

**THE EFFECT WEBPAGE BODY KEYWORD
LOCATION HAS ON RANKING IN SEARCH ENGINE
RESULTS: AN EMPIRICAL STUDY**

by

Wouter Thomas Kritzinger

DISSERTATION

Submitted in partial (50%) fulfilment
of the requirements for the degree

MAGISTER TECHNOLOGIAE

in

BUSINESS INFORMATION SYSTEMS

in the

FACULTY OF BUSINESS INFORMATICS

at the

CAPE PENINSULA UNIVERSITY OF TECHNOLOGY

Supervisor: Prof M Weideman

November 2005

DECLARATION

I, the undersigned, hereby declare that the work done towards this qualification has been my own work, except where stated otherwise. All references used were accurately recorded, and any opinions expressed herein are my own and not necessarily those of the Cape Peninsula University of Technology. This dissertation has not been submitted before for any degree or assessment at any other university.

STUDENT:

Signature: _____

W.T. Kritzinger

Date: _____

ACKNOWLEDGEMENTS

For the development of this research output the author would like to thank the following, without whose help this project would not have been possible.

- My Creator for giving me the staying power.
- My wife Eleanor, for her understanding, love and support all the way.
- To my parents, Wouter and Marina for their constant love, support and financial assistance.
- To my supervisor, Prof. Weideman for his guidance, patience and understanding during busy times.
- Prof. Watkins for his assistance in the proofreading of this document.
- To the Cape Peninsula University of Technology and National Research Foundation for financial assistance.

ABSTRACT

The growth of the World Wide Web has spawned a wide collection of new information sources, which has also left users with the daunting task of determining which sources are valid. Most users rely on the web because of the low cost of information retrieval. Other advantages of the web include the convenience in terms of time and access as well as the ability to easily record results.

It is also claimed that the web has evolved into a powerful business tool. Examples include highly popular business services such as Amazon.com and Kalahari.net. It is estimated that around 80% of users utilise search engines to locate information on the Internet. This of course places emphasis on the underlying importance of webpages being listed on search engines indices.

It is in the interest of any company to pursue a strategy for ensuring a high search engine ranking for their e-Commerce website. This will result in more visits from users and possibly more sales. One of the strategies for ensuring a high search engine ranking is the placement of keywords in the body text section of a webpage. Empirical evidence that the placement of keywords in certain areas of the body text will have an influence on the websites' visibility to search engines could not be found. The author set out to prove or disprove that keywords in the body text of a webpage will have a measurable effect on the visibility of a website to search engine crawlers.

From the findings of this research it will be possible to create a guide for e-Commerce website authors on the usage, placing and density of keywords within their websites. This guide, although it will only focus on one aspect of search engine visibility, could help e-Commerce websites to attract more visitors and to become more profitable.

RESEARCH OUTPUTS

Research outputs on related topics produced by the author during this study are listed below.

JOURNAL ARTICLES

Authors	Title	Journal	Status
Kritzinger, W.T.	Keywords in webpage body text to increase website visibility to search engines.	SAJIM	To be submitted December 2005
Lubbe, S., de Kock, D., Kritzinger, W.T.	Software Piracy: Some aspects for SA managers to keep in mind.	South African Journal of Economic and Management Science, Vol 6, No 4.	Published 2003, pp 785 – 801.

CONFERENCE PAPERS

Authors	Title	Institution / Event	Status
Kritzinger, W.T., Weideman, M.	A study on the correct usage of webpage keywords to improve search engine ranking	7 th Annual Conference on WWW applications (WWW2005)	Published in proceedings of WWW2004 (http://www.zaw3.co.za)
Kritzinger, W.T., Weideman, M.	The role keyword location plays in website visibility to search engines with special reference to KEWL, KRIS and HictE.	HictE mini-conference (Carnegie of New York Project): Innovation, Collaboration and Transformation in Higher Education: an ICT perspective.	Published in proceedings of the first HictE Carnegie mini conference (2004).
Kritzinger, W.T., Weideman, M.	The role keyword location plays in website visibility to search engines: an empirical study.	6 th Annual Conference on WWW applications (WWW2004)	Published in proceedings of WWW2004 (http://www.zaw3.co.za)
Weideman, M., Kritzinger, W.T.	Knowledge Repository interfacing: hyperlinks versus relationship-based links.	HictE mini-conference (Carnegie of New York Project): Innovation, Collaboration and Transformation in Higher Education: an ICT perspective.	Published in proceedings of the first HictE Carnegie mini conference (2004).
Weideman, M., Kritzinger, W.T.	Concept mapping vs. web page hyperlinks as an information retrieval interface – preferences of postgraduate culturally diverse learners.	South African Institute of Computer Scientists and Information Technologists (SAICSIT): IT Research in Developing Countries.	Published in proceedings of SAICSIT 2003.
Weideman, M., Kritzinger, W.T.	Search Engine Information Retrieval: empirical research on the usage of Metatags to enhance website visibility and ranking of e-Commerce websites.	7 th World Multiconference on Systemics, Cybernetics and Informatics (SCI2003), Information Systems, Technologies and Applications.	Published in Proceedings of SCI2003.

BOOKS

Authors	Title	Publication details	Status
Weideman, M., Kritzinger, W.T., Mbikiwa, F., Chambers, R. (Editors).	ICT Research Forum	Cape Town, South Africa, 01 May 2005, ISSN: 1814-9812	Published
Weideman, M., Kritzinger, W.T., Theart, B. (Editors).	Proceedings of The first HictE Mini-Conference (Carnegie Corporation of New York)	Bellville, South Africa, 19 October 2004.	Published

POSTERS

Authors	Title	Institution / Event	Status
Kritzinger, W.T., Weideman, M.	Website Visibility: As you seen?	6 th Annual Conference on WWW Applications (WWW2004)	Published in Proceedings of WWW2004 (http://www.zaw3 .co.za)

WORKING PAPERS

Authors	Title	Institution / Event	Status
Weideman, M., Kritzinger, W.T.	Concept Mapping – a proposed theoretical model for implementation as a knowledge repository.	Carnegie Corporation Project (2003)	Published on http://www.uwc.a c.za/ems/is/hicte

SUPERVISOR:

Signature: _____

Date: _____

Prof. M. Weideman

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

1.1	INTRODUCTION	1
1.2	RESEARCH PROBLEM	2
1.2.1	Background	2
1.2.2	Statement of problem	3
1.2.3	Research hypotheses	4
1.3	METHODOLOGY	4
1.4	RESULTS & CONCLUSIONS	6
1.5	LIMITATIONS	6
1.6	SUMMARY	7

CHAPTER 2 – LITERATURE REVIEW

2.1	INTRODUCTION	8
2.2	WWW AND BUSINESS	11
2.2.1	Publicity, marketing & advertising	12
2.2.2	Direct online selling	13
2.2.3	Research & Development	13
2.2.4	Communication	13
2.2.5	Collaboration	14
2.3	SEARCH ENGINES	14
2.3.1	Introduction	14
2.3.2	Indexing	15
2.3.3	Ranking	16
2.4	WEBSITES	18
2.4.1	HTML Coding	18
2.4.2	Metatags	20
2.4.2.1	Introduction	20
2.4.2.2	Metadata	21

2.4.2.3	Metatag studies	22
2.4.2.4	Optimizing warning	23
2.5	KEYWORDS	23
2.5.1	Webpage title	27
2.5.2	Full-text	27
2.5.3	Webpage title and full-text	28
2.5.3	Design features	28
2.6	WEBSITE VISIBILITY	29
2.6.1	Introduction	29
2.6.2	Contributing factors	29
2.6.2.1	Internal factors	31
2.6.2.2	External factors	31
2.6.3	Search engine optimization	31
2.6.3.1	Strategies & guidelines	32
2.7	SPAM	33
2.8	IMPROVING WEBSITE VISIBILITY	34
2.8.1	Search Engine Optimisation	34
2.8.1.1	Irrelevant keywords.....	35
2.8.1.2	Keyword tag density	35
2.8.1.3	Keyword repetition	36
2.8.1.4	Doorway pages	36
2.8.1.5	Hidden text	36
2.8.1.6	Graphics	37
2.8.1.7	NOFRAME tag	37
2.8.1.8	Page Cloaking	37
2.8.1.9	Automatic submission tools	38
2.8.1.10	Submitting frequency	38
2.8.1.11	SEO time investment	38
2.8.2	Paid Placement	39

CHAPTER 3 – METHODOLOGY

3.1	INTRODUCTION	41
------------	---------------------------	-----------

3.2	RESEARCH HYPOTHESES	41
3.3	RESEARCH METHODS	42
3.3.1	Quantitative Research	42
3.3.1.1	Exploratory Research	42
3.3.1.2	Descriptive Research	43
3.3.1.3	Experimental Research	43
	I. Empirical field/natural experimental	43
3.3.2	Experiment Details	44
3.3.2.1	Search Engines	44
3.3.2.2	WebPosition Gold 2	45
3.3.2.3	Spearman Rank Correlation	45
3.3.2.4	Test Results	46
3.4	SUMMARY	46

CHAPTER 4 – RESULTS & ANALYSIS

4.1	RESULTS & ANALYSIS	47
4.1.1	First Experiment	47
4.1.2	Second Experiment	49

CHAPTER 5 – CONCLUSION

5.1	INTRODUCTION	53
5.2	SIGNIFICAANCE OF THE STUDY	53
5.3	FURTHER RESEARCH	54
5.4	CONCLUSION	54

REFERENCES

REFERENCES	59
-------------------------	----

TABLES

Table 4.1: The statistical results of the first experiment	48
Table 4.2: The statistical results of the second experiment	52

APPENDICES

Appendix A – Google Keyword Position Results – 1st Experiment ...	66
Appendix B – Yahoo! Keyword Position Results – 1st Experiment ..	67
Appendix C – MSN Keyword Position Results – 1st Experiment	68
Appendix D – AskJeeves Keyword Position Results	
1st Experiment	69
Appendix E – Yahoo! Keyword Position Results – 2nd Experiment ..	70
Appendix F – AskJeeves Keyword Position Results	
2nd Experiment	71
Appendix G – AltaVista Keyword Position Results	
2nd Experiment	72
Appendix H – MSN Keyword Position Results – 2nd Experiment	73
Appendix I – AOL Keyword Position Results – 2nd Experiment	74
Appendix J – Google Keyword Position Results – 2nd Experiment ..	75
Appendix K – Lycos Keyword Position Results – 2nd Experiment ...	76

GLOSSARY

Glossary	77
-----------------------	----

CHAPTER 1

INTRODUCTION

1.3 INTRODUCTION

The growth of the World Wide Web has resulted in the conception of a wide collection of new information sources, which has also left users with the daunting task of determining which sources are valid and which are not (Abels, White & Hahn, 1997). Most users rely on the web because of the low cost of information retrieval, unlike having to buy a book or make use of a library. Other advantages of the web include the convenience in terms of time and access as well as the ability to easily record results.

‘The web is not just about promoting one’s work, but is also the interactive exchange of information, which has now evolved into a powerful business tool (Green, 2000).’

Amazon.com and Kalahari.net are two very good examples of the web being a powerful business tool. With regards to scientific communication, the web is a key medium, with access not only to e-journals and conference proceedings, but also information of the particular researchers (need it be the individual scholars, research groups or/and departments) (Aguillo, 2000).

An increasing number of sophisticated tools have been developed to aid with the retrieval of all the information that is available on the Internet (Brinkley & Burke, 1995). Some tools that the Internet offers are commercial search engines and subject gateways; seen as two of the most important tools for locating/retrieval of information (Thelwall, 2002).

The Internet has seen a significant transformation in the mid-1990s with the development of search engines. Some authors claim that there is an estimated 1.3 billion sites on the web, each providing publicly accessible

information - over 1 million new websites are being added annually (Zhang & Dimitroff, 2005a). These search engines provides access to this overwhelmingly complex information resource.

Thelwall (2001) estimates that around 80% of users utilise search engines to locate information on the Internet. This places emphasis on the underlying importance of webpage owners being listed with search engines. An important strategy for any website owner is planning how a visitor would/could find their way to their particular site (Thelwall, 2001). A good example of this is the business strategies that a lot of large companies are using to ensure that their website obtains a high ranking in certain Google searches. Some companies opt for paid inclusion, to have their sites ranked higher than they normally would have (Smith, 2003).

‘Basically, every Internet web publisher wants good webpage visibility in search engine results so as to increase accessibility of their webpages. Unfortunately, many websites have poor visibility in search engine rankings or may not be listed at all due to various reasons’ (Zhang & Dimitroff, 2004).

website developers have to ensure that they have good rankings with search engines (Podesta, 2000). Ultimately, there can only be ten (typically) websites appearing in the top ten results page of search engines, with most of them vying for the first position. However, being listed in a search engine index is no guarantee that a user will be able to find the website (Weideman & Kritzinger, 2003).

1.4 RESEARCH PROBLEM

1.4.1 Background

According to Ritchie (2000), there are a number of reasons why dot com’s fail. Dot com (also dotcom or redundantly dot.com) companies is the collection of start-up companies selling products or services using or is

somehow related to the Internet (Wikipedia, 2005a). These reasons include a faulty business model, technology (for example, errors on the website), management, too many office parties, advertising waste and lastly the fact that users are not ready (Ritchie, 2000). According to Bazac (2002) the main reasons why dot com's fail are:

- A poor business plan.
- Poor company promotion.
- Poor financial management.
- Poor human resources management.
- Errors in the company's website.

A study have been done on the usability of e-Commerce websites (Benbunan-Fich, 2001), but literature on the visibility of these sites specifically could not be found.

These authors do not state that the usage of keyword placing and density will have an impact on whether an e-Commerce website will be successful or not. However, the literature does seem to indicate that keyword usage (including density, location, etc.) does play a part in visibility (Sullivan, 2003a).

1.4.2 Statement of problem

The research problem is the lack of empirical evidence on the effect keyword usage has on website visibility. Extensive searching has been done to determine if the lack of correct usage of keywords plays a part in the high rate of website failures. The author has found no empirical evidence thus far that the two issues are linked.

It is the intention of the author to investigate whether there is a link between the correct usage of keywords and the visibility of a website to search engines. Visibility is defined as the ranking position in a search

engine results list. The closer to the top of the search results list, the better its visibility, and vice versa.

1.4.3 Research hypotheses

H₁: The use of keywords in body text of a website does have a measurable effect on website visibility to search engines.

The literature has highlighted that keywords should be placed within the HTML Meta tag; title tag; and throughout the webpage itself so ensure good visibility to search engines. The author however did not find any empirical proof that the placement of keywords within the body text of a webpage has a measurable effect on a website's visibility to search engines. The author's aim in this dissertation is to prove or disprove this hypothesis.

1.5 METHODOLOGY

The research design that was used in conducting this Masters research was an empirical field/natural experimental design. According to Mouton (2003:157), the definition of this kind of research design is:

‘Studies that are usually quantitative in nature and which aim to provide a broad overview of a representative sample of a large population’.

Field experiments, according to Mouton (2003:157), distinguish themselves from ‘true’ or ‘classical’ experiments by the fact that they occur in natural setting (in this case, the web) rather than laboratory or artificial settings. Also, because of practical and ethical concerns, they do not involve random assignment of subjects to experimental and comparison groups (Mouton, 2003:157).

The author decided to conduct an experiment to test whether the prominence of a keyword has a measurable effect on a website's visibility to search engines. According to Nielsen, 97% of all web searches are performed on AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo! (Nielsen, 2004). The author chose AskJeeves, Google, MSN and Yahoo as tools to use in the experiment.

The author executed did a single-word search on these search engines, using the keyword 'books'. The top ten websites from the respective search engines were then visited and inspected. All the text in the 'body area' of the source code within the webpages was divided into three equal areas, namely top, middle and bottom. This was done by counting the number of characters in the 'body area' and dividing it by three. The keyword 'books' was then counted in each of the three areas and a percentage was calculated of the occurrence of the keyword in each of the three areas.

Furthermore, the author used an industry standard program to calculate the visibility percentage of each of the top ten search results in the respective search engines using the keyword 'books'. If the website had a first place ranking for the keyword 'books' then the visibility percentage will be 100%.

Since the data was not normally distributed, the best way to determine whether a relationship between two variables (e.g. keyword prominence and visibility percentage) exists was to use the Spearman Rank Correlation Coefficient.

Using the figures in Appendix A - D, a Spearman Correlation was calculated for each of the four search engine result sets. The author repeated the calculation on all possible groupings of two and three search engines. The Spearman Correlation calculated will indicate whether there is a significant relationship between the visibility percentage of the website and the location of the keywords. Finally the author grouped all four

search engines together which provided a large enough sample to calculate a Pearson Correlation, which produced an overall result.

Furthermore, to test the results the author then decided to do the same experiment again using seven search engines namely AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo!. For this second experiment the top 20 search results of each to the respective search engines were used.

1.6 RESULTS & CONCLUSIONS

As a result of this finding a general conclusion can be stated that keywords should be concentrated at the top of a webpage, and thinned towards the bottom.

It was found that for all the search engines used in this research, with exception (Yahoo!), that there were significant negative relationships between the visibility percentage and the keywords listed at the bottom area of the body text of a webpage. This is a very good indication that the ranking of websites listed with these search engines will decrease if the keywords are listed in the bottom area of the body text area.

The experiments in this research produce comparable results. This research has highlighted some definite trends in the prominence of keywords to be used on webpages, with special importance to commercial sites.

1.7 LIMITATIONS

Only keywords, keyword placing and density as website visibility determining factors to search engines were considered for this research. There are many other techniques to make websites more visible to search engines, but due the scope of this study made it impossible to test all of these techniques.

Only the 10 search engines that, according to literature, are the most popular were chosen for this research.

Only the 50 first results of the search engines were inspected. Literature suggests that most search engines only display the top 10 search results on the first page. Users tend to examine only the first page of search results and once they find a good match for their search, they tend not to look further down the list (Introna & Nissenbaun, 2000; Henzinger et al, 2002).

1.8 SUMMARY

The issue examined in the research described here is a universal one insofar as the Internet could potentially be used by anyone. The issue of site ranking within a results list is of interest to website publishers.

The aim of the research was to test to what extent keywords affect an e-Commerce website's visibility to search engines. This would include keyword placing and keyword density within an e-Commerce website.

In pursuance of this aim, the objective of the study is to provide guidelines to SMME e-Commerce website designers on effective keyword-selection, -placing and -density within the e-Commerce website.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

According to Weideman and Kritzinger (2003:231), the amount of data available on the Internet cannot be measured. New and existing webpage authors constantly add more by uploading new and revised webpages to web servers, some on an hourly basis. It is estimated that more than 1.3 billion websites are available on the Internet, and over 1 million new websites are added to it every year (Ambergreen Internet Marketing Ltd, 2005). The amount of information on the Internet continues its trend to grow exponentially according to Kobayashi & Takeda (2000) as cited by Zhang & Dimitroff (2005b).

The abundance of computing and networking techniques has made it possible for users across the world to access Internet sources and electronically publish information on the Internet. The world of the Internet was transformed with the development in the mid-1990s of search engines (Zhang & Dimitroff, 2005a:665).

Also according to Weideman & Kritzinger (2003:231) there is no central body responsible for categorising, validating or censoring data on the Internet. These authors further state that these factors contribute to the rather chaotic situation Internet users face when attempting to retrieve relevant information from the Internet (Weideman & Kritzinger, 2003:231). Internet information is characterized by the instability of its contents, decentralization of locations, multiplicity of forms, diversity of user groups, and dynamics of the environment (Zhang & Dimitroff, 2005b:691). These features make the problems of web resource discovery and description more complex and sophisticated. It is apparent that it has been an increasingly complex challenge for Internet searchers to find relevant

information in the digital universe through the use of Internet search engines (Zhang & Dimitroff, 2005b:691).

As a result of the around 80% of users utilising search engines to locate information on the Internet (Zhang & Dimitroff, 2005a:665), places emphasis on the underlying importance of webpage owners being listed with search engines. An important strategy for any website owner is planning how a visitor would/could find their way to their particular site (Thelwall, 2001:114-124).

In a similar vein, other authors claim:

‘Basically, every Internet web publisher wants good webpage visibility in search engine results so as to increase accessibility of their webpages. Unfortunately, many websites have poor visibility in search engine rankings or may not be listed at all due to various reasons’ (Zhang & Dimitroff, 2004:310-320).

Internet users can be categorized into two broad groups according to Zhang & Dimitroff (2005a:666): end users searchers and webpage publishers. The first group’s priority is to locate information on the Internet conveniently and accurately. Information browsing and information searching are two primary means. The former relies on a well-organized subject directory system while the latter rests on a search engine. Most of the time, these users prefer to employ a search engine to find relevant information. The second group’s focus is the creation of webpages and their publication on the Internet. This group’ priority is to maximize the probability that their published websites are indexed by search engines and that they appear high on searchers’ search engine results lists (Zhang & Dimitroff, 2005a:666).

The commercial potential of the web is a subject of widespread discussion, with many predictions of continuing rapid growth in the future (Simeon, 1999). A web presence for a company should mean that potential clients

using the web to search for a product or service that they provide, will see the Company's site appear whichever search engine the client uses. If there are a lot of competitors on the web, the company would like their own name to appear at the top of the list of search results.

Of great concern, therefore, is the question of how to produce a successful web intervention for a business. One useful model summarising the findings of previous research is Simeon's attracting, informing, positioning and delivering (AIPD) approach to evaluating websites (Simeon, 1999:299-300). The main assertion of the AIPD model is that in order for a firm to develop or maximize the strategic potential of its website, it should enhance and integrate the site's overall capacity to attract, inform, position, and deliver (Simeon, 1999:299).

According to Simeon (1999:299), after having established a presence on the Internet, one of the primary objectives of a new website is to attract a variety of interested parties to visit the company's website. This is being done in a variety of ways. Some of the most popular techniques include registration with search engines, providing free software, locating entertaining activities on the site, developing advertising campaigns, or establishing general information and service portals (Simeon, 1999:299).

The informing ability of the website, according to Simeon (1999:300), is also the most fundamental capability. It is possible to evaluate the extent to which an organization fulfils the informing function by examining the efforts to exchange information with key stakeholders. In addition to providing information on the company's vision, history, products, and services, many websites include information on organizational structure, financials, recruiting, executive teams and customer surveys (Simeon, 1999:300).

The positioning function relates to the ways in which the services that an organization provides help create an image or market position. Factors which can be used to examine this strategy include an analysis of website

design, transaction types, market targeting, community relations, and links to domestic and international activities (Simeon, 1999:300).

The delivery capability of the website is tightly linked to its technical infrastructure. The extent to which the web technology provides for interactivity, reliability, security, and speed, helps to determine the strategic potential of the site. Because of its inherent speed, global reach, and constant accessibility, the Internet provides unparalleled delivery flexibility (Simeon, 1999:300).

According to Thelwall, the first hurdle for any website is the first aspect: attracting visitors. 'If the site does not get visited then its content is irrelevant' (Thelwall, 2000b:150). It is clear that certain steps need to be taken to make websites visible.

2.2 WWW AND BUSINESS

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

Simeon (1999:287) claims that the Internet has become a powerful business tool. This new approach to the communication and distribution of information and services has transformed the fundamental dynamics behind many social and business interactions. The barriers and obstacles,

which often accompanied traditional commerce, are giving way to new business approaches. Consumers, producers and distributors now all have flexible, fast and inexpensive ways of participating in the market for products and services around the world (Simeon, 1999:297).

According to Podesta (2000:73), what really matters about the Internet can be summarised in one word – speed – and speed is the lifeblood of business success in the 21st century.

‘Product development cycles are measured in months; product ramp-up is measured in days; and product lifetimes are approaching that of the mosquito. The Internet can help us keep up, and that is why our industry needs to use the Internet to keep pace with the new demands for speed’ (Podesta, 2000:73).

The Internet and, more particularly according to Cockburn and Wilson (1995) the WWW, are attracting businesses in their thousands, with the main application areas being the five discussed below:

2.2.1 Publicity, marketing & advertising

The WWW appears to be an ideal medium for businesses attempting to promote themselves and their wares. Setting up a site on the WWW, and thus providing instant access to your good/services to millions of potential customers, can be achieved at a small fraction of the cost than using more conventional methods according to Watson (1994) as cited by Cockburn and Wilson (1995). According to Wikipedia (2005b) e-Commerce consists primarily of the distributing, buying, selling, **marketing**, and servicing of products or services over electronic systems such as the Internet and other computer networks. In early 1999, it was widely recognised that because of the interactive nature of the Internet, companies could gather data about prospects and customers in unprecedented amounts – through site registration questionnaires, and as part of taking orders (Hall, 1999).

2.2.2 Direct online selling

It is already possible to visit virtual malls full of virtual shops, browse through catalogues and examine various products in vast detail, all courtesy of the web. This has all been made possible by the multi-media capabilities that the web provides according to Minio (1994) as cited by Cockburn and Wilson (1995). As a place for direct retail shopping, with its 24-hour availability, a global reach, the ability to interact and provide custom information and ordering, and multimedia prospects, the web is rapidly becoming a multibillion-dollar source of revenue for the world's businesses according to Hall (1999). Online shopping can be defined as the process consumers go through to purchase products or services over the internet (Wikipedia, 2005c). Hall (1999) further states that thousands of companies that sell products to customers have discovered that the web provides not only a 24-hour-a-day showcase for their products but a quick way to reach the right people in a company for more information.

2.2.3 Research & Development

Companies, especially those involved in research and development, can use the Internet as an additional resource for collecting information. Tetzeli (1994) as cited by Cockburn and Wilson (1995) explains how it is possible to post a query on a bulletin board or join a discussion group and receive advice on how to solve the problem. Alternatively, there are millions of webpages, some of which contain access to searchable databases of information relating to particular subjects.

2.2.4 Communication

The use of low-cost electronic mail is the Internet service used most extensively by business according to Rosen (1994) as cited by Cockburn and Wilson (1995). Other forms of communication include facsimile or fax, and the emerging use of telephone calls over the Internet. Most of this according to Hall (1999) is business-to-business, with some companies

attempting to use e-mail and fax for unsolicited ads to consumers and other business prospects. Hall (1999) further states that an increasing number of business websites offer e-mail newsletters for subscribers. A new trend is 'opt-in' e-mail in which web users voluntarily sign up to receive e-mail, usually sponsored or containing ads, about product categories or other subjects they are interested in (Hall, 1999).

2.2.5 Collaboration

When links are formed between companies, it can be easy for them to communicate through the Internet. One example of this is the collaboration between IBM and Bellcore who use Internet links to share a workstation (Tetzeli, 1994 as cited by Cockburn and Wilson, 1995).

2.3 SEARCH ENGINES

2.3.1 Introduction

According to Weideman & Kritzing (2003:231-236) search engines provide the average Internet user with a (mostly) free, apparently easy way to find general information on the Internet. Search engines are used to locate information on the web, whether relevant or not (Alimohammadi, 2003: 238-242).

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

websites to include in their database (Synder & Rosenbaum, 1999:375-384).

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

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

2.3.2 Indexing

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

Search engines are intended to assist in helping people find information that is at least slightly relevant. With the estimated 1.3 billion websites available (Zhang & Dimitroff, 2004:310-320), search engines are intended to assist searchers sort through the large amount of information that is available on the Internet, playing an important role in the process of information retrieval. One author provides a very basic definition of a search engine as follows:

‘A database that contains massive amounts of data about websites.’
(Green, 2000:124-137).

This author further states that the database is compiled using a program known as a crawler, which visit sites and indexes them in a predetermined way.

The index is updated regularly either by human editors or by crawlers. Both humans and crawlers simply collect information of new websites by visiting as many websites as possible, and then building them into the index (Weideman, 2005).

2.3.3 Ranking

Some search engines allocate weights to where the keywords are located within the website, while others evaluate how many keywords appear on a webpage. The number of in-links (links to a webpage) also features, for example Google's PageRank algorithm (Weideman, 2004b:904-915). Keywords which appear in the HTML title tag are often assumed to be more relevant than others (Sullivan, 2003a).

The sheer volume of webpages available places the obligation on website owners of having to draw customers. As a result of the claimed 80% of traffic being generated by search engines, these owners often turn to these services as their major marketing route (Beal, 2003).

As the WWW expanded however, they have started falling victim to extensive criticism. Most users now feel that search engines tend to retrieve information that is totally irrelevant and contrary to what the user is looking for. They have also been criticized for the tendency to retrieve duplicates (Green, 2000:124-137). There are many different types of search engines, and the issue of what ranking criteria they use to rank websites surfaces. Most search engines have their own algorithms for this purpose - these algorithms are often trade secrets, and only some superficial information about their operation is publicly available (Synder and Rosenbaum, 1999:375-384).

Other search engines index every word on every page of a website thereby increasing the number of results retrieved, while decreasing the relevance of those results (Rowland, 1998:222-229).

Most web users employ search engines as part of at least one of their strategies to find new websites (CyberAtlas, 1999). Any website owner who wants to attract new visitors should therefore be concerned whether or not his site is registered with the major search engines.

Having a page indexed, the essential first stage of being recognized by search engines is extremely important according to Introna and Nissenbaum (2000:169-186).

'Without much exaggeration one could say that to exist is to be indexed by a search engine' (Introna and Nissenbaum, 2000: 172). If a webpage is not in the index of a search engine, a person wishing to access it must know the complete Uniform Resource Locator (URL) – also known as the webpage address (Introna and Nissenbaum, 2000:169-186).

There is a number of design and other issues that need to be addressed in order to get the page registered with search engines, states Thelwall (2000a:152). Some search engines (such as Yahoo!) have a large number of human site reviewers, and the business needs to submit the address of the site in order to have it reviewed and registered so that it can be found in a search. Other search engines such as HotBot use automated programs known as 'spiders' to trawl the web and index pages (Thelwall, 2000a:152).

Search engines create a map of the web by indexing webpages according to keywords and then build those into an enormous database that links page content to keywords to URLs. Keywords are not determined a priori by the designers of the search engines' databases nor, explicitly, by some other authority, but rather they are 'deduced' from webpages themselves

in the process of indexing (Introna and Nissenbaum, 2000:169-186). In a particular webpage a keyword can be any of the following:

- Actual keywords indicated by the webpage designer in an HTML meta tag.
- All or some of the words appearing in the title that is indicated by the HTML title tag.
- The first x number of words in a webpage.
- All the words in the webpage.

Being listed in a search engine index is, however, no guarantee that a user will be able to find the website, even if the website qualifies as a candidate for the user's search. websites that are not ranked highly are less likely to be visited (Courtois & Berry, 1999:39-46; Notess, 1999:84-86). Users tend to examine only the first page of search results and once they find a good match for their search, they tend not to look further down the list. Most search engines display only 10 of the most relevant results on the first page.

Thus, exclusion from the top 10 results means that only a small number of search engine users will actually see a link to the website (Introna & Nissenbaum, 2000:169-186; Henzinger et al, 2002). Yet another study indicated that searchers spent a relatively short amount of time searching for one topic: the average search session seemed to last between five and 10 minutes only (Cooper, 2001). It is thus clear that there is serious competition for those top ten seats.

2.4 WEBSITES

2.4.1 HTML Coding

The coding of a typical HTML webpage consists of a header and a body section. The header section contains instructions and information which is not displayed as part of the webpage, while the body coding determines

what and how the user will see the webpage. A simplified version of the HTML code of an imaginary webpage follows:

```
<HTML>
<HEAD>
  <TITLE>Wouter's Star Wars Novels Site</TITLE>
</HEAD>
<BODY>
  <h1 align='center'>
    Welcome to Wouter's Star Wars Novels Site!</h1>
</BODY>
</HTML>
```

It is this HTML coding which search engine crawlers inspect and index when they visit a webpage. Metatags (see next section) are located inside the header section. This research focuses on the location of keywords used in the body section of a webpage.

The body area is the main area of an HTML page which contains all the visible text and images that appear in the browser window. The body area begins with the body tag, which is coded <body>, with a closing tag of </body>, and is divided into a number of sub areas. A simplified version of the HTML code of the body section of an imaginary webpage follows:

```
<body>

<div id='down'>
  This is the first sentence a web-crawler will see.
</div >

<div id='top'>
This is the first sentence a person will see.
  </div >

</body >
```

Search engines constantly change the tags on which they place emphasis. One aspect about crawlers seldom changes; they index title tags and body text, hence the importance of placing keywords throughout the visible-body text on webpages. This guarantee that search engines can find and record all the relevant keywords (Thurow, 2003:21, 70-86).

2.4.2 Metatags

2.4.2.1 Introduction

More and more people and organizations now publish on the Internet according to Zhang & Dimitroff (2005b:692). Web publishers are a growing population of Internet users. Most Internet users employ search engines as a primary tool to tap Internet resources according to Haltley (2002) as cited by Zhang & Dimitroff (2005b:692). Every Internet web publisher wants good webpage visibility in search engine results lists so as to increase accessibility of their webpages.

website visibility and accessibility are extremely important when web publishers are targeting the typical Internet user, one who usually does not know a web address before searching. Unfortunately, many websites have poor visibility in search engine rankings or may not be listed at all due to various reasons. Note, too, that it is a challenge for Internet publishers to insure that their websites appear at the top of a search engine return list according to Zhang & Dimitroff (2005b:692).

It is widely recognized that many variables contribute to a successful and highly visible website in a search engine results list. A well-designed, content-rich, and easily navigated site is very important. However, if a website is not crawled and indexed by search engines or it is crawled and indexed but not well optimized, only a limited number of Internet searchers will likely access it. All of the effort made with regard to webpage content would be wasted according to Zhang & Dimitroff (2005b:692). In other

words, simply posting a webpage on a public domain or submitting it to hundreds of search engines does not necessarily lead to a successful presence of that website on Internet search engine results lists.

Web authors are usually not experts in information retrieval – in fact many are content specialists with only the technical skills needed to transfer content to this medium. Therefore an information retrieval standard for improving accessibility should be designed for use by web designers and publishers with varying backgrounds. The introduction of metadata may be a positive step towards such a standard (Zhang & Dimitroff, 2005b:692).

2.4.2.2 Metadata

Metadata attempts to facilitate understanding, identifying, describing, utilizing, and retrieving Internet information sources and their contents. Metadata is broadly defined as data about data. The index of a book (indicating the physical location of chapters) is one simple form of metadata. Metadata has been used over many years to assist librarians in locating books and other stored materials. A variety of schemes exist - the Dublin Core system is one noted example of a metadata scheme that has been married to web technologies (El-Sherbini and Klim, 2004:238-248). In principle, metadata provides an effective mechanism for describing and locating data that is relevant to a particular user (Burnett *et al* 1999, as cited by Zhang & Dimitroff, 2005b:693).

The HTML 'language' makes provision for the inclusion of metadata as part of the coding of a website. This is achieved through the definition of a variety of metatags, the content of which is not displayed for user viewing, but could be read by other programs (such as search engine crawlers). Including metadata in a webpage is not mandatory: it is voluntary and ad hoc in its implementation. Internet publishers can select metadata elements and embed them directly in their webpages at will. Search engines use the metadata supplied in webpages to extract keywords and

other related information for webpages and use them as index terms in their databases. Over the past several years, more and more web publishers have realized the importance of metadata and have added it to their webpages (Zhang & Dimitroff, 2005b:693). A large number of studies addressing the usage of metatags were found. Two studies by Weideman and Kritzingler focused on the use of metatags in very specific types of websites, as discussed below

2.4.2.3 Metatag studies

Firstly, South African Higher Education Institute websites were inspected. It was found that they do not employ metatags properly on their websites - only 27.3% made effective use of HTML metatags to enhance their visibility (Weideman, 2002). Secondly, a selection of e-Commerce based websites were dissected, and findings show that the TITLE, KEYWORD and DESCRIPTION metatags were used effectively 46%, 63% and 55% of the time respectively (Weideman and Kritzingler, 2003:231-236). The clear difference between metatag usage on academic and commercial websites can be ascribed to the fact that commercial ventures are more aware of the use of these tags. Educational institutes do not rely solely on their websites to draw paying clients.

In a similar study on the use of metatags in 346 Iranian websites, Alimohammadi found that 31.5% and 24.6% of these websites made use of the 'keywords' and 'description' metatags respectively (Alimohammadi, 2004:220-223).

These studies indicate that e-Commerce websites are aware of the use of these tags, but does not necessarily use them effectively. The authors of e-Commerce websites should spend more time and effort when deciding on which title, keywords and description will best describe the website. Educational institutes should also start using metatags more to ensure that potential overseas students/clients will be able to find the institute's website, thus, drawing in more foreign students/clients.

2.4.2.4 Optimizing warning

Other authors warn that the coding of 'good' metatags is not the most important factor influencing website visibility and ranking. One of them notes that a content-rich site is more important than website submission to search engines and writing metatags (Clay, 2004). Literature is starting to appear, which seems to indicate that search engines are attaching less and less value to metatags, as a result of spamming. Spending too much time and energy on metatags to the detriment of other, more important factors could be counter-productive.

There is a big difference between creating metadata tags and optimizing your site for high ranking and visibility in search engine results lists. In other words, adding metadata to a webpage is one thing and getting a satisfactory ranking position is another. It is not enough to simply add metatags and to submit a website to search engine indices and directories. Successful use of metadata to communicate meaning of information relies on users' understanding or awareness of other's interpretation of the domain and how this interpretation is reflected in the metadata statement (Brasethvik, 1998 as cited by Zhang & Dimitroff 2005b:693).

Furthermore, metadata users lack a clear picture of how metadata elements in webpage behave with respect to visibility in search engine results lists. In other words, the effect of metadata application for improving information discovery cannot be maximized until users understand the domain clearly, interpret it in the context of metadata correctly, and comprehend the behaviour of metadata elements fully (Zhang & Dimitroff, 2005b:693).

2.5 KEYWORDS

According to Sullivan, one of the main rules in a ranking algorithm involves the location and frequency of keywords on a webpage (Sullivan, 2003b).

Sullivan refers to it as the 'location/frequency method'. Pages with the search terms appearing in the HTML title tag are often assumed to be more relevant than others to the topic. Search engines will also check to see if the search keywords appear near the top of a webpage, such as in the headline or in the first few paragraphs of text. The focus of this project is on the usage of keywords in the body text of a webpage.

The search query refers to the word/s that a search engine user types into the input box of a search engine to start a search. Most search engine algorithms try to match the search query to the keywords (found on webpages) it has stored in its index. One author claims that multi-word keyword phrases could produce higher rankings in search engine results (Konia, 2002).

Keyword prominence refers to how often and where keywords appear in a webpage. One recommendation is to place important keywords at or near the start of a webpage, sentence, and in the TITLE or DESCRIPTION metatag (Wong, 2004). The placement of keyword near the start of a webpage is central to this research. The author, however, did not find any empirical results to confirm or refute the recommendation made by Wong (2004).

According to Konia (2002), keyword prominence calculations include the following:

- If a keyword appears at the beginning of an area, its prominence will be 100%.
- If a keyword appears exactly in the middle of an area, its prominence will be 50%.
- If a keyword appears at the end of the area, the prominence will be 0%.
- If a keyword appears at the beginning of the area, with a repetition of the keyword at the end of the area, the prominence will be 50%.

- If an area consists of multiple parts (such as having three headings tags on the page), all three areas are treated as a single adjoining area when prominence is calculated.

Konia also states that the keyword weight for a page area will increase if the number of keywords in the area increases or if the word count in the area decreases (Konia, 2002).

One of the foundations of an effective search engine marketing campaign is to select the best keywords that potential customers could use to find the site (Weideman and Kritzing, 2003:231-236). However, ranking in the search engines is determined by how relevant the algorithms consider a webpage to be for a particular keyword (Nobles and O'Neil, 2000). According to Meadhra (2004), the following list orders the possible keyword locations from the highest to the lowest priority:

- Domain Name.
- Page title.
- Headings (enclosed within HTML H1, H2, H3 tags).
- Body text.
- Metatags.
- Links.
- Alt text (Meadhra, 2004).

This author further claims that a page that contains multiple instances of a keyword will generally rank higher than a page on which the keyword appears only once. However, the interaction of keyword placement and repetition is one of the areas where it is almost impossible to second-guess the ranking algorithms (Meadhra, 2004). No empirical evidence could be found that indicates the value of locating keywords at given points on a webpage.

The importance of placing carefully chosen keywords in HTML body text has been stressed by many authors in the past. One author lists the following as being important issues regarding the use of keywords:

- How many times they are repeated.
- Where they appear.
- How they are positioned relative to each other.
- Which tags surround the keywords (Bowman, 2004a).

The same author also lists areas where keywords should be included as being: domain names, webpage titles, metatags, header content, body content, text links, filenames, ALT text, named anchors, ordered lists, directory names and CSS classes (Bowman, 2004b).

Suggestions are offered by other authors on how to ensure that descriptive keywords and key phrases will satisfy crawlers. Keywords such as 'shoes' and 'women's shoes' should be replaced by 'imported Italian shoes' and 'women's aerobic sneakers' respectively (Microsoft, 2004).

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

Pages with the search terms appearing in the HTML title tag, according to Sullivan, are often assumed to be more relevant than others to the topic. Search engines assume that any page relevant to the topic will mention those words right from the beginning. Accordingly, a document with a high frequency of keywords in the beginning of a document is seen as more relevant (relative to the keyword entered) than one with a low frequency

lower down in the document. However, the author did not find any empirical evidence that substantiate Sullivan's statement.

Thurrow states that all the search engines consider the words at the top of a webpage more important than the words on the rest of the webpage. How high up a keyword is on a webpage is called keyword prominence (Thurrow, 2003:71-73). The author again did not find empirical evidence to substantiate this statement by Thurrow.

These last two references are central to this research project, and the author has set out to provide evidence or contradict exactly these claims by Sullivan and Thurrow. No empirical evidence could be found in the literature to substantiate their premises.

Zhang & Dimitroff (2005a) that tested the impact of webpage content characteristics on webpage visibility in search engine results. These are some highlighted findings and suggestions from their research:

2.5.1 Webpage title

When the number of duplicated keywords in a webpage title increases, its visibility in a search engine results list increases up to three duplications. When the duplications exceed three, there is a downturn in terms of visibility performance in search engine results list. Therefore, a point of diminishing returns has been identified at four duplicated keywords (Zhang & Dimitroff, 2005a:687).

2.5.2 Full-text

As the number of duplicated keywords in the full-text of a webpage increases, the visibility in the results list of a search engine increases. No diminishing returns were found with full-text keywords (Zhang & Dimitroff, 2005a:688).

2.5.3 webpage title and full-text

webpages with keywords in both title and full-text achieved better visibility performance than the webpages with keywords only in full-texts and the webpages with keywords only in titles in light of returned position in a search engine results list. webpages with keywords only in fulltexts achieved better performance than webpages with keywords only in titles (Zhang & Dimitroff, 2005a:688).

2.5.4 Design features

There is no significant difference between the original webpage and webpages with font color changes, font case changes, font size changes, plural form changes, or adjectival changes in terms of their visibility performance. Search engines are apparently blind to design features that, while not important in terms of retrieval, are important in terms of positive affective response to webpage design (Zhang & Dimitroff, 2005a:687).

The foundation of an effective search engine marketing campaign is selecting the best keywords that potential customers use to find a website. Thus, selecting the right keywords requires research (Thurow, 2003:49-70). Some search engines are case sensitive, which means that searching for 'Chinese' might yield different search results than searching for 'chinese' (Thurow, 2003:49-70).

According to Beal, like any successful marketing strategy, it is vital to ensure that one knows your audience and how to reach them. In the same way traditional advertising agencies survey their demographic audience, search engine marketers must ensure that their search engine optimization campaign targets the correct keywords or search phrases. Target the wrong search phrase and one could end up with good search engine rankings for keywords that have no search requests (Beal, 2003).

2.6 WEBSITE VISIBILITY

2.6.1 Introduction

The issue of visibility is one that is easy for the inexperienced to ignore. An otherwise excellent website may be completely ignored because few potential customers ever find it (Thelwall, 2000a:152). Search engine Optimization (SEO), or search engine positioning, is the process of identifying factors in a webpage which would impact search engine accessibility to it and fine-tuning the many elements of a website so it can achieve the highest possible visibility when a search engine responds to a relevant query (Zhang & Dimitroff, 2005a:666).

Search engine optimization aims at achieving good search engine accessibility for webpages, high visibility in a search engine results, and improvement of the chances the webpages are retrieved. Search engine optimization is a difficult task, far more intricate and complex than one would expect, particularly since different search engines have different indexing strategies and ranking algorithms (Zhang & Dimitroff, 2005a:666).

One author has pinpointed the stark contrast between getting a website *publicized* as opposed to having it *recognized* (Guenther, 2004). Due to an overabundance of technologies and freely available programs, it has become a routine task to design and publish a basic website. However, it is not a given that large numbers of visitors will start frequenting this site. If there is no financial gain involved, this presents no problem. In many cases, however, large numbers of online visitors is a must - an e-trading website is a case in point.

2.6.2 Contributing factors

When designing webpages, the Internet web publisher's underlying intention is to draw as many visitors to the site as possible. This requires

that the site has a good ranking with top search engines (Ford *et al*, 2002:124-137). To ensure a top ranking with search engines, the webpage must have good visibility to search engine crawlers (Lowley, 2000:190-211).

According to some authors there are various factors that can contribute to the visibility of a website. These include webpage metadata structure, webpage content, hyperlink cited status, search query expansion, as well as a number of other factors (Zhang and Dimitroff, 2004:310-320) (Zhang & Dimitroff, 2005a:666). These authors also state that many websites have poor visibility or may not be listed at all due to various reasons.

A metadata system is a system used to describe a webpage for a variety of reasons. webpage content is simply determined by words on the webpage itself. Hyperlink cited status of a webpage refers primarily to the number of webpages on the Internet that hyperlink or cite to a particular webpage: the more pages hyperlink a webpage, the better the hyperlink cited status, and vice versa. Hyperlink cited status of a webpage is a variable that may affect its viability in a search return list. Since a webpage with high hyperlink status usually is considered to be more important or influential than other pages with low hyperlink status, some search engine ranking algorithms take it into consideration, making results ranking appear to be more relevant. In other words, a returned webpage with a better hyperlink cited status would be ranked higher than other returned pages (Zhang & Dimitroff, 2005a:666).

Query expansion also affects webpage visibility in a search engine from a different perspective. The Internet search process is an interactive process between a human being and a search engine. It is a complex process affected by multiple variables. During this interactive process, an initial query may be changed, modified, or revised, moving toward a more effective, and well-defined query. Some search engines monitor, analyze, and use users' query expansion information as a factor for webpage visibility calculation (Zhang & Dimitroff, 2005a:666).

These factors can be grouped into two basic categories, as discussed below.

2.6.2.1 Internal factors

The first group includes webpage metadata structure and webpage content. These factors are internal and are determined by the webpage itself. They can be obtained or parsed from a webpage. The second group includes hyperlink cited status, query expansion, and possibly others. These factors are external to the webpage and cannot be obtained from the webpage itself. The factors in the first group can be controlled and manipulated by webpage designers or developers due to their internal nature. They should be primary factors in optimizing the visibility of a webpage in a search engine results list (Zhang & Dimitroff, 2005a:667).

2.6.2.2 External factors

The factors in the second group cannot be controlled and managed by the webpage designers or developers because of their external nature. Hyperlink cited status of a webpage totally depends on whether other websites cite or hyperlink to a webpage. Query expansion relies on users' search behaviour (Zhang & Dimitroff, 2005a:667).

Obviously, a webpage designer cannot control an Internet searcher's behaviour and cannot change webpage hyperlink cited status. He/she can only control the internal factors identified in the first group (Zhang & Dimitroff, 2005a:667).

2.6.3 Search engine optimization

An increasing number of websites are turning to search engines as their primary marketing route (Centaur Communication, 2002 as cited by Zhang

& Dimitroff, 2005a:667). Driven by this trend, search engine optimization is a booming field for entrepreneurs. Many companies offer search engine optimization services to help enhance customer's online experiences by pushing relevant websites to the fore (Kanaley, 2002 as cited by Zhang & Dimitroff, 2005a:667).

These services range from free webpage optimization submission to paid optimization software to free webpage submission. Free website submission cannot assure that the submitted website will end up in a good position. On the other hand, many web publishers, especially non-profit institutes or organizations, cannot afford to pay for optimization software and pricey website submissions. A growing industry has blossomed that offers advice (for a fee in most cases) on maximizing webpage placement. This advice about which techniques will provide optimal ranking results is hinted at on the Internet itself but none of those offering advice provide details about any empirical research on which their recommendations might be based.

While a common theme among these advice givers is "location, location, location" the specific advice is fairly generic and based on conventional wisdom, not on tested hypotheses. Research on this emerging topic, on the other hand, has not been reported in research-oriented publications. Some websites offer search engine optimization tips based on their experiences while others merely provide a basic introduction to the topic (Greenberg, 2000 as cited by Zhang & Dimitroff, 2005a:668) (Sullivan, 2002b).

2.6.3.1 Strategies & guidelines

There are various ways of ensuring website visibility to search engines. Keyword location is one of them, and it is critical in ensuring that websites are visible to search engines (Nobles and O'Neil, 2000). These authors also list the following as some of the strategies that can be used as a guide to creating top-ranking pages:

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

These authors also provide the following guidelines:

- Fine-tune keywords to narrow down the competition. By fine-tuning keyword phrases, the website designer is targeting traffic to a particular webpage.
- Pair a general keyword with a more specific one.
- Consider combining keywords that are distantly related.
- If a specific keyword is commonly miss spelt, include the miss spelt version as one or more of the keywords (e.g. accommodation, acommodation, accomodation, acomodation).

2.7 SPAM

Spam in search engine context differs from traditional unsolicited e-mail spam. Search engine spam is the use of any ranking technique which confuses the search engine crawlers and manipulates the quality of the results produced.

Unfortunately, many website authors with dubious motives have found various ways to trick search engine crawlers into attaching a higher ranking value to a given website than what it deserves based on actual content. Some of these methods include:

- Repeat keywords in the body text in a way that does not resemble a normal English sentence.
- Use the same colour for text as for the background, making repetitive sequences invisible to the human visitor (but not to the crawler).
- Multiple instances of the same tag.
- Submit identical versions of the same page to search engines.
- Create doorway pages - crawler friendly, text-stuffed pages which link to the page containing the human-friendly page.
- Bait-switching techniques (cloaking) - one page is served to the crawler, and another one to the human visitor.
- Link farms - endless lists of hyperlinks to webpages, serving no purpose other than to artificially increase the in link count for those pages (Weideman, 2004b:904-915), (Sullivan, 2003a).

2.8 IMPROVING WEBSITE VISIBILITY

2.8.1 Search Engine Optimisation

Search Engine Optimisation (SEO) has been defined as:

‘The process of identifying factors in a website which could impact search engine accessibility to it and fine tuning the many elements of a website so that it can achieve the highest possible visibility when a search engine responds to a relevant query.’ (Zhang and Dimitroff, 2004:310-320).

These authors further state that search engine optimisation aims at achieving good search engine accessibility for webpages, high visibility in

search engine results and improvement of the chances that webpages are retrieved. It has become a very complex, sophisticated practice that requires constant research, practice and re-evaluation to be effective.

According to Roy (2005) when there are certain common mistakes that are often repeated during search engine optimization. He further states that if people avoid these mistakes they will avoid a lot of anguish and frustration in the long run.

2.8.1.1 Irrelevant keywords

The first step according to Roy (2005) in any SEO campaign is to choose the right keywords for which to optimize the site. If the wrong keywords are initially chosen, all the time and effort that are devoted in trying to get the site a high ranking with search engines will be wasted.

‘If you choose keywords which no one search for, or if you choose keywords which won’t bring in targeted traffic to your site, what good will the top rankings do?’ (Roy, 2005).

2.8.1.2 Keyword tag density

A lot of sites according to Roy (2005) have hundreds of keyword listed in the meta keyword tag, in the hope that by listing the keywords in the meta keyword tag, they will be able to get a high ranking for those keywords. Roy (2005) further states that contrary to popular opinion, the meta keyword tag has almost completely lost its importance as far as search engine positioning is concerned. Hence, just by listing keywords in the meta keywords tag, a website author will not be able to get a high ranking with search engines. To get a high ranking for those keywords, one needs to put the keywords in the actual body content of the site (Roy, 2005).

2.8.1.3 Keyword repetition

Another common mistake according to Roy (2005) is that designers endlessly repeat their target keywords in the body of their pages and in the meta keyword tags. Because so many designers have used this tactic in the past, the search engines keep a lookout for this and may penalize a site. According to Roy (2005) a site should have a repetition of important keywords a number of times. However, the way the keywords are placed in the pages needs to make grammatical sense. Simply repeating the keywords endlessly no longer works.

2.8.1.4 Doorway pages

Another myth prevalent among website authors according to Roy (2005) is that since the algorithm of each search engine is different, they need to create different pages for different search engines. While this sound acceptable in theory, it is counter-productive in practice. Furthermore, although the pages are meant for different engines, they will actually end up being very similar to each other. The search engines are often able to detect when a site has created such similar pages, and may penalize or even ban this site from their index (Roy, 2005). Instead of creating different pages for different search engines, create one page which is optimized for one keyword for all the search engines.

2.8.1.5 Hidden text

According to Roy (2005), hidden text is text with the same colour as the background colour of a page. Many webmasters, in order to get high rankings in the search engines, try to make their pages as keyword rich as possible. However, there is a limit to the number of keywords one can repeat in a page without making it appear odd to human visitors. This, in order to ensure that the human visitors to a page do not perceive the text to be odd, but that the page is still keyword rich, many webmasters add text with the same colour as the background colour (Roy, 2005). Search

engines, however, have long since caught up with this technique, and ignore or penalize the pages which contain such text. They may also penalize the entire site if even one of the pages in that site contains such hidden text (Roy, 2005).

2.8.1.6 Graphics

Search engines only understand text, they do not understand graphics according to Roy (2005). If a site contains lots of graphics but little text, it is unlikely to get a high ranking in the search engines.

2.8.1.7 NOFRAME tag

Many search engines do not understand frames. For sites which have used frames, these search engines will only consider what is present in the NOFRAMES tag. Yet, many webmasters according to Roy (2005) make the mistake of adding something like this to the NOFRAMES tag: 'This site uses frames, but your browser doesn't support them'. For the search engines which do not understand frames, this is all the text that they ever get to see in this site, which means that the chances of this site achieving a high ranking in these search engines are non-existent (Roy, 2005).

2.8.1.8 Page Cloaking

According to Roy (2005), page cloaking is a technique used to deliver different web pages under different circumstances. Designers generally use page cloaking for two reasons:

- In order to hide the source code of their search engine optimized pages from their competitors.
- In order to prevent human visitors from having to see a page which looks good to the search engines but does not necessarily look good to humans (Roy, 2005).

The problem with this according to Roy (2005) is that when a site uses cloaking, it prevents the search engines from being able to spider the same page that their users are going to see. If the search engines cannot do this, they can no longer be confident of providing relevant results to their users. Thus, search engines will probably ban the site from their index.

2.8.1.9 Automatic submission tools

Many website authors use an automatic submission software or service to submit their sites to the major search engines. However, the search engines do not like automatic submission tools and may ignore the pages if one uses them. Roy (2005) is of the opinion that the major search engines are simply too important for one not to spend the time to submit the site manually.

2.8.1.10 Submitting frequency

According to Roy (2005), designers often make the mistake of submitting too many pages per day to the search engines. This often results in the search engines simply ignoring many of the pages which have been submitted from that site. Ideally, designers should submit no more than one page per day to the search engines. By limiting oneself to a maximum of one page per day, designers ensure that they stay within the limits of all the search engines (Roy, 2005).

2.8.1.11 SEO time investment

Search engine optimization is a means to an end – not the end itself. The end is to increase the sales of products and services. Apart from trying to improve search engine ranking, website authors also need to spend time on all other factors which determine the success or the failure of the

website – the quality of the products and services, the quality of customer service, etc (Roy, 2005).

2.8.2 Paid Placement

It has been predicted that search engine companies would turn to micro-payments as a means of generating income, since the basic search function has been and still is free to the average user ((Green, 2000:124-137) quoting Nielsen). However, this prediction has not as yet materialised in the search engine industry. Instead, a large amount of money has been and still is generated by a variety of schemes which produce actual income through advertising.

Paid placement is the first one that comes to mind. This service is also called PFP (Pay For Placement) or PPC (Pay Per Click). It involves the payment by a website owner or webmaster to a search engine company a certain bid price to have their website listed in search engine results. This bidding is made for specific keywords or keyword phrases. Every potential visitor who then clicks on this listing generates this bid charge (normally a few cents per click) to the owner, hence the name PPC.

One major benefit of PFP is the fact that the website does not have to be optimised for keyword usage, saving expensive resources. Secondly, any website owner can have any website listed for (a) certain keyword(s), regardless of the relevance of these keywords to the content, as long as the bidding price is paid! This fact leads to some ethical questions being raised. PPC search engines include Kanoodle and Overture. Overture estimates that 40% of daily searches done generate some form of revenue (Moxley *et al*, 2004:61-65).

A second possibility is paid inclusion (PI). This service is also referred to as PFI (Pay For Inclusion). webmasters or website owners pay a set price for webpages to be included in search engine databases for a certain period. It is also guaranteed that the crawler will revisit these webpages

for refreshing within a specified time period. However, no guarantee is given that the relevant webpages will be listed on the result screen of the user. PFI search engines include AskJeeves, Inktomi and AltaVista.

PI does offer the advantage to the client that a website will be regularly revisited by the crawler, ensuring that recent changes to the website will be reflected correctly in the database. Some PI systems also provide regular reports on clickthroughs, allowing owners to trace the effectiveness of certain keywords and/or phrases. A further advantage of this system is the fact that dynamically generated webpages, to which search engine crawlers are notoriously averse, can benefit by this type of promotion.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

Through the literature study it has become clear that there is no easy solution to the problem of significantly increasing 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 lengthy and continual process.

In the following paragraphs the author states the research hypotheses, discusses some research methods and provides an overview of the experiment details.

3.2 RESEARCH HYPOTHESES

H₁: The use of keywords in body text of a website does have a measurable effect on website visibility to search engines.

Literature has highlighted that keywords should be placed within the HTML meta tag; title tag; and throughout the webpage body-text so ensure high visibility to search engines. The author however did not find any empirical evidence that the placement of keywords, more particularly, the placement of keywords within the body text of a webpage, has a measurable effect on a websites visibility to search engines. The author's aim in this dissertation is to confirm or contradict this hypothesis.

3.3 RESEARCH METHODS

3.3.1 Quantitative Research

Struwig and Stead (2001:7-8) states that quantitative research can be defined as a form of conclusive research, involving large representative samples and fairly structured data collection procedures. The emphasis of quantitative research is placed on the methodology, since it relies on the measurement and analysis of statistical data to determine relationships between entities, which could eventually culminate in quantifiable conclusions (Struwig and Stead, 2001:7-8).

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 according to Eldabi, Irani, Paul and Love (2002:65).

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

3.3.1.1 Exploratory research

Cooper and Schindler (2003:148) stated that a study may be viewed as either exploratory or formal in 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. These authors further state that formal research begins where exploratory left off by continuing to test the

hypotheses or answer the research questions posed (Cooper & Schindler, 2003:148).

3.3.1.2 Descriptive research

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

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).

3.3.1.3 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.

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.

I. Empirical field/natural experimental

The research design that will be used in conducting this Masters research will be an empirical field/natural experimental design. According to Mouton (2003:157), the definition of this kind of research design is:

'Studies that are usually quantitative in nature and which aim to provide a broad overview of a representative sample of a large population'. Field experiments, according to Mouton (2003:157), distinguish themselves from 'true' or 'classical' experiments by the fact that they occur in natural setting (in this case, the web) rather than laboratory or artificial settings. Also, because of practical and ethical concerns, they do not involve random assignment of subjects to experimental and comparison groups (Mouton, 2003:157).

3.3.2 Experiment Details

3.3.2.1 Search Engines

The author decided to conduct an experiment to test whether the prominence of a keyword has a measurable effect on a website's visibility to search engines. Before doing this research, a decision had to be taken on which search engines must be chosen to conduct this experiment on. According to Nielsen (2004), 97% of all web searches are performed on AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo! (Nielsen, 2004). Based on this information and on the author's own experience, the author chose AskJeeves, Google, MSN and Yahoo! to conduct the experiment on.

The author then did a single-word search, using the keyword 'books'. This keyword was chosen because when a search is done with this keyword most of the results listed are e-Commerce type sites. The top ten search results from the respective search engines were then visited and inspected. All the text in the 'body area' of the source code within the webpages was divided into three equal areas, namely top, middle and bottom. This was done by counting the number of characters in the 'body area' and dividing it by three. The keyword 'books' was then counted in each of the three areas and a percentage was calculated of the occurrence of the keyword in each of the three areas.

3.3.2.2 WebPosition Gold 2

Sullivan, editor of Search Engine Watch, had to following to say about WebPosition Gold:

"WebPosition is already an outstanding package, making it easy for web marketers to quickly discover how they rank in the various search engines. Now WebPosition Gold makes life even easier, by introducing a suite of utilities aimed at removing the mystery from search engine submissions. Professionals and novices alike will find tools to help them submit to search engines, check their ranks and even see the actual terms people use to find their web sites, valuable marketing data that many web marketers have no access to."
(Sullivan, 2000)

The author used WebPosition Gold 2 to calculate the visibility percentage of each of the top ten search results in the respective search engines, using the keyword 'books'. WebPosition Gold 2 is a search engine software program that optimizes web pages, submits to top search engines worldwide, and tracks search engine rankings and web site traffic. If the website had a first place ranking for the keyword 'books' the visibility percentage should be 100%. As indicated in Appendix A - D, WebPosition Gold 2 visibility percentage decreases as each respective search engine's ranking decreases. This implies that the search engines rankings agree with ratings provided by WebPosition Gold 2.

3.3.2.3 Spearman Rank Correlation

Since the data is not normally distributed, the best way to determine whether a relationship between two variables (e.g. keyword prominence and visibility percentage) exists, is to use the Spearman rank correlation coefficient.

'The Spearman's correlation analysis compares the order, rather than the numeric magnitude, of the variables and is mostly used to examine the strength of the relationship between two variables' (Hirsch & Riegelman, 1996).

Using the figures in Appendix A - D, a Spearman correlation was calculated for each of the four search engine result sets. The author repeated the calculation on groupings of two and three search engines. The Spearman Correlation calculated will indicate whether there is a significant relationship between the visibility percentage of the website and the location of the keywords. Lastly, the author grouped all four search engines together which provided a large enough sample to calculate a Pearson correlation, which produced an overall result.

3.3.2.4 Test Results

To test the results, the author then repeated the experiment using seven search engines namely Google, Yahoo!, MSN, Lycos, AskJeeves, AltaVista and AOL. These seven search engine's were mentioned by Nielsen for producing 97% of web searches (Nielsen, 2004). For this second experiment the top 20 search results of each to the respective search engines were used. See Appendix E – K, for a summary of the results.

3.4 SUMMARY

In this chapter the author gave a brief overview of the quantitative research approach and several research methods, namely; exploratory research, descriptive research and experimental research. Also, the author chose the empirical field/natural experimental research design because of the reasons as indicated in paragraph 3.3.1.3 by Mouton (2003:157). Finally, details regarding the experiments were supplied and summarized.

CHAPTER 4

4.1 RESULTS & ANALYSIS

4.1.1 First Experiment

Appendix A - D represents the results of the top 10 search listings of the search engines for the keyword 'books'. From the results it is clear that the rank of a website corresponds with the visibility percentage which WebPosition Gold 2 calculated for each of the websites for the keyword 'books'. It should also be noted that the keywords of the website with the number one rank are more or less evenly distributed throughout the webpage. Fifty percent of the time the keyword 'books' appeared at the top of the webpage; the remaining 50% is evenly distributed between the middle and bottom sections of the webpage. Two of the findings in Appendix A did not meet the instinctive expectation of the author. The websites ranked at number five and eight have all of the keywords located in the bottom and middle sections respectively. However, it must be stated that the position of the keywords in the body text of a webpage is but one of many ways of assuring high ranking in search engine results.

As seen in Appendix B, the website with the number one ranking has placed all its keywords at the top of the webpage. Also to be noted is that the website at the number 10 ranking has placed all its keywords at the bottom of the webpage. However, Yahoo! makes use of human indexers, and the placement of keywords within the body text of a webpage might have no bearing on the website's ranking. The rest of the websites all have their keywords placed more or less evenly across the three sections of their webpage. However, it would seem that the majority of the websites favours a higher percentage of keyword placements in the top section of their webpages.

In Appendix C, the number one ranked website also have all the occurrences of the keyword 'books' in the top section of the webpage as in

Appendix B. The rest of the websites all have their keywords spaced more or less evenly across the three sections of their webpage. Again, it would seem that the majority of the websites favours a higher percentage of keyword placements in the top section of their webpages.

Appendix D shows that the search engine's number one ranked website also have all the occurrences of the keyword 'books' in the top section of the webpage. The rest of the websites all have their keywords placed more or less evenly across the three sections of their webpage.

Table 4.1: The statistical results of the first experiment

Search Engine	Top	Middle	Bottom
Yahoo!	-	-	Negative
Google	-	-	-
MSN	-	-	-
AskJeeves	-	-	Negative
Yahoo! and Google	Positive	-	-
Yahoo! and MSN	-	-	-
Yahoo! and AskJeeves	Positive	-	Negative
Google and MSN	-	-	-
Google and AskJeeves	-	-	-
MSN and AskJeeves	-	-	-
Yahoo!, Google and MSN	Positive	-	-
Yahoo!, Google and AskJeeves	Positive	-	Negative
Yahoo!, MSN and AskJeeves	Positive	-	Negative
Google, MSN and AskJeeves	-	-	-
Yahoo!, Google, MSN and AskJeeves	Positive	-	-

Table 4.1 represents the statistical results of the first experiment. Since the data of Appendix A – D is not normally distributed; the best way to determine whether a relationship between two variables (e.g. keyword prominence and visibility percentage) exists is to use the Spearman rank correlation coefficient.

The Spearman's correlation analysis produces three possible answers, namely: positive, negative and inconclusive. As can be seen in Table 4.1, all the answers in the bottom area are negative and all the answers in the top area are positive. This can be interpreted that if a website designer places keywords in the top section of the webpage it will have a positive

impact on the website's visibility to search engines. Similarly, if a website designer places keywords in the bottom section of the webpage it will have a negative impact on the websites visibility to search engines.

The author then grouped the results of the search engines together, because it is assumed that a website designer would like to have his/her website rank well with all search engines and not just in one. Again all the results in the top section were positive and all the results in the bottom section were negative.

4.1.2 Second Experiment

Appendix E – K represents the results of the top 20 search listings of the search engines AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo! for the keyword 'books'. From the results it is clear that the rank of a website corresponds with the visibility percentage which WebPosition Gold 2 calculated for each of the websites for the keyword 'books'. From Appendix E it can be seen that the sites which ranked numbers 11 and 13 have most of their keyword placed at the bottom of the webpages.

However, it must be stated that the position of the keywords in the body text of a webpage is but one way of many other ways of assuring high ranking in search engine results. A result the author did not expect is the site which ranks number 15. This site had no occurrence of the keyword 'books' in the webpage. This result could be contributed to the fact the keyword in the body-text of a webpage only plays a small role in a website's visibility to search engines.

In Appendix F the author observed again that another webpage contained no occurrence of the keyword 'books'. This time the website had even a higher ranking, which was number two. The rest of the results from Appendix F appear to have an even spread of the occurrence of the keyword 'books' across the top, middle and bottom sections of the

webpages. Both websites number 15 and 16 had all the keywords in the top of the webpage.

Appendix G the author found that a site with a high ranking, in this case number five, has no occurrence of the keyword 'books' on the webpage. Website number 20 had all of its keywords in the top section of the webpage. The rest of the results from Appendix G appears to be an even spread of the occurrence of the keyword 'books' across all the sections of the webpages.

Appendix H indicates that the results are approximately the same as the previous results. However, the website that ranked number 19 has a visibility percentage of 0%. The author concluded that the website might not have been available on the Internet at the time of the experiment. Therefore it was impossible for webPositioning Gold 2 to calculate a visibility percentage. The author visited the website again a week later and it was found that the URL no longer exists on the internet.

From Appendix I, website number 17 had an interesting spread of the keyword. All of the occurrences of the keyword 'books' are located in the middle section of the webpage. Another odd result is the websites ranked numbers one and two. Here it was found that the number one ranked website's visibility percentage was 93.33% while the number two ranked website's visibility percentage was 100%. A possible reason for this could be that changes were made on the websites, but that the search engine has not yet updated its ranking.

The results that stood out in Appendix J were the websites that ranked numbers one and two, and website ranked number 16. Again, like in the previous search engine, it was found that the number two ranked website's visibility percentage was 100% and the number one ranked website's visibility percentage was 93.33%. Also, the website that ranked number 11 have most of it occurrence of the keyword 'books' at the bottom of the webpage.

Appendix K one has more or less that same results as in the previous search engines. However, one thing that stood out was the fact that there were a number of websites that had the same visibility percentage. The reason for this could be that this particular search engine uses a difference ranking algorithm than the other search engines.

Table 4.2 represents the statistical results of the second experiment. Since the data of Appendix E – K is not normally distributed; the Spearman correlation analysis was used.

The Spearman's correlation analysis gives three possible answers, namely: positive, negative and inconclusive. As can be seen in Table 4.2 all the answers in the bottom area are negative and all the answers in the top area are positive. This can be interpreted that if a website designer places keywords in the top section of the webpage it will have a positive impact on the websites visibility to search engines. The same can be said that if a website designer places keywords in the bottom section of the webpage it will have a negative impact on the websites visibility to search engines.

The author then grouped the results of the search engines together, because it is assumed that a website designer would like to have his/her website rank well with all search engines and not just in one. Again all the results in the top section were positive and all the results in the bottom section were negative.

Table 4.2: The statistical results of the second experiment

Search Engine	Top	Middle	Bottom
Yahoo!	-	-	-
Google	-	-	Negative
MSN	-	-	Negative
AskJeeves	-	-	Negative
AOL	-	-	Negative
Lycos	-	-	Negative
AltaVista	-	-	Negative
Yahoo! and Google	-	-	-
Yahoo! and MSN	-	-	-
Yahoo! and AskJeeves	-	-	-
Yahoo! and AltaVista	-	-	-
Yahoo! and AOL	-	-	-
Yahoo! and Lycos	-	-	-
Google and MSN	-	-	Negative
Google and AskJeeves	-	Negative	Negative
Google and AltaVista	-	-	Negative
Google and AOL	-	-	Negative
Google and Lycos	-	-	Negative
MSN and AskJeeves	-	-	Negative
MSN and AltaVista	Positive	-	Negative
MSN and AOL	-	-	Negative
MSN and Lycos	Positive	-	Negative
AskJeeves and AltaVista	-	-	Negative
AskJeeves and AOL	-	Positive	Negative
AskJeeves and Lycos	-	-	Negative
AltaVista and AOL	-	-	Negative
AltaVista and Lycos	Positive	-	Negative
AOL and Lycos	-	-	Negative

CHAPTER 5

CONCLUSION

5.1 INTRODUCTION

The purpose of this chapter is to draw the final conclusions from this research in this dissertation.

5.2 SIGNIFICANCE OF THE STUDY

Users tend to examine only the first page of search results and once they find a good match for their search, they tend not to look further down the list. Most search engines display only 10 of the most relevant results on the first page. And, according to Ambergreen Internet Marketing Ltd, most users usually examine only the top 10 websites in a search engine results list and only 1% of users check beyond the third page of a search engine results list (Ambergreen Internet Marketing Ltd, 2005). Thus, exclusion from the top 10 results means that only a small number of search engine users will actually see a link to the website (Introna & Nissenbaun, 2000; Henzinger et al, 2002).

It is therefore in the interest of the company to pursue a strategy for ensuring a high search engine ranking for their e-Commerce website. This will result in more visits from users and possibly more sales via the e-Commerce website. The focus of this study was on the effect of keyword usage on e-Commerce website visibility to search engines.

As such it will allow e-Commerce ventures to enhance the visibility of its website to search engines.

5.3 FURTHER RESEARCH

It is recommended that further research be undertaken using larger samples. In this paper the research was limited to the top twenty search results of only seven search engines. More search engines should be explored with more keyword search results. The use of multiple keywords could also be inspected. An initial literature survey on the apparently diminishing value of metatags as viewed by search engines should be done. This survey could be substantiated by a series of empirical studies.

5.4 CONCLUSION

It was noted that when Yahoo! and Google were grouped there was a positive significant relationship between the visibility percentage and keywords listed at the top area of the webpage. **This implies that a website's ranking will increase with Yahoo! and Google when the keywords are more densely grouped in the top area of the webpage.**

It was found that for all but one of the search engines used in this experiment, there were negative significant relationships between the visibility percentage and the keywords listed at the bottom area of the body text. The exception was Yahoo! and combinations involving Yahoo!. **This indicates that the ranking of websites listed with these combinations will decrease if the keywords are listed in the bottom area of the body text area.**

The statistical results of the first experiment are summarised in Table 4.1.

When the statistical results were viewed individually it was found that with Yahoo! and AskJeeves there was a negative significant relationship between the visibility percentage and keywords listed at the bottom area of the webpage. **This indicates that the ranking of websites listed at Yahoo! or AskJeeves will decrease if their keywords are listed in the bottom area of the webpage.**

When Yahoo! and AskJeeves were grouped it was found that there was a positive significant relationship between the visibility percentage and keywords listed at the top of the webpage and a negative significant relationship between the visibility percentage and keywords listed at the bottom of the webpage. **This indicates that the ranking of websites listed with Yahoo! and AskJeeves will increase when the keywords are listed at the top area of the webpage and that ranking will decrease when the keywords are listed at the bottom area of the webpage.**

When investigating Yahoo!, Google and MSN, a positive significant relationship exists between the visibility percentage and keywords placed at the top area of a webpage. **Thus, ranking will increase at these three search engines when the keywords appear at the top area of the webpage.**

When investigating Yahoo!, Google and AskJeeves as a group and Yahoo!, MSN and AskJeeves as a group, it was found that there is a positive significant relationship between the visibility percentage and keywords placed at the top area of the webpage and a negative significant relationship between the visibility percentage and keywords placed at the bottom of the webpage. **Thus, it can be claimed that ranking will increase when keywords are placed at the top of the webpage and ranking will decrease when keywords are placed at the bottom of the webpage.**

Lastly, a Pearson correlation was calculated with all four of these search engines grouped. It was found that there exists a positive significant relationship between the visibility percentage and keywords placed at the top of the webpage. **Thus, ranking/visibility will increase when the keywords are placed in the top area of a webpage.**

For the first experiment, it therefore appears as if the concentration of keywords should be at the top rather than the bottom of a webpage to enhance visibility to search engine crawlers.

The statistical results of the second experiment are summarised in Table 4.2.

These results indicated no effect for Yahoo!, or any other search engine that was grouped with Yahoo!. This makes it difficult to determine what effects keyword prominence on body text has on Yahoo!. One possible conclusion is that Yahoo! ignores keyword prominence, since human editors are used.

When MSN and AltaVista were grouped it was found that there was a positive significant relationship between the visibility percentage and keywords listed at the top of the body text area and a negative significant relationship between the visibility percentage and keywords listed at the bottom of the body text area. **This indicates that the ranking of websites listed with MSN and AltaVista will increase when the keywords are listed at the top area of the body text area and that the ranking will decrease when the keywords are listed at the bottom area of the body text area. The same applied for MSN and Lycos.**

When investigating AskJeeves and AOL, a significant positive relationship was evident between the visibility percentage and keywords that are listed in the middle of the body text area and a negative relationship between the visibility percentage and keywords that are listed in the bottom of the body text area. **This indicates that the ranking of websites listed with AskJeeves and AOL will increase if the keywords are placed in the middle section of the body text area, and will decrease if they are placed at the bottom section of the body text area.**

The statistical results showed a negative relationship between the visibility percentage and keywords that are listed in the bottom of the body text

area for AltaVista and Lycos, for AltaVista and AOL, for Google and MSN, Google and AltaVista, Google and AOL, Google and Lycos as well as MSN and AskJeeves. **This indicates that the ranking with these search engine combinations decreases if the keyword density is concentrated in the bottom section of the body text area.**

When experimenting with Google and AskJeeves, there was a negative relationship between visibility and keywords that are listed in the middle section of the body text, as well as a negative relationship between visibility and keywords that are listed at the bottom of the body text area.

It was found that for all the search engines used in this experiment, with Yahoo! being the only exception, that there were significant negative relationships between the visibility percentage and the keywords listed at the bottom area of the body text. **This indicates that the ranking of websites listed with these search engines will decrease if the keywords are listed in the bottom area of the body text area.** Yahoo!, when tested alone did not display any results.

As a general conclusion it can be stated that keywords should be concentrated at the top of a website, and diluted towards the bottom. website authors must apply these rules when viewing the actual HTML coding, not the website as it is rendered by the browser on the screen.

In conclusion, both experiments produced similar results. When keywords are used at the bottom area of the body text the ranking of websites will decrease. It therefore appears as if the concentration of keywords should be at the top rather than the bottom of a webpage to enhance visibility to search engine crawlers.

Designers of e-commerce based websites should pay close attention to the use of keywords on webpages. For every separate HTML page, the relevant keywords should be identified and place inside the top text areas.

Care should be taken to assure that the keyword density does not lead to spam penalties by search engine algorithms. Finally, alternative spelling of common words should also be included, to match spelling mistakes made by searchers.

It was the intention of the author to investigate whether there is a link between the correct usage of keywords and the visibility of a website to search engines. The author is of the opinion that the research hypotheses – *The use of keywords in body text of a website does have a measurable effect on website visibility to search engines* – has been successfully proven. All the results show that a websites ranking to search engines increase when the keywords are placed at the top of the webpage. Similarly, the results show that a website's ranking in search engines decreases when the keywords are placed at the bottom of the webpage. It is finally concluded that this research has highlighted some definite trends in the prominence of keywords to be used on webpages, with special importance to commercial sites.

REFERENCES

- Abels, G., White, DM., Hahn, K. 1997. Identifying user-based criteria for web pages. *Internet Research: Electronic Networking Applications and Policy*. 7(4), 252 –262.
- Aguillo, I. 2000. A new generation of tools for search, recovery and quality evaluation of World Wide web medical resources. *Journal of Management in Medicine*. 14(3/4), 240-248.
- Alimohammadi, D. 2004. Measurement of the presence of keywords and description meta-tags on a selected number of Iranian web sites. *Online Information Review*. 28(3), 220-223.
- Alimohammadi, D. 2003. Meta-tag: a means to control the process of web indexing. *Online Information Review*. 27(4), 238-242.
- Ambergreen Internet Marketing Ltd. 2005.
<http://www.ambergreenInternetmarketing.com/index.shtml> [28 September 2005].
- Beal, A. 2003. Optimizing Keywords for Search Engines.
<http://searchenginewatch.com/webmasters/article.php/2168011>.
- Benbunan-Fich, R. 2001. Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*. 39(2),151-163.
- Bowman, J.L. 2004a. Optimization with Keywords. <http://spider-food.net/keywords.html>.
- Bowman, J.L. 2004b. Search Engine Optimization and Coding.
<http://spider-food.net/keywords-c.html>.

Brinkley, M., Burke, M. 1995. Information Retrieval from the Internet: an evaluation of the tools. *Internet Research: Electronic Networking Applications and Policy*. 5(3), 3-10.

Clay, B. 2004. Search Engine Optimization.
http://www.bruceclay.com/web_rank.htm.

Cockburn, C., Wilson, T.D. 1995. Business use of the World Wide web. *Information Research*, 1(1) Available at: <http://informationr.net/ir/1-2/paper6.html>

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

Cooper, M.D. 2001. Usage patterns of a web-based library catalog. *Journal of the American Society for Information Science and Technology*, 52(2),37.

Courtois, M.P. & Berry, M.W. 1999. Results ranking in web search engines. *Online*, 39-46

CyberAtlas. 1999. [Online]. Available WWW: <http://www.cyberatlas.com>

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

El-Sherbini, M., Klim, G. 2004. Metadata and cataloguing practices. *The Electronic Library*. 22(3), 238-248.

Ford, N., Miller D., Moss N. 2002. web Search Strategies and retrieval effectiveness: an empirical study. *Journal of Documentation*. 58(1), 30-48.

- Green, D. 2000. The evolution of web Searching. *Online Information Review*. 24(2), 124-137.
- Guenther, K. 2004. Getting your website Recognized. *Online*. 28(3).
- Hall, P. 1999. The Internet as business enabler. *South African Journal of Information Management*. 1(2/3).
- Hämäläinen, M., Whinston, A. B., & Vishik, S. 1996. Electronic markets for learning: education brokerage on the Internet. *Communications of the ACM*, 39(6), 51-58.
- Henzinger, M.R., Motwani, R. & Silverstein, C. 2002. Challenges in web search engines. *SIGIR Forum*, 36(2).
- Hirsch, R.P. & Riegelman, R.K. 1996. *Statistical Operations: Analysis of Health Research Data*. Cambridge, MA: Blackwell Science.
- Introna, L. & Nissenbaum, H. 2000. Shaping the web: Why the Politics of Search Engines Matters. *Information Society*, 16(3):169-186.
- Konia, B.S. 2002. *Search Engine Optimization with webPosition Gold 2*. Wordware Publishing Inc, Plano, Texas.
- Lowley, S. 2000. The evaluation of WWW search engines. *Journal of Documentation*. 56(2), 190-211.
- Meadhra, M. 2004. How keywords impact search engine rankings. <http://www.builder.com>.
- Microsoft. 2004. Optimizing for the correct Keywords. http://www.submit-it.com/subopt.htm?tipq=2&sell_page=1.

Mouton, J. 2003. *How to succeed in your Master's & Doctoral Studies: A South African guide and resource book*. Van Schaik Publishers, Pretoria, South Africa.

Moxley, D., Blake, J., Maze, S. 2004. web search engine advertising practices and their effect on library service. *The Bottom line: Managing Library Finance*. 17(2), 61-65.

Nielsen, J. 2004. Statistics for Traffic Referred by Search Engines and Navigation Directories to Useit. <http://www.useit.com>.

Nobles, R., O'Neil, S. 2000. *Maximize web Site Traffic: Build web Site Traffic fast and free by optimizing search engine placement*. Adams Media Corporation, Holbrook, MA.

Notess, G.R. 1999. Rising relevance in search engines. *Online*, May/June, 23(3):84-86.

Podesta, G. 2000. e-Commerce: Helping Customers Gain the Competitive Edge. *Plastics Engineering*, July:73-74.

Ritchie, K. 2000. Reasons why dot coms fail: how to avoid the pitfalls of e-Business. <http://www.ktb.net> [accessed 21 November 2003].

Rowland, G. 1998. Getting more from the Internet. *New Library World*. 99(6), 222-229.

Roy, S. 2005. Top 10 Search Engine Positioning Mistakes. *Search Engine Guide*. <http://www.searchengineguide.com>

Simeon, R. 1999. Evaluating Domestic and international web-site strategies. *Internet Research: Electronic Networking Applications and Policy*. 9(4), 297-308.

Singh, A.M. 2002. The Internet – strategies for optimal utilization in South Africa. *South African Journal of Information Management*. 4(1).

Smith, A. 2003. Think local, search global? Comparing search engines for searching geographically specific information. *Online Information Review*, 27(2), 102-109.

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

Synder, H, Rosenbaum, H. 1999. Can Search Engines be used as tools for web-Link Analysis? A Critical Review. *Journal of Documentation*. 55(4), 375-384.

Sullivan, D. 2003a. Ending the debate over cloaking.
http://www.searchenginewatch.com/sereport/print.php/34721_2165231.

Sullivan, D. 2003b. How Search Engines rank web pages.
<http://www.searchenginewatch.com/webmasters/article.php/2167961>.

Sullivan, D. 2002a. Search engine placement tips.
<http://www.searchenginewatch.com/webmasters/article.php/2168021>.

Sullivan, D. 2002b. Intro to Search Engine Optimization.
<http://searchenginewatch.com/webmasters/article.php/2167921>

Sullivan, D. 2000. WebPosition Gold.
<http://www.lambdesign.co.uk/wpgold/product-reviews-dlr-LambDesign.htm>

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

Thelwall, M. 2000b. Commercial web sites: lost in cyberspace? *Internet Research: Electronic Networking Applications and Policy*, 10(2):150-159.

Thelwall, M. 2001. Commercial web site links. *Internet Research: Electronic Networking Applications and Policy*, 11(2):114-124.

Thelwall, M. 2002. Subject gateway sites and search engine ranking. *Online information Review*. 26(2), 101 – 107.

Thurrow, S. 2003. Search engine visibility. New Riders, IND,

Tyner, R. 2001. Sink or Swim: Internet Search Tools and Techniques. <http://www.ouc.bc.ca/libr/connect96/search.htm>.

Weideman, M. 2005. [Online]. Available WWW: <http://www.mwe.co.za/seaarticles.htm> (Accessed 13 April 2005).

Weideman, M. 2004a. Empirical evaluation of one of the relationships between the user, search engines, metadata and web sites in three-letter .com web sites. *South African Journal of Information Management*. 6(3).

Weideman, M. 2004b. Ethical issues on content distribution to digital consumers via paid placements as opposed to website visibility in search engine results. *In proceedings of ETHICOMP* (Syros, Greece 14-16 April), Volume 2, 904-915.

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

Weideman, M., Kritzing, W. 2003. Search Engine Information Retrieval: Empirical Research On The Usage Of Metatags To Enhance website Visibility And Ranking Of E-Commerce websites. *In proceedings of The*

7th World Conference on Systemics, Cybernetics and Informatics
(Orlando, Florida, USA, 28-30 July), Volume VI, 231-236.

Wikipedia. 2005a. Dot-Com. http://en.wikipedia.org/wiki/Dot_com

Wikipedia. 2005b. Electronic commerce.
http://en.wikipedia.org/wiki/Electronic_commerce

Wikipedia. 2005c. Online shop.
http://en.wikipedia.org/wiki/Online_shopping

Wong, M. 2004. Mike's marketing tools. <http://www.Mikes-Marketing-Tools.com>.

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

Zhang, J., Dimitroff, A. 2005b. The impact of metadata implementation on webpage visibility in search engine results (Part II). *Information Processing and Management*, 41, 691-715.

Zhang, J., Dimitroff, A. 2004. Internet search engines' response to metadata Dublin Core implementation. *Journal of Information Science*. 30(4), 310-320.

APPENDICES

APPENDIX A

Google Keyword Position Results – 1st Experiment

GOOGLE					
website	Rank	Keyword Position			Visibility Percentage
		Top	Middle	Bottom	
www.barnesandnoble.com/	1	50%	25%	25%	100.00%
www.amazon.com	2	100%	0%	0%	96.67%
digital.library.upenn.edu/books	3	40%	20%	40%	93.33%
www.powells.com	4	46%	37%	17%	90.00%
www.bartleby.com	5	0%	0%	100%	86.67%
www.abebooks.com	6	36%	21%	43%	83.33%
www.borders.com	7	66%	34%	0%	80.00%
www.scholastic.com	8	0%	100%	0%	76.67%
www.oreilly.com	9	50%	17%	33%	73.33%
www.cs.cmu.edu/web/books.html	10	40%	20%	40%	70.00%

APPENDIX B

Yahoo! Keyword Position Results – 1st Experiment

Yahoo!					
website	Rank	Keyword Position			Visibility Percentage
		Top	Middle	Bottom	
www.amazon.com	1	100%	0%	0%	100.00%
www.barnesandnoble.com/	2	50%	25%	25%	96.67%
www.borders.com	3	66%	34%	0%	93.33%
www.bookfinder.com	4	43%	43%	14%	90.00%
www.booksense.com	5	43%	0%	57%	86.67%
onlinebooks.library.upenn.edu	6	40%	20%	40%	83.33%
www.allbookstores.com	7	45%	28%	27%	80.00%
www.barnesandnoble.com/book browser	8	57%	14%	29%	76.67%
www.powells.com	9	46%	37%	17%	73.33%
www.bartleby.com	10	0%	0%	100%	70.00%

APPENDIX C

MSN Keyword Position Results – 1st Experiment

MSN					
website	Rank	Keyword Position			Visibility Percentage
		Top	Middle	Bottom	
www.amazon.com	1	100%	0%	0%	100.00%
www.abebooks.com	2	36%	21%	43%	96.67%
www.books.com	3	50%	25%	25%	93.33%
digital.library.upenn.edu/books	4	40%	20%	40%	90.00%
www.nytimes.com/pages/books	5	56%	12%	32%	86.67%
www.powells.com	6	46%	37%	17%	83.33%
www.addall.com	7	50%	25%	25%	80.00%
www.promo.net/pg	8	64%	27%	9%	76.67%
www.alibris.com	9	86%	14%	0%	73.33%
www.ivillage.com/books	10	61%	33%	6%	70.00%

APPENDIX D

AskJeeves Keyword Position Results – 1st Experiment

AskJeeves					
website	Rank	Keyword Position			Visibility Percentage
		Top	Middle	Bottom	
www.amazon.com/	1	100%	0%	0%	100.00%
www.bookfinder.com/	2	43%	43%	14%	96.67%
www.barnesandnoble.com	3	50%	25%	25%	93.33%
www.borders.com	4	66%	34%	0%	90.00%
www.abebooks.com	5	36%	21%	43%	86.67%
www.powells.com	6	46%	37%	17%	83.33%
www.amazon.co.uk/	7	48%	26%	26%	80.00%
www.allbookstores.com	8	45%	28%	27%	76.67%
www.freebooknotes.com/	9	50%	0%	50%	73.33%
www.bythebooks.com/	10	23%	54%	23%	70.00%

APPENDIX E

Yahoo! Keyword Position Results – 2nd Experiment

Yahoo!	Keyword Position			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.amazon.com	43%	43%	14%	1	100.00%
www.barnesandnoble.com	55%	36%	9%	2	96.67%
www.borders.com	40%	20%	40%	3	93.33%
www.bookfinder.com	67%	11%	22%	4	90.00%
www.booksense.com	56%	22%	22%	5	86.67%
www.onlinebooks.library.upenn.edu	50%	50%	0%	6	83.33%
www.allbookstores.com	24%	64%	12%	7	80.00%
www.barnesandnoble.com/bookbrowser	29%	42%	29%	8	76.67%
www.powells.com	41%	37%	22%	9	73.33%
www.bartleby.com	100%	0%	0%	10	70.00%
www.nybooks.com	25%	0%	75%	11	66.67%
www.nytimes.com/pages/books/index.html	45%	55%	0%	12	63.33%
www.bibliofind.com	0%	29%	71%	13	60.00%
www.bookpage.com	14%	57%	29%	14	56.67%
www.bomc.com	0%	0%	0%	15	53.33%
www.oprah.com	29%	71%	0%	16	50.00%
www.digital.library.upenn.edu	60%	40%	0%	17	40.00%
www.abebooks.com	37%	43%	20%	18	46.67%
www.books.com	55%	36%	9%	19	43.33%
www.addall.com	35%	6%	59%	20	36.67%

APPENDIX F

AskJeeves Keyword Position Results – 2nd Experiment

ASKJEEVES	Keyword Position			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.amazon.com	43%	43%	14%	1	100.00%
www.books.bc.ca	0%	0%	0%	2	96.67%
www.bookfinder.com	67%	11%	22%	3	96.67%
www.barnesandnoble.com	55%	36%	9%	4	93.33%
www.borders.com	40%	20%	40%	5	90.00%
www.abebooks.com	37%	43%	20%	6	86.67%
www.allbookstores.com	24%	64%	12%	7	76.67%
www.amazon.co.uk	43%	43%	14%	8	80.00%
www.powells.com	41%	37%	22%	9	83.33%
www.freebooknotes.com	29%	14%	57%	10	73.33%
www.bythebooks.com	27%	66%	7%	11	70.00%
www.digital.library.upenn.edu/books	60%	40%	0	12	90.00%
www.digital.library.upenn.edu/books/banned-books.html	33%	18%	49%	13	90.00%
www.alibris.com	24%	48%	28%	14	73.33%
www.ilab-lila.com	100%	0%	0%	15	66.67%
www.bartleby.com	100%	0%	0%	16	70.00%
www.bibliofind.com	0%	29%	71%	17	60.00%
www.sunsite.unc.edu/ibic/IBIC-homepage.html	25%	30%	45%	18	50.00%
www.bookbrowse.com	18%	27%	55%	19	76.67%
www.ccel.org	27%	46%	27%	20	43.33%

APPENDIX G

AltaVista Keyword Position Results – 2nd Experiment

ALTAVISTA	Keyword Position			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.amazon.com	43%	43%	14%	1	100.00%
www.powells.com	41%	37%	22%	2	93.33%
www.abebooks.com	37%	43%	20%	3	96.67%
www.books.com	55%	36%	9%	4	86.67%
www.k-books.co.jp	0%	0%	0%	5	63.33%
www.digital.library.upenn.edu/books	60%	40%	0	6	90.00%
www.books.or.jp	0%	40%	60%	7	46.67%
www.nytimes.com/pages/books	24%	63%	13%	8	86.67%
www.addall.com	35%	6%	59%	9	80.00%
www.amazon.co.uk	40%	30%	30%	10	70.00%
www.promo.net/pg	82%	9%	9%	11	73.33%
www.gutenberg.net	47%	33%	20%	12	76.67%
www.onlinebooks.library.upenn.edu	33%	67%	0%	13	66.67%
www.booksamillion.com	20%	20%	60%	14	50.00%
www.borders.com	40%	20%	40%	15	80.00%
www.arts-books.com	57%	14%	29%	16	73.33%
www.alibris.com	24%	48%	28%	17	56.67%
www.netlibrary.com	0%	50%	50%	18	53.33%
www.calendarlive.com/books	63%	31%	6%	19	53.33%
www.bartleby.com	100%	0%	0%	20	43.33%

APPENDIX H

MSN Keyword Position Results – 2nd Experiment

MSN	Keyword Position			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.amazon.com	43%	43%	14%	1	100.00%
www.abebooks.com	37%	43%	20%	2	96.67%
www.powells.com	41%	37%	22%	3	90.00%
www.books.com	55%	36%	9%	4	93.33%
www.digital.library.upenn.edu/books	60%	40%	0	5	86.67%
www.addall.com	35%	6%	59%	6	83.33%
www.nytimes.com/pages/books	24%	63%	13%	7	80.00%
www.promo.net/pg	82%	9%	9%	8	76.67%
www.gutenberg.net	47%	33%	20%	9	73.33%
www.booksamillion.com	20%	20%	60%	10	56.67%
www.netlibrary.com	0%	50%	50%	11	60.00%
www.calendarlive.com/books	63%	31%	6%	12	53.33%
www.salon.com/books	40%	40%	20%	13	50.00%
www.oreilly.com	38%	38%	24%	14	70.00%
www.amazon.com/exec/obidos/ats-query-page	0%	33%	67%	15	66.67%
www.ivillage.com/books	81%	0%	19%	16	46.67%
www.alibris.com	24%	48%	28%	17	63.33%
www.bartleby.com	100	0%	0%	18	36.67%
www.ncbi.nlm.nih.gov/entrez/query.	50%	50%	0%	19	0.00%
www.magickeys.com/books	59%	23%	18%	20	43.33%

APPENDIX I

AOL Keyword Position Results – 2nd Experiment

AOL	Keyword Position			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.barnesandnoble.com/	55%	36%	9%	1	93.33%
www.amazon.com/exec/obidos/subst/home/home.html	43%	43%	14%	2	100.00%
www.digital.library.upenn.edu/books/	60%	40%	0	3	96.67%
www.powells.com/	41%	37%	22%	4	90.00%
www.abebooks.com/	37%	43%	20%	5	83.33%
www.bartleby.com/	100%	0%	0%	6	86.67%
www.scholastic.com/	60%	20%	20%	7	80.00%
www.borders.com/	40%	20%	40%	8	76.67%
www.oreilly.com/	38%	38%	24%	9	73.33%
www.nybooks.com/index	20%	0%	80%	10	66.67%
www.promo.net/pg/	82%	9%	9%	11	63.33%
www.magickeys.com/books	59%	23%	18%	12	60.00%
www.netlibrary.com/	0%	50%	50%	13	56.67%
www.ipl.org/div/books/	50%	50%	0%	14	0.00%
www.addall.com/	35%	6%	59%	15	40.00%
www.bookfinder.com/	67%	11%	22%	16	46.67%
www.cnn.com/SHOWBIZ/	0%	100%	0%	17	53.33%
www.alibris.com/	24%	48%	28%	18	43.33%
www.gutenberg.net/	47%	33%	20%	19	0.00%
www.nytimes.com/pages/books/	24%	63%	13%	20	86.67%

APPENDIX J

Google Keyword Position Results – 2nd Experiment

Google	Keyword position			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.barnesandnoble.com	55%	36%	9%	1	93.33%
www.amazon.com	43%	43%	14%	2	100.00%
www.digital.library.upenn.edu	60%	40%	0	3	96.67%
www.powells.com	41%	37%	22%	4	90.00%
www.abebooks.com	37%	43%	20%	5	83.33%
www.bartleby.com	100%	0%	0%	6	86.67%
www.scholastic.com	60%	20%	20%	7	80.00%
www.borders.com	40%	20%	40%	8	76.67%
www.oreilly.com	38%	38%	24%	9	73.33%
www.cs.cmu.edu	50%	50%	0%	10	70.00%
www.nybooks.com	25%	0%	75%	11	66.67%
www.promo.net	82%	9%	9%	12	63.33%
www.magickeys.com	59%	23%	18%	13	43.33%
www.cnn.com	0%	100%	0%	14	56.67%
www.netlibrary.com/	0%	50%	50%	15	53.33%
www.ipl.org/div/books/	75%	25%	0%	16	0.00%
www.addall.com/	35%	6%	59%	17	46.67%
www.bookfinder.com	67%	11%	22%	18	40.00%
www.alibris.com/	24%	48%	28%	19	50.00%
www.gutenberg.net/	47%	33%	20%	20	76.67%

APPENDIX K

Lycos Keyword Position Results – 2nd Experiment

Lycos	Keyword Location			Rank	Visibility Percentage
	Top	Middle	Bottom		
www.barnesandnoble.com	55%	36%	9%	1	93.33%
www.store.aetv.com/html/catalog/bp01_1.html?id=1103	38%	31%	31%	2	93.33%
www.powells.com	41%	37%	22%	3	90.00%
www.alibris.com	24%	48%	28%	4	76.67%
www.borders.com	40%	20%	40%	5	86.67%
www.barnesandnoble.com	55%	36%	9%	6	93.33%
www.powells.com	41%	37%	22%	7	90.00%
www.alibris.com	24%	48%	28%	8	76.67%
www.gutenberg.org	47%	33%	20%	9	73.33%
www.booksinprint.com/bip/	5%	20%	75%	10	70.00%
www.booksamillion.com	20%	20%	60%	11	56.67%
www.amazon.com/exec/obidos/tg/browse/-/283155	43%	43%	14%	12	66.67%
www.netlibrary.com	0%	50%	50%	13	53.33%
www.oreilly.com	38%	38%	24%	14	50.00%
www.salon.com/books	40%	40%	20%	15	50.00%
www.bartleby.com	100%	0%	0%	16	53.33%
www.ivillage.com/books	81%	0%	19%	17	46.67%
www.alibris.com	24%	48%	28%	18	76.67%
www.calendarlive.com/books	63%	31%	6%	19	53.33%
www.magickeys.com/books	59%	23%	18%	20	40.00%

GLOSSARY

Body Text

The body area is the main area of an HTML page which contains all the visible text and images that appear in the browser window.

Bot

See Crawler.

Crawler

A computer program designed to travel across the Internet automatically, gathering information about websites in the process. Also called robot, bot or spider.

e-Commerce

e-Commerce consists primarily of the distributing, buying, selling, marketing, and servicing of products or services over electronic systems such as the Internet and other computer networks.

Front end

That part of a computer program which interfaces with the user.

HTML

Hypertext Markup Language. The lingua franca that enables browser programs to display webpages in an understandable format.

Index

A file which contains all the data collected by the robot(s) of a search engine.

Indexing

The automatic selection and compilation of 'meaningful' words from a website into a list (often referred to as an index) that can be used by a search engine to retrieve pages.

Internet

A worldwide collection of connected computers.

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.

Keyword search

A search for documents containing one or more words that are specified by a user.

Listing

When a page is entered into the database of a search engine it is 'listed'.

Metadata

Data about data. Often used to refer to data identifying various basic sets of data about a 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.

Query

A word, number, phrase(s), operator(s) or sentence(s) which expresses the user's information need in a language which the search engine can understand.

Ranking

A method used by a search engine to sort and display search results in such a way that the most relevant answer appears first, the second most relevant one second, etc.

Relevance

How well a retrieval system provides documents which contain the information a user is looking for, as measured by the user.

Robot

See Crawler.

Search engine

A program which allows a user to specify a query, and then attempts to find information in its index file.

Searching

The process of attempting to find useful information within a large base of unordered data.

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.

Spider

See Crawler.

Traffic

The number of unique visitors to a single webpage.

URL

Universal Resource Locator – the standard name for the address of a website on the Internet.

Webpage

A single page of a website; it will commonly include text, graphics, and links to other web pages.

Website

The entire collection of webpages and other information (such as images, sound, and video files, etc.).