freight Transport Services	www.fortunefreight.co.za	1-10	yes	0	0	0	0	0	0	0	0
255	Frama Smart Mailing	Supply Servicing Mail Room Equipment	www.framaswiss.com	1-10	yes	0	0	0	0	0	0	0	0
256	Francois Terblanche CPA	Practicing Member Institute Certified Accountants	www.projectsountants.co.za	1-10	yes	0	0	0	0	0	0	0	0
257	Franki Africa	Building Civil Contractors Foundation Specialists	www.franki.co.za	21-30	yes	0	0	0	0	0	0	0	0
258	Freshgold S.A. Exports	Export Warehousing Prepacking Fresh Produce	www.freshgoldsa.co.za	21-30	yes	0	0	0	0	0	0	0	0
259	Fruits Unlimited	Fresh Produce Export Producer Exporter	www.fruitsunlimited.c.za	1-10	yes	0	0	0	0	0	0	0	0
260	Future Managers	Education Development Resource Services	www.futuremanagers.net	1-10	yes	0	0	0	0	0	0	0	0
261	G K N Sinter Metals	Powder Metallurgical Components Automotive Electrical	www.gknsintermetals.co.za	81-100	yes	0	0	0	0	0	0	0	0
262	G R Concept SA	Fashion Design Retail Event Management	www.gavinrajah.com	1-10	yes	0	0	0	0	0	0	0	0
263	G.M.P. S.A.	Laminating Encapsulating Extruded Polyesters	www.gmpsa.com	21-30	yes	0	0	0	0	0	0	0	0
264	G.R. Pharmaceuticals	Pharmaceutical Cosmetics Manufacturing	www.grpharmaceuticals.co.za	41-60	yes	0	0	0	0	0	0	0	0
265	GAC Shipping	Ships Agents Bunker Traders	www.gulfagencycompany.com	1-10	yes	0	0	0	0	0	0	0	0
266	Gaff Nabee Properties	Estate Agency Residential Commercial Properties	www.gaffnabee.co.za	1-10	yes	0	0	0	0	0	0	0	0
267	GE Security	Securtiy trader African International	www.gesecurity.com	1-10	yes	0	0	0	0	0	0	0	0
268	Gearhouse South Africa - Cape Town	Supplier Technical Equipment Event Staging	www.gearhouse.co.za	11-20	yes	0	0	1	0	0	0	0	1
269	Gemini Marine	Manufacturers all types Inflatables	www.gemini-inflatables.com	41-60	yes	0	0	0	0	0	0	0	0
270	General Management and Marketing Services	General Management Marketing Consultancy	www.managementservices.co.za	1-10	yes	0	0	1	0	0	1	0	2
271	Geos Cape Town Language Centre	English Language Tuition Foreigners	www.geosct.co.za	1-10	yes	0	0	0	0	0	0	0	0
272	Geratec	Nutritionally Focussed Catering	www.geratecsa.com	1-10	yes	0	0	0	0	0	0	0	0
273	Gifts International	Worldwide Gift Wine Distribution	www.giftsint.co.za	1-10	yes	0	0	0	0	0	0	0	0
274	Gillan & Veldhuizen Inc.	Attorneys	www.gvinc.law.za	1-10	yes	0	0	0	0	0	0	0	0
275	Glass Brick Warehouse	Glass Brick Distributor	www.glassbrickwarehouse.co.za	1-10	yes	0	0	0	0	0	0	0	0
276	Global Conferences	Organizing Co-Ordinating Supervising Large Conferences	www.globalconf.co.za	1-10	yes	0	0	0	0	0	0	0	0
277	Global Spec	Non Destructive Testing Inspection Services	www.miois.co.za	21-30	yes	0	0	0	0	0	0	0	0
278	GLS Human Performance Technology Solutions	Consulting Human Performance Skills Development	www.workcentric.com	1-10	yes	0	0	0	0	0	0	0	0
279	GMS Systems	Design Manufacture Installation Overhead Conveyors	www.gms-systems.com	11-20	yes	0	0	1	0	0	0	0	1
280	Gnu World	Web Design Promotion Hosting	www.gnuworld.com	1-10	yes	0	0	0	0	0	0	0	0
281	Gooie Hoop Onderdele	Motor Spares Accessories	www.gooiehooponderdele.co.za	1-10	yes	0	0	0	0	0	0	0	0
282	Golding Commercial Properties	Property Managers Consultants	www.pamgolding.co.za/comm	21-30	yes	0	0	0	0	0	0	0	0
283	Good Hope International Beverages	Manufacturer Exporter Fruit Juice Soy	www.ghib.com	1-10	yes	0	0	0	0	0	0	0	0
284	Gordon's Bay Security Trust	Installations Monitoring Alarm Systems	www.gbsec.co.za	21-30	yes	0	0	0	0	0	0	0	0
285	Grab-A-Student	Labour Broker	www.grabastudent.co.za	1-10	yes	0	0	0	0	0	0	0	0

286	Grandy Pies	Pie Biscuit Confectionery Manufacturer	www.grandypies.co.za	11-20	yes	0	0	0	0	0	0	0
287	Grant Thornton	Auditors Accountants Business Advisors	www.gt.co.za	101-150	yes	0	0	0	0	0	0	0
288	Graphicraft	Manufacturers Promotional Gifts Diary Producers	www.graphicraft.co.za	11-20	yes	0	0	0	0	0	0	0
289	Grenco	Industrial Refrigeration	www.grasso.co.za	81-100	yes	0	0	0	0	0	0	0
290	Grinaker-LTA	Building Contractors	www.grinaker-lta.com	151-200	yes	0	0	0	0	0	0	0
291	Grove	International Sales Centre	www.grovegroupltd.com	1-10	yes	0	0	0	0	0	0	0
292	H.S.E. Supplies	Offshore Suppliers Electrical Wholesaler Affiliated	http://www.hse.co.za	1-10	yes	0	0	0	0	0	0	0
293	H2O International	Water Treatment Systems	http://www.h20.co.za	1-10	yes	0	0	0	0	0	0	0
294	Hamrad Electronics	Dealers Electronic Spares Communication Equipment	http://www.nis.za/homepgs/hamrad/har	11-20	yes	0	0	0	0	0	0	0
295	Handi-Pak	Manufacture Blow Moulded Articles	http://www.handi-pak.co.za	1-10	yes	0	0	0	0	0	0	0
296	Hannibal Safari Equipment	Manufacturer Exporter 4x4 Accessories Tents	http://www.hannibal.co.za	1-10	yes	0	0	0	0	0	0	0
297	HBD Venture Capital	Seed Start up Venture Capital	http://www.hbd.com	1-10	yes	0	0	0	1	0	0	1
298	Heibre Auctioneering & Marketing	Auctioneering Marketing	http://www.heibre.com	1-10	yes	0	0	0	0	0	0	0
299	Heitor Projects	Electronic Development Manufacturing	http://www.heitorprojects.com	1-10	yes	0	0	0	0	0	0	0
300	Helder Water	Water Treatment Purification Dispensing	http://www.helderwater@infor4africa.c	1-10	yes	0	0	0	0	0	0	0

APPENDIX I
SCREENSHOT OF PACKAGE USED TO RECORD CRITERIA RESULTS

Keyword No	Search E	Ranking	URL	Section 1					Section 2						Section 3		Section 4		Section 5	Section 6											
				1.1	1.2.1	1.2.2	1.2.3	1.2.4	2.1.1	2.1.2	2.1.3	2.1.4	2.2	2.3	3.1	3.2	4.1	4.2	5.1	6.1.1	6.2.1	6.2.2	6.2.3	6.2.4	6.2.5	6.2.6	6.3.1	6.3.2	6.3.3	6.3.4	6.3.5
1	Google	1																														
	Google	2																														
	Google	3																														
	Google	4																														
1	Yahoo	1																														
	Yahoo	2																														
	Yahoo	3																														
	Yahoo	4																														
1	MSN	1																														
	MSN	2																														
	MSN	3																														
	MSN	4																														
1	Ananzi	1																														
	Ananzi	2																														
	Ananzi	3																														
	Ananzi	4																														
2	Google	1																														
	Google	2																														
	Google	3																														
	Google	4																														
2	Yahoo	1																														
	Yahoo	2																														
	Yahoo	3																														
	Yahoo	4																														
2	MSN	1																														
	MSN	2																														
	MSN	3																														
	MSN	4																														
2	Ananzi	1																														
	Ananzi	2																														
	Ananzi	3																														
	Ananzi	4																														

Eugene Visser (Student 1)

APPENDIX J
SHEET USED TO CONDUCT INTERVIEW

1. **Domain Names (www.sa-cycling.com)**

2. **HTML Naming Conventions or File Name**

3. **Frames**

4. **Hypertext / Anchor Text**

5. **HTML**

6. **Meta-tags**
   6.1 **Dublin Core**

   6.2 **Meta-title tag**

   6.3 **Meta-description tag**

   6.4 **Meta-keyword tag**

   6.5 **Meta-header tag**

   6.6 **Meta-robot Tag**

   6.7 **Alt Tags**

7. **Spamming**

8. **Flash**

9. **Banner advertising**

10. **Link popularity**

11. **JavaScript**

CLARIFICATION OF TERMS

Crawler

Software used by a search engine to find and retrieve webpages to include in its index. Also known as a robot, bot or crawler.

Directory

A Website that lists webpages or sites by specific categories, using human editors to manually place websites or webpages into the categories. This is commonly known as a "human-based" search engine.

e-Business

If a company is set up and deals solely online it is said to be an e-Business. For an established company, e-Business means transforming internal processes using Internet technologies.

e-Commerce

This refers to trading over the Internet. It could also refer to electronic data interchange, where purchase orders can be transmitted between computers and payments made.

Frames

An HTML technique that enables website designers to divide the browser screen into two or more sections. Each section or frame, is a single webpage. Frames are often used to centralize navigation controls on a webpage.

HTML (Hypertext Markup Language)

The coded format language used for creating hypertext documents on the World Wide Web and controlling how the Web pages appear to the user.

Hypertext link

A word, or set of words, linked to a different location on the same webpage, or a different webpage on the same website, or a webpage hosted elsewhere. By clicking on a hyperlink, users are transported to the destination location.

Index

A searchable database of words pointing to documents created by search engine software.

JavaScript

An open-source scripting language developed by Netscape that enables web designers to create more animated and dynamic webpages.

Keyword

A single word or phrase typed into a search engine query. In a different context it may also be a single word that accurately describes the contents of a single webpage or website.

Meta tag

An HTML tag, placed between the <head> and </head> tags, that supplies information about the content of a webpage, such as what HTML specifications a webpage follows, or description of a webpage's content. A Meta tag however, does not effect how a webpage is displayed on a browser.

Optimization

The process of designing, writing, coding, and submitting webpages to search engines to increase the probability that webpages will appear at the top of search engine queries for selected keywords or key phrases.

SMME

SMME is the abbreviation for Small, Medium and Micro Enterprises and can be defined as:

Small Companies   *- Employing fewer than 50 people.*

- Total annual turnover is fewer than R10m.

- Total cross asset value is fewer than R3.75m.

Medium   *- Employing fewer than 200 people.*

- Total annual turnover is fewer than R40m.

- Total cross asset value is fewer than R15m.

Micro   *- Employing fewer than five people*

- Total annual turnover is fewer than R0.15m

- Total cross asset value is fewer than R0.10m

Spamming

Spam in search engine context differs from traditional unsolicited e-mail spam. Search engine spam is the use of any search engine ranking technique which manipulates the quality of the results produced by the search engines. Examples of spamming include excessive repetition of a keyword in a page, optimizing a page for a keyword which is unrelated to the contents of the site, using invisible text, etc. Most search engines will penalize a page which uses spamming.

Traffic

The number of unique visitors to a single webpage.

URL (Uniform Resource Locator)

The address which identifies the location of a Web page on the World Wide Web.