

**THE CROSSOVER POINT BETWEEN KEYWORD RICH WEBSITE TEXT AND
SPAMDEXING**

by

HERBERT ZUZE

Thesis

**Submitted in fulfilment
of the requirements for the degree**

MAGISTER TECHNOLOGIAE

In

BUSINESS INFORMATION SYSTEMS

in the

FACULTY OF BUSINESS

at the

CAPE PENINSULA UNIVERSITY OF TECHNOLOGY

Supervisor: Prof M Weideman

**Cape Town Campus
2011**

DECLARATION

I, Herbert Zuze, declare that the contents of this thesis represent my own unaided work, and that the thesis has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.

Signed

Date

ABSTRACT

THE CROSSOVER POINT BETWEEN KEYWORD RICH WEBSITE TEXT AND SPAMDEXING

With over a billion Internet users surfing the Web daily in search of information, buying, selling and accessing social networks, marketers focus intensively on developing websites that are appealing to both the searchers and the search engines. Millions of webpages are submitted each day for indexing to search engines. The success of a search engine lies in its ability to provide accurate search results. Search engines' algorithms constantly evaluate websites and webpages that could violate their respective policies. For this reason some websites and webpages are subsequently blacklisted from their index.

Websites are increasingly being utilised as marketing tools, which result in major competition amongst websites. Website developers strive to develop websites of high quality, which are unique and content rich as this will assist them in obtaining a high ranking from search engines. By focusing on websites of a high standard, website developers utilise search engine optimisation (SEO) strategies to earn a high search engine ranking.

From time to time SEO practitioners abuse SEO techniques in order to trick the search engine algorithms, but the algorithms are programmed to identify and flag these techniques as spamdexing. Search engines do not clearly explain how they interpret keyword stuffing (one form of spamdexing) in a webpage. However, they regard spamdexing in many different ways and do not provide enough detail to clarify what crawlers take into consideration when interpreting the spamdexing status of a website. Furthermore, search engines differ in the way that they interpret spamdexing, but offer no clear quantitative evidence for the crossover point of keyword dense website text to spamdexing. Scholars have indicated different views in respect of spamdexing, characterised by different keyword density measurements in the body text of a webpage. This raised several fundamental questions that form the basis of this research.

This research was carried out using triangulation in order to determine how the scholars, search engines and SEO practitioners interpret spamdexing. Five websites with varying keyword densities were designed and submitted to Google, Yahoo! and Bing. Two phases of the experiment were done and the results were recorded. During both phases almost all of the webpages, including the one with a 97.3% keyword density, were indexed. The aforementioned enabled this research to conclusively disregard the keyword stuffing issue, blacklisting and any form of penalisation. Designers are urged to rather concentrate on usability and good values behind building a website.

The research explored the fundamental contribution of keywords to webpage indexing and visibility. Keywords used with or without an optimum level of measurement of richness and poorness result in website ranking and indexing. However, the focus should be on the way in which the end user would interpret the content displayed, rather than how the search engine would react towards the content. Furthermore, spamdexing is likely to scare away potential clients and end users instead of embracing them, which is why the time spent on spamdexing should rather be used to produce quality content.

ACKNOWLEDGEMENTS

Without your support this work would not be successful:

- To my God who gave me a revelation and a conquering spirit.
- To my wife, Patience, who provided support and a long suffering heart for me to continue working hard.
- To my supervisor, Professor M. Weideman, who went the extra mile and constantly encouraged me.
- To Mrs Uys, for assistance with the statistical analysis and write-up.
- To my faculty and my colleagues who supported me during the times I needed them most.
- To Tanya, Patience, Derby, Sarah and Rueben for the proofreading of my work.

DEDICATION

This research is dedicated to my first born child with whom God has blessed me.

RESEARCH OUTPUTS

The author produced the following research outputs during this study.

Output Type	Authors	Title	Institution/Event	Status
Poster	Zuze, H., Weideman, M.	Keyword density in website body text- search engine spamdexing.	12th Annual Conference on World Wide Web Applications.	Published: www.zaw3.co.za
Journal Article	Zuze, H., Weideman, M.	An empirical study on the degree of website keyword density accepted by search engines.	South African Journal of Information Management.	Submitted August 2011.
Poster	Zuze, H., Weideman, M.	Cloaking on Google SERP – search engine spamdexing?	13th Annual Conference on World Wide Web Applications.	Submitted July 2011, accepted.
Conference Paper	Zuze, H., Weideman, M.	A comparative analysis of search engine indexing time	13th Annual Conference on World Wide Web Applications.	Submitted June 2011, accepted.

TABLE OF CONTENTS

CHAPTER ONE: BACKGROUND TO THE RESEARCH PROBLEM

1.1	INTRODUCTION	1
1.2	STATEMENT OF RESEARCH PROBLEM	2
1.2.1	Research question	2
1.2.2	Research objectives	2
1.3	BACKGROUND TO RESEARCH PROBLEM	2
1.3.1	Entities of research problem	4
1.3.1.1	Search engines	4
1.3.1.2	Indexing	4
1.3.1.3	Keyword density (richness)	4
1.3.1.4	Keyword stuffing	5
1.3.1.5	Spamdexing	5
1.4	RESULTS AND CONCLUSION	6
1.5	LIMITATIONS OF THE STUDY	7
1.6	CHAPTER SUMMARY	7

CHAPTER TWO: LITERATURE REVIEW: SEO AND KEYWORD RICH WEBSITE TEXT

2.1	INTRODUCTION	9
2.2	SEARCH ENGINES	12
2.2.1	Search engine elements	14
2.2.2	Crawlers	14
2.2.3	Indexing	15
2.2.4	Ranking	16
2.3	WEBSITES	18
2.3.1	Website visibility	18
2.3.2	Webpage usability	20
2.3.3	Sitemaps	22
2.3.4	Keyword placement	23
2.4	SEARCH ENGINE MARKETING	25
2.4.1	The growth of SEM	26
2.4.2	Search engine optimisation	27
2.4.2.1	Search engine spamdexing	28
2.4.2.2	Keywords in body text	30
2.4.3	SEO white hat techniques	31
2.4.3.1	Links	31
2.4.3.2	Inbound and outbound links	32
2.4.3.3	Cross-linking	33
2.4.4	SEO black hat techniques	33
2.4.4.1	Keyword metatag abuse	34
2.4.4.2	Hidden text	34
2.4.4.3	Cloaking	35
2.4.4.4	Keyword stuffing	35
2.4.4.5	Link farm	37
2.4.4.6	Content duplication	37

2.4.4.7	Extra title tags	37
2.4.5	Commercial websites and spamdexing	38
2.4.6	Keyword density	38
2.5	PAY-PER-CLICK	39
2.6	SEARCH ENGINES AND KEYWORD STUFFING	40
2.6.1	Google and keyword stuffing	40
2.6.2	Yahoo! and keyword stuffing	41
2.6.3	Bing and keyword stuffing	41
2.7	CONCLUSION	42

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

3.1	INTRODUCTION	44
3.2	RESEARCH QUESTIONS	45
3.3	ONTOLOGICAL STANCE	45
3.4	EPISTEMOLOGICAL STANCE	46
3.5	CONCEPTUAL FRAMEWORK	46
3.6	RESEARCH METHODOLOGY	48
3.6.1	Quantitative research	48
3.6.2	Qualitative research	48
3.6.3	Triangulation	49
3.6.4	Sampling	50
3.6.5	Types of sampling	51
3.6.6	Interviews	52
3.7	INTERVIEWS WITH SEO PRACTITIONERS	53
3.8	SEARCH ENGINES' POLICES	53
3.9	EXPERIMENTAL RESEARCH	53
3.9.1	Phase 1 website structure	54
3.9.2	Homepage keyword and content structure	54
3.9.3	Websites domain registration	55
3.9.4	Websites submission to search engines	55
3.10	PHASE 2 EXPERIMENT	57
3.11	SUMMARY	58

CHAPTER FOUR: RESULTS AND ANALYSIS

4.1	INTRODUCTION	59
4.2	RESULTS FROM SEO PRACTITIONERS	59
4.3	INTERVIEW QUESTIONS AND RESPONSES	59
4.3.1	How many years have you been in the SEO industry and why did you consider being in this field?	59
4.3.2	What SEO tools do you regularly use and how do they work?	61
4.3.3	What areas do you think are currently the most important in organically ranking a site?	62
4.3.4	If a submitted site is not showing up on SERP, does this mean it has been banned and how long does it take for a site to be indexed and show up on the SERP?	65
4.3.5	Should one optimise for the singular or long tail form of keywords and why?	67
4.3.6	What do you understand by the phrase keyword stuffing?	67
4.3.7	How often do you carry out experiments on search engines to check if their search and indexing algorithms have changed?	68

4.3.8	Do you follow what forums and bloggers say regarding the SEO industry?	70
4.3.9	Have you ever seen a blacklisted or banned website and what does it look like?	71
4.3.10	How do you measure the richness or poorness of a keyword in the body of a webpage?	72
4.3.11	How do Google, Yahoo! and Bing interpret keyword stuffing and do their algorithms stick with their respective document guidelines and procedures?	72
4.3.12	Do you think SEO practitioners and website developers understand spamdexing?	73
4.3.13	Do you think there is a crossover point from keyword rich website text to spamdexing and how do you interpret it?	74
4.3.14	Are sitemaps helpful for webpage indexing?	75
4.3.15	Where do you think the SEO industry is heading?	76
4.3.16	Interviews analysis	77
4.4	INTERVIEWEES' WEBSITES INDEXING ANALYSIS	79
4.5	RESULTS OF E-MAILS SENT TO GOOGLE, YAHOO! AND BING PROFESSIONAL EMPLOYEES	80
4.6	RESULTS AND ANALYSIS OF SEARCH ENGINES' GUIDELINES	81
4.6.1	Google	81
4.6.1.1	Design and content guidelines	82
4.6.1.2	Technical guidelines	82
4.6.1.3	Quality guidelines	82
4.6.2	Yahoo! search content quality guidelines	83
4.6.2.1	Pages Yahoo! wants included in the index	83
4.6.2.2	What Yahoo! considers unwanted	84
4.6.2.3	Bing Webmaster center results	84
4.6.2.3	Guidelines for successful indexing	84
4.6.2.4	Technical recommendations for a website	85
4.6.2.5	Content guidelines for a website	85
4.6.2.6	Techniques that may prevent a website from appearing in Bing results	86
4.6.3	Google, Yahoo! and Bing Webmaster centers	86
4.6.4	Websites' experimental results	86
4.6.4.1	String search	87
4.6.4.2	Site search	89
4.6.4.3	Webmaster tools	89
4.7	PHASE 1 RESULTS AND ANALYSIS	89
4.7.1	Analysis of scraped websites	91
4.8	PHASE 2 RESULTS AND ANALYSIS	93
4.9	INDEXING STATISTICAL ANALYSIS	93
4.9.1	Analysis 1: comparing indexing time between Google, Yahoo! and Bing	94
4.9.1.1	Analysis of Google indexing time	95
4.9.1.2	Survival functions for Google, Yahoo! and Bing	96
4.9.2	Analysis 2: comparing indexing time between two keyword groups	97
4.9.2.1	Mean and medians for survival time	98
4.9.2.2	Overall comparisons	98
4.9.2.3	Cox regression	100
4.9.2.4	Case processing summary	101
4.9.2.5	Omnibus tests of model coefficients	102
4.10	SUMMARY	104

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

5.1	INTRODUCTION	108
5.2	WEBPAGE INDEXING	108
5.2.1	Scraping	109
5.2.2	Search engine algorithm	110
5.2.3	Keyword stuffing	110
5.2.4	Keyword density	112
5.3	SIGNIFICANCE OF THE STUDY	112
5.4	FINAL CONCLUSION	113
5.5	RECOMMENDATIONS	116
5.6	FUTURE RESEARCH AREAS	117
5.7	RESEARCH SUMMARY	117

LIST OF FIGURES

Figure 1.1:	The adaptation of Malaga’s conceptual framework for the generation of site profits	6
Figure 2.1:	World Internet penetration rate by geographic regions – 2010	9
Figure 2.2:	Internet users in the world by geographic regions – 2010	10
Figure 2.3:	Internet users in the world by world regions – 2010	10
Figure 2.4:	Web search elements	14
Figure 2.5:	Paid placement and natural results	17
Figure 2.6:	Google central home page	21
Figure 2.7:	Yahoo! central home page	21
Figure 2.8:	Bing central home page	22
Figure 3.1:	Relationship between various research elements	47
Figure 3.2:	Google site submission console	55
Figure 3.3:	Bing site submission console	56
Figure 3.4:	Yahoo! site submission console	56
Figure 3.5:	Model of partitioning of scientific and non-scientific claims	57
Figure 4.1:	Interviewee SEO industry experience in years	61
Figure 4.2:	SEO tools used by interviewees for decision making process	62
Figure 4.3:	Minimum and maximum indexing time as stated by each Interviewee	66
Figure 4.4:	Experimenting SE algorithm change	70
Figure 4.5:	Webpages indexing predictions by interviewees	80
Figure 4.6:	Getlaptops1 results using a search string on Google	88
Figure 4.7:	Getlaptops2 results using a search string on Yahoo!	88
Figure 4.8:	Getlaptops3 results using a search string on Bing	89
Figure 4.9:	Screenshot of Getlaptops1 scraped by an Iranian website	90
Figure 4.10:	Screenshot of Getlaptops1, Getlaptops2 and Getlaptops3 scraped by an Iranian website	91
Figure 4.11:	Translating a webpage to another language	92
Figure 4.12:	Duration of scraping per website	92
Figure 4.13:	Webpage indexing survival function	96
Figure 4.14:	Webpage indexing hazard function	97
Figure 4.15:	KeywordGroup1 survival function	99
Figure 4.16:	KeywordGroup1 hazard function	100
Figure 4.17:	The survival function for patterns 1-2	103
Figure 4.18:	The hazard function for patterns 1-2	104

Figure 4.19:	Phase 1 website indexing time recorded over a period of 35 Days	106
Figure 4.20:	Phase 2 website indexing time recorded over a period of 35 Days	106
Figure 5.1:	A snapshot of cached result error for Getlaptops1	110
Figure 5.2:	Website indexing time	111
Figure 5.3:	Google SERP showing keyword stuffing results (1)	113
Figure 5.4:	Google SERP showing keyword stuffing results (2)	114
Figure 5.5:	The Mega Business Solution off-line page	114

LIST OF TABLES

Table 2.1: Percentage of U.S. upstream traffic from search engines among verticals	13
Table 2.2: Crawlers for Google, Bing and Yahoo!	15
Table 3.1: Sample size for identified entities	51
Table 3.2: A summary of Phase 1 websites content and keyword densities	54
Table 3.3: A summary of Phase 2 websites content and keyword densities	58
Table 4.1: Phase 1 and Phase 2 search strings	87
Table 4.2: Phase 1 website homepage indexing time	90
Table 4.3: Phase 2 website homepage indexing time	93
Table 4.4: Case processing summary	94
Table 4.5: Means and medians for survival time on webpage indexing	95
Table 4.6: Overall comparison using Chi-Square	96
Table 4.7: Keyword case processing summary	97
Table 4.8: Means and medians for survival time on KeywordGroup1	98
Table 4.9: Overall comparison of survival distributions for the different levels of KeywordGroup1	98
Table 4.10: The case processing summary for the indexing time	101
Table 4.11: The categorical variable coding	101
Table 4.12: Omnibus tests of model coefficients	102
Table 4.13: KeywordGroup1 equation variables	102

APPENDICES

Appendix A:	Letter to Interviewees	127
Appendix B:	SEO Questionnaire	128
Appendix C:	E-mails to Google, Yahoo! and Bing professional employees	129
Appendix D:	Phase 1 websites snapshots	132
Appendix E:	Phase 2 websites snapshots	142
Appendix F:	Iranian scraping site snapshots	151
Appendix G:	Websites Phase 1 sample code	161
Appendix H:	Websites Phase 2 sample code	165
Appendix I:	Phase 1 indexing results recordings	171
Appendix J:	Phase 2 indexing results recordings	173
Appendix K:	Webmaster tools snapshots	175
Appendix L:	Transformed survival data	176
Appendix M:	Survival table	178
Appendix N:	Keyword survival table	180
Terms/Acronyms/Abbreviations		182

CHAPTER ONE

BACKGROUND TO THE RESEARCH PROBLEM

1.1 INTRODUCTION

With over a billion Internet users daily surfing the Web in search of information, buying, selling and accessing social networks, marketers focus intensively on developing websites that are appealing to both the searchers and the search engines (SE). Millions of webpages are submitted each day to search engines for indexing. The openness of the Web makes it easier for speedy growth and success and these developments paved the way for many of the webpages' lack of authority and quality (Castillo, Chellapilla and Davison 2008:68-72).

The success of a search engine lies in its ability to provide accurate search results. Search engines' algorithms search for websites and webpages that violate their respective policies. Several websites and webpages are subsequently blacklisted (Yung 2011:38) and excluded from their index list. According to Yung (2011:38-39), search engines share the list of blacklisted websites, which means that if a website is blacklisted by one search engine, other search engines will do so as well. The increasing use of websites as marketing tools has resulted in major competition amongst websites. Website developers strive to develop websites of a high quality, which are unique and rich in content as this will assist them in obtaining a good reputation from search engines. By focusing on websites of a high standard, website developers utilise search engine optimisation (SEO) strategies and occasionally abuse these strategies to earn a high search engine ranking.

Search engines have revealed that they differ in the way in which they interpret spamdexing (a technique of fooling a search engine's indexing algorithm with the intention of increasing the webpage ranking in the search results of a search engine). As far as keyword stuffing is concerned, there is no clear quantitative evidence being shown as the crossover point of keyword dense website text and spamdexing.

The purpose of this study was to provide experimental evidence towards the different claims regarding the optimum level of keyword density and spamdexing. Various sources, ranging from scholars to SEO practitioners, indicated different views in respect to keyword stuffing, characterised by different quantity measurements deemed correct and optimum levels of keywords' in the body text of a webpage.

1.2 STATEMENT OF RESEARCH PROBLEM

No consistent guidance exists on search engine interpretation of the difference between keyword rich website text and spamdexing, which could result in legitimate websites being blacklisted.

1.2.1 Research question

How do search engines interpret natural language text as being keyword rich or spamdexing?

1.2.2 Research objectives

This research study has the following objectives:

- To determine how search engines interpret spamdexing.
- To determine how SEO practitioners view spamdexing.
- To investigate search engine interpretation of websites with varying keyword density.
- To determine how search engines index webpages with varying keyword density.
- To determine the action which search engines take when they interpret a website as containing spamdexing features.

1.3 BACKGROUND TO RESEARCH PROBLEM

In such an expanding pool of information, reaching the top in search engine rankings is becoming a difficult issue for information providers (Zhao 2004:108-119). Specialised index and directories are used to store information in certain subject areas and due to the increasing number of available webpages, users' experiences are increasingly becoming difficult in finding documents relevant to their interest (Raisinghan 2005a:7).

According to Wilson and Pettijohn (2008:133-149) and Raisinghan (2005a:7), spamdexing techniques, such as repeating keywords several times to improve the rank of the page without adding value to the content of the page, are being employed by numerous savvy website developers. It is, however, claimed that search engines penalise the pages that appear to use the technique. Inevitably, legitimate pages are often unduly penalised or completely removed from a search engine index.

User search behaviour is continuously changing to short time style (average time spent searching for information on a page) as patience is being lost each time a page is opened. This is supported by Nielsen and Loranger (2006:30-34), who strongly expressed their view on interpreting content effects to visitors; they estimated that visitors often take 45-60 seconds to scan through the webpage content. These authors claim that during this space of time a visitor is capable of reading a maximum of 200 words. This gives a clear indication of the extent to which user satisfaction is prioritised, compared to search engine views. In this regard, the facts surrounding presenting content to the user should be clearly defined in 200 words. By doing so the user's interest would be captured for a longer viewing time, which may result in conversion if the website is marketing oriented.

Following the user search behaviour notion, the best measurement in terms of the number of words should be established – it is sufficient to permit search engines a rich harvest of keywords, but not too many as this might scare off human readers (Visser and Weideman 2011).

This growing short attention span led SEO practitioners to employ various tactics in order to have their websites highly ranked as well as meeting usability principles. According to Ron and Zsolt (2011), website owners invest substantial resources in order to influence their online visibility. If the quality of sites corresponds with their estimation for visitors, then SEO aids as a mechanism that improves the ranking by correcting measurement errors. However, Zhao (2004:108-119) found that if the most relevant results are not returned in the first 20 results of the search engine result page (SERP), users tend to lose faith in the search and usually give up and cancel the search as they look for an alternative search.

According to Zhao (2004:108-119), in the past HTML metatags assisted the webpages in achieving higher rankings; however, most of the search engines crawlers now ignore these tags. Nevertheless, Wilson and Pettijohn (2008:133-149), indicated that the following two primary factors are important in today's search engine rankings of a webpage:

- on page keywords: the more focused and clear the content on the webpage is, the greater are the chances of it being ranked high for the main keywords on the page.

- link patterns of a website: that is, the quantity, quality and context of incoming links to the site's domain name.

Zhang and Dimitroff (2005a:665-690) also found that websites need to have keywords appearing both in the page title and throughout the page body text in order to attain better search engine results.

1.3.1 Entities of research problem

The following were identified as sub-topics relative to the research problem and formed the backbone of the research:

- search engines,
- indexing,
- keyword density (richness),
- keywords stuffing and
- spamdexing.

1.3.1.1 Search engines

The author selected Google, Bing and Yahoo! to use for the research, as they are currently the biggest search engines with the greatest market share (Snack 2011).

1.3.1.2 Indexing

According to Weideman (2009:192), indexing is the process of reading and recording the weight-carrying words in a search format to an index file, with the goal of getting indexed quickly (and hopefully ranked well) by the search engine. The search engines "discover" a new site when the spiders find a link to that site from other sites (Malaga 2009:132-139). A user can also submit a website manually for indexing.

1.3.1.3 Keyword density (richness)

According to Malaga (2009:132-139) keyword density measures the extent to which a certain word or phrase appears on a site or a webpage. There is considerable discussion among SEO practitioners as to the optimum level of keyword density. Most agree that if the keyword density is too high the search engine can penalise a site. However, since the search engine does not disclose this level, determining the best

keyword density is often difficult and it has become a guessing game for the SEO practitioners.

In relation to keyword density, Zhao (2004:108-119) considered keyword frequency to be one of the major ranking factors which search engines apply in determining the relevance of the webpage content. In this regard, a search engine analyses the frequency of the appearance of keywords compared to other words in the webpage - the higher the frequency of a keyword the better the chances of the word being deemed relevant to the webpage. Like several other scholars, Zhao is of the opinion that a keyword density of 6 - 10% in the body text of a page is the best for satisfying crawler demands.

1.3.1.4 Keyword stuffing

Keyword stuffing is regarded as one of the black hat techniques (unethical techniques that are used to get higher search rankings) that attempt to achieve a highly ranked site by tricking the search engine algorithms. However, the problem of the ratio of the keyword density to content count has not been clarified.

1.3.1.5 Spamdexing

Activities that were viewed as the best practices a number of years ago can today result in a website being blacklisted on some search engines (Wilson and Pettijohn 2008). Malaga (2007:68-82) also agreed that the introduction of an action research approach, enabling exclusive reliance on the academic literature, is not appropriate. Therefore, the researcher conducted extensive research on the practitioners' best practices and their interpretation of spamdexing.

Spamdexing became a challenge to practitioners as the SEO tactics are centred on attracting the search engine to quickly index and highly rank the site (Abernethy, Chapelle and Castillo 2009:2). High ranking can positively affect the generation of profit. This is supported by Malaga's research, where he designed his conceptual framework in which he defines indexing, on-site factors and backlinks as the major factors that affect websites Return on Investment (ROI). Figure 1.1 shows Malaga's conceptual framework on the generation of site profits.

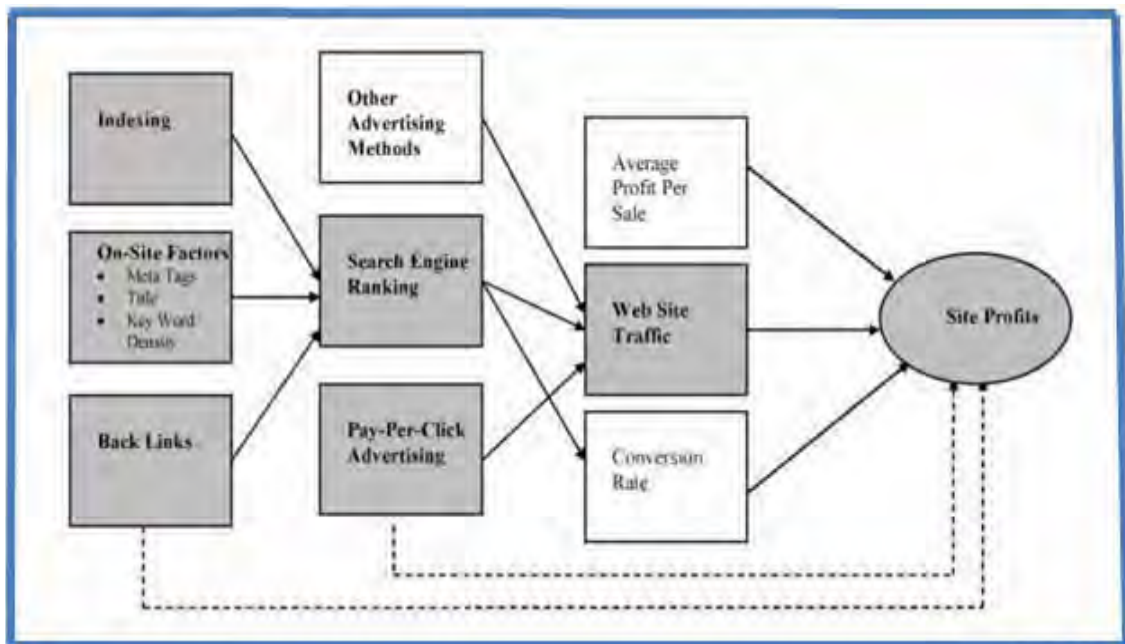


Figure 1.1: The adaptation of Malaga's conceptual framework for the generation of site profits (Source: Malaga 2007).

1.4 RESULTS AND CONCLUSION

The crossover point of the keyword richness of a website to spamdexing is of no significance to the penalisation by search engines experimented in this study. Search engine practitioners should not focus on the reaction of the search engines to the use of keywords in their websites; they ought to focus on designing highly attractive websites that answer users' search questions and satisfy their requirements. This study further indicated that cloaking (pages that give the search engine different content to what the end user sees) still exists and affects website indexing.

Among other elements, the following were identified to be the basic key elements to a good website:

- user satisfaction,
- retention of search users,
- content richness and being semantically synthesised and
- marketing oriented.

This study has further revealed that indexing is relative to crawlers' visitation and basing on two experimental phases conducted during the study, the webpages with the highest keyword densities were the ones that were indexed first. This may be due to the fact that higher keyword densities provided strong emphasis on the importance

of the content published; hence, crawlers gave them first preference as compared to webpages with low keyword densities.

The more content a website has the more weight-carrying key phrases the website could possibly rank for. Search engines reward both qualitative and quantitative websites with solid, informative and useful content, as well as good rankings for specific search terms or phrases (Visser and Weideman 2011).

1.5 LIMITATIONS OF THE STUDY

The research is subjected to the following limitations:

- it considered keyword density only and all other SEO techniques are assumed to be held constant,
- the experiments were based on three search engines only (Google, Yahoo! and Bing),
- only five e-marketing websites were used and
- it did not focus on page ranking.

1.6 CHAPTER SUMMARY

The study established a balanced use of keywords in order to avoid the under-use of keywords, which could result in website under-ranking. The research also determined that keyword density is not a major factor as far as search engines are concerned. Beside the websites in this study, there are other websites that are still making use of the aforementioned method. The research further proved the existence of cloaking on Google search results, which was observed from one of the websites of this study.

Furthermore, the study outcome will enable SEO practitioners to have a clear understanding of the optimum use of keywords, hence preventing the overuse of keywords whose ultimate result is spamdexing. This could increase the productivity of Web developers, content providers and SEO practitioners, since it will eliminate the uncertainty of the outcome of their development.

In addition, an added value is saving SEO practitioners and content providers time as it enables a quick estimation of the correct keyword density. This could ultimately increase user efficiency, given that search engines will be faster as there will be less spamdexing to sort through. The study assisted in obtaining an understanding of the crossover point to spamdexing and full utilisation of keywords in the webpages.

Furthermore, the research contributes to the domain of knowledge base and effectively supplements the existing knowledge base of keywords' usability, distribution, density and spamdexing.

CHAPTER TWO

LITERATURE REVIEW: SEO AND KEYWORD RICH WEBSITE TEXT

2.1 INTRODUCTION

Internet usage is dramatically increasing daily in all sectors of the world, as is Web development which is enhanced by the emergence of new Web technologies. Social networks are continuously dominating as the central pieces of our Web interactions with Facebook, Twitter, LinkedIn and MySpace drawing a considerable number of Internet users. See Figure 2.1, Figure 2.2 and Figure 2.3 for June 2010 World Internet usage statistics.

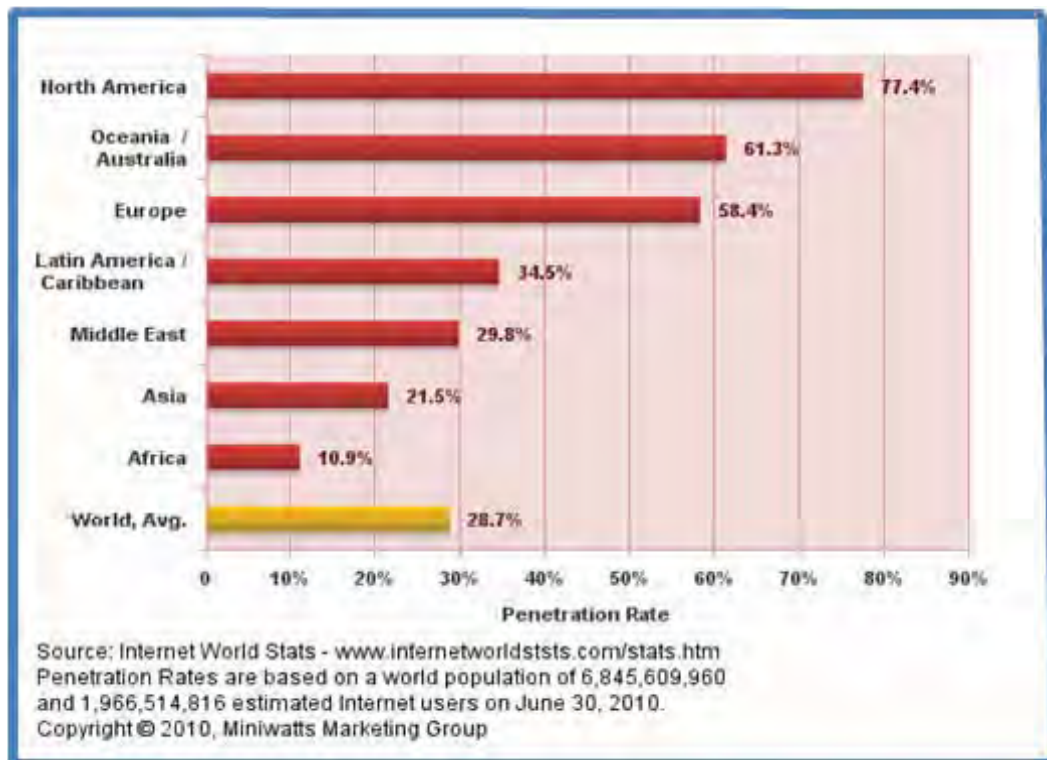


Figure 2.1: World Internet penetration rate by geographic regions - 2010 (Source: Internet World stats organization 2010a).

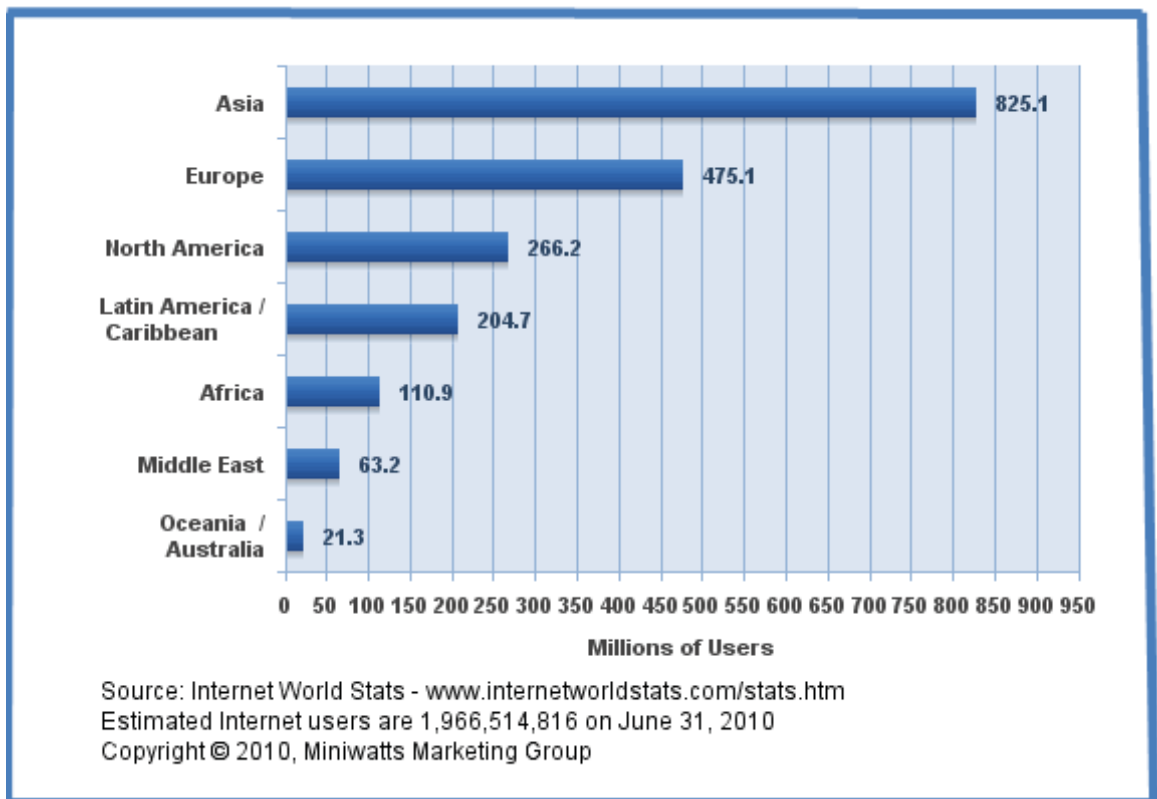


Figure 2.2: Internet users in the world by geographic regions - 2010
 (Source: Internet World stats organization 2010a).

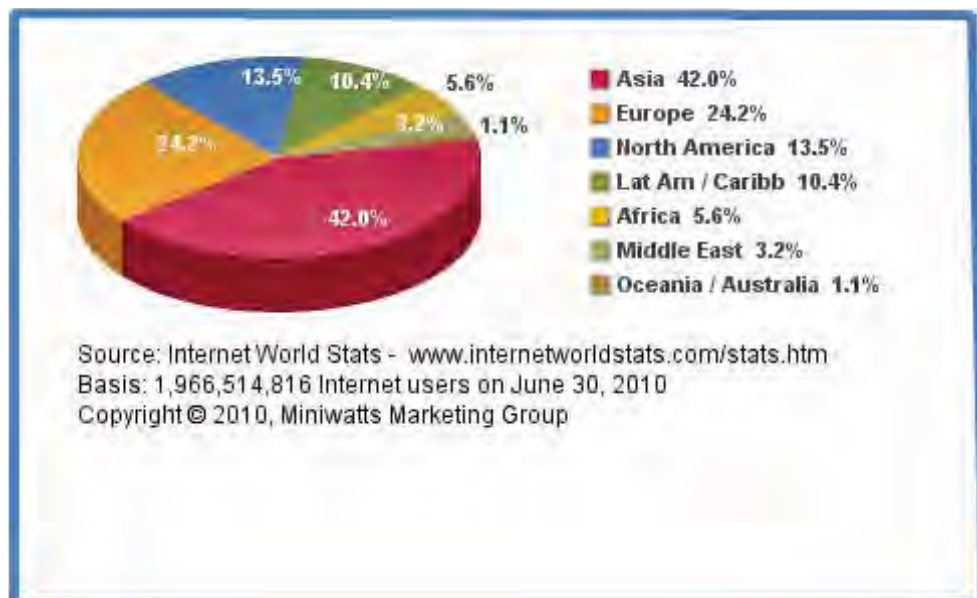


Figure 2.3: Internet users in the world by world regions - 2010
 (Source: Internet World stats organization 2010a).

The dominance of the Web as the library and repository (Gori and Witten 2005:115) of vital information has reinforced the competition in the e-commerce industry. The Internet is both an enabling tool for business and the new business environment that is transforming the economy. According to Smith (2002:6), the Internet is the most

global, borderless, cost-effective and open business application and communication infrastructure. Interactions and relationships between Small to Medium-sized Enterprises (SMMEs) and their customers have changed as a result of the Internet, as the Internet now offers customers many more choices.

Like several other scholars, Singh (2002) identified the Internet as the fastest growing technology in the world, and further described it as having taken only seven years to reach a 25% market share. This is as opposed to the telephone that took 35 years and the television which took 26 years. The geographical distribution of Internet users increased due to the increase in equipment types that can be used to access the Internet. Users visit the Internet via mobile phones, netbooks other than desktops and laptops. This Internet usage increase has resulted in a number of Web marketing platforms and strategies gaining an entry into marketing competition.

As a result, around 80% of users utilise search engines to locate information via the Internet (Zhang and Dimitroff 2005a:665). This fact emphasises the underlying importance of webpage owners being listed with search engines. An important strategy for any website owner is planning how visitors would find their way to their particular site (Thelwall 2001:119). Within this context, Zhang and Dimitroff (2004:310) observed that, "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".

Stolz and Barth (2007:37-47) show how business models for a Web presence can be categorised into one or more of the following categories:

- E-commerce websites: sell products and services,
- Content-based websites: these aim at providing and delivering potentially useful information to their users,
- Communication-based websites: facilitates the exchange of information and contact of users with each other, especially by the use of the Web 2.0 websites that are based on this business variant (Maceviciute 2010:246) and
- Context-based websites: these establish and reorganise content within a new context to develop the entropy for their users, like search engines do.

In line with competitiveness and design, a good marketable website positively defines a site that returns good ranking to be designed in search engine friendly HTML; this

will enable a search engine spider to crawl and index every page of the site (Bruemmer 2001a; Leonard-Wilkinson 2002:30-31).

With ceased euphoria and growing realistic judgement about the possibilities of Internet based businesses, companies focus on the profit generated by their websites. More than ten years after the creation of the World Wide Web, Jacoby and Luqi (2007:43-50) continue searching for the measurable success indicators in their work, in order to evaluate and improve websites and their business models.

2.2 SEARCH ENGINES

Weideman (2004a:2-3), defined a search engine as a program that offers users interaction with the Internet through a front end, where the user can insert a search term or make successive selections from relevant directories. Hereafter, the search engine compares the search term against an index file, which contains information concerning webpages. Matches found are then returned to the user via the front end.

When users search for items or information on the Web they make use of the terms called keywords and/or keyword phrases by entering them into the search engine. Search engines and directories measure both 'link popularity' (quality and quantity of links) and 'click-through popularity' to determine the overall popularity components of a website.

Search engines provide the ordinary Internet user with a (mostly) free and seemingly easy way to locate wide-ranging information on the Internet (Weideman and Kritzinger 2003:231-236). Search engines are used to find information on the Web, whether relevant or not (Alimohammadi 2003:238-242). The index is updated frequently either by human editors or by computerised programs (called spiders, robots or crawlers) (Weideman 2004a:3).

In its process of determining the most relevant webpages, a SE selects a set of candidates' pages that comprises of some or all of the query terms and calculates a score for every webpage. Lastly, a list of webpages are sorted by their respective scores and returned to the end-user (Egele, Kolbitsch and Platzer 2009:51-62).

Website designers often spend a great deal of time concentrating on the visual presentation of a website, which at times result in them failing to take into consideration the content. The value of a webpage's content to users is inspected through search engines by using several complex algorithms. Search engines

achieve this by making use of crawlers to identify keywords in the natural readable content of a webpage (Ramos and Cote 2004). From time to time SEO practitioners overuse SEO techniques in order to trick the search engine algorithm; however, these practices are occasionally caught as spamdexing. Search engines do not clearly explain how they interpret spamdexing in a webpage. Search engines regard spamdexing in many different ways and do not provide enough detail to clarify what crawlers take into consideration when interpreting the status of spamdexing in a website.

More than a decade ago, according to Nielsen (2004), 97% of Web searches were performed on AltaVista, AOL, AskJeeves, Google, Lycos, MSN and Yahoo!. Kennedy (2009) stated that Google has roughly a 70% share and had turned Web searching to resemble a monopoly. However, the world of search has seen a dramatic change in the previous year, with new competitors like Microsoft's Bing, whose new classifications of search are changing the way in which information is discovered.

Recently, data produced by Sterling (Nielsen 2010) revealed that Bing became the number two search engine in the U.S. after overtaking Yahoo!, which subsequently became number three, as depicted in Table 2.1 below. Nielsen claims that Google's August share amounted to 65% (and growth is flat), but that Bing and Yahoo! had switched places and Bing combined with Yahoo! powered search accounted for 24.56 %.

Percentage of U.S. upstream traffic from search engines among verticals									
	Google			Yahoo! Search			Bing**		
Domain	August 2009	August 2010	Year-over-year percentage change	August 2009	August 2010	Year-over-year percentage change	August 2009	August 2010	Year-over-year percentage change
Automotive	22.15%	23.13%	4%	4.61%	4.49%	-3%	1.88%	3.10%	65%
Health	31.62%	30.60%	-3%	5.74%	5.08%	-11%	3.09%	3.82%	24%
Shopping	18.71%	20.07%	7%	3.99%	3.98%	0%	1.66%	2.75%	66%
Travel	28.97%	30.05%	4%	4.63%	4.19%	-10%	2.27%	3.33%	47%
<i>Note: Data is based on monthly upstream traffic from the Hitwise sample of 10 million U.S. Internet users. Figures are for Web searches only.</i>									
<i>**This includes executed searches on Bing.com, Live.com and MSN Search but does not include searches on Club.Live.com.</i>									
Source: Experian Hitwise									

Table 2.1: Percentage of U.S. upstream traffic from search engines among verticals (Sterling 2010).

2.2.1 Search engine elements

Figure 2.4 shows the basic elements of the retrieval of information from the WWW, as illustrated by Langville and Meyer (2006). These authors distinguished the four modules (crawler, page repository, indexers, and indexes) and established that their corresponding data files exist and operate independent of users and their queries.

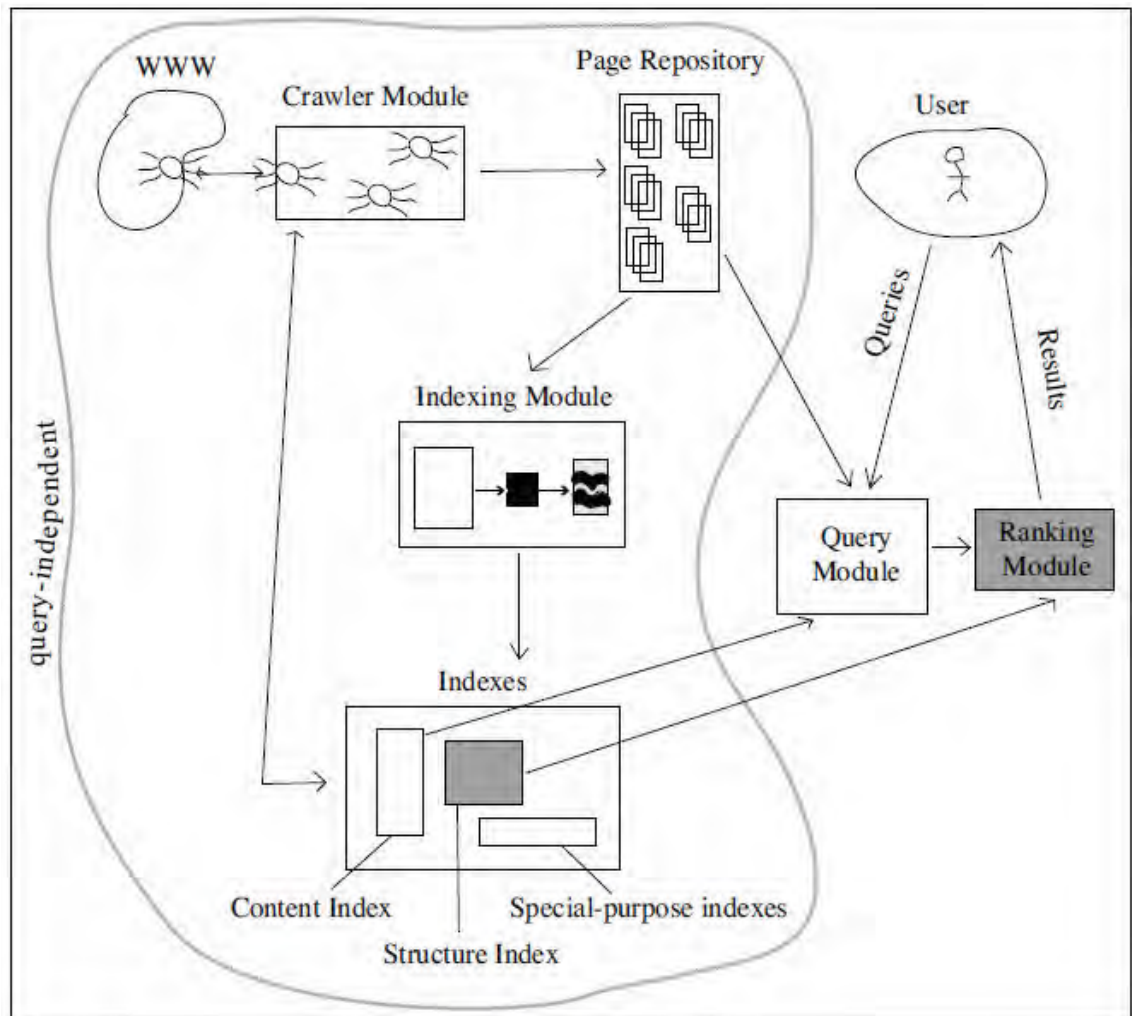


Figure 2.4: Web search elements (Source: Langville and Meyer 2006).

2.2.2 Crawlers

A spider is a software program operated by a search engine that visits a website, records all the words on the pages, and notes links to other sites (Bruemmer 2001a). Spiders do two things, namely text indexing and link following, and this means that if a spider fails to find content or links on a site it leaves the site without noting anything (Leonard-Wilkinson 2002:30-31). Bruemmer further states that if a site's targeted keywords are not "amplified" to appear as a significant component of the webpage,

the search engine will not assign them much importance, which might result in the page not attaining high rankings for these keywords. Spiders track links from one page to another and indexes everything they come across on their way. Nevertheless, they do not see images, Flash movies, JavaScript, frames, password-protected pages or directories.

Web crawlers specialise in downloading Web content, analysing and indexing the content, and they start from the newly surfaced Web to interlinked HTML pages. Web crawlers have limitations if the data is behind the query interface. In order to engage in a dialogue and negotiate for the information, the response depends on the querying party's context (Sharma and Sharma 2010:1). This proves that an effective combination of a spider-happy and user friendly text will aid a webpage in earning a high ranking and can convert the targeted traffic to clients. To ensure top rankings with search engines, the webpage must have good visibility to search engine crawlers. Table 2.2 shows a list of crawlers for Google, Bing and Yahoo!, respectively.

Search Engine	Crawler
Google	Googlebot
Bing	Bingbot
Yahoo!	Inktomi Slurp

Table 2.2: Crawlers for Google, Bing and Yahoo!

Spiders constantly crawl the Web, returning with new and updated pages to be indexed and stored. According to Benczur, Erdelyi, Masanes and Siklosia (2009), when sharing knowledge across different domains, the linkage and the crawl strategies in use differ. For example, Erdelyi, Garzo and Benczur (2011) found that recent results and crawlers' visitations have concentrated on the definition of new features, hence ignoring other important factors in the domain of machine learning techniques that affect SEO results.

2.2.3 Indexing

With an estimated 1.3 billion websites indexed at the time (Zhang and Dimitroff 2004:314), search engines are intended to assist searchers in sorting through the large amount of information that is available on the Internet, hereby playing an

important role in the process of information retrieval. 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 build them into the index (Weideman 2005). Search engines create a map of the Web by indexing webpages according to keywords and then by building those into a database that links page content to keywords and URLs.

Thurrow (2003) described the strategy of placing keyword-rich text on webpages as useless if the “search engine spiders” have no way of finding such text. Thurrow furthermore identified the following website navigation schemes as problematic to ‘spiders crawling’ a website:

- poor HTML coding,
- image maps,
- frames,
- JavaScript, dynamic or database-driven webpages and
- flash components.

Search engines focus on content richness, amongst other things, when indexing websites. Website designers who implement and abuse SEO tactics risk their websites being temporarily or permanently removed from the index. However, if a page cannot be viewed by crawlers then it cannot be indexed and will, therefore, not exist on the search engine’s results page.

Based on the results from the study carried out by Malaga (2007:68-82), a conclusion was made that acquiring a link from a high PageRank (PR) site will result in the Google spider finding a site. However, Google's spider seems to take an inordinately long time to index sites. Furthermore, it appears that links from well known sites lead to a modest improvement in search engine rankings on both Yahoo! and Bing. When discussing Google’s perspective, Mathews (2011: SEO [Search Engine Optimization], what exactly is it again?), mentioned that it should take up to weeks or months for a site to be indexed.

2.2.4 Ranking

It is estimated that nearly 80% of users utilise search engines to locate information on the Internet. This, by implication, places emphasis on the underlying importance of webpages being listed on search engines’ indices (Kritzing and Weideman 2007).

The driving forces behind the high ranking of a site are relevant content and regular updates. A frequently updated site has a high probability of crawler revisits. In an attempt to include fresh content on the site, duplicate content should be avoided since search engines have the ability to identify whether or not the content is an exact copy from another site.

Theoretically, the better a particular website ranks, the more traffic that website should receive and the more visitors ought to convert. This is supported by the fact that on average 67% of search engine users does not look beyond the first SERP (Weideman 2009:32).

It is a great achievement and a motivational factor for many of the website owners to see their websites being highly ranked on SERPs. The high ranking of a website can be achieved through using either or both of the search engine marketing strategies: paid placement and/or SEO. Figure 2.5 shows the location of paid placement and SEO results on a SERP of Google.



Figure 2.5: Paid placement and natural results (Source: Google 2010a).

Ranking higher in search engines have economic benefits, bearing in mind that a highly ranked site is strongly correlational to more traffic that often gives a positive feedback of better revenue (Castillo, Chellapilla and Davison 2007).

2.3 WEBSITES

A website is a collection of related webpages or files and documents hosted on a Web server and accessed on the World Wide Web of the Internet (Cho 2008). A website contains one or more related documents called 'webpages' that are linked to each through the use of hyperlinks. Each webpage should present different and diverse information depending on the characteristics of its content. Fiol-Roig, Miro-Julia and Herraiz (2011:61) stated that the "uncontrolled nature of the Web content presents additional challenges to web page classification as compared to traditional text classification, but the interconnected nature of the hypertext also provides features that can assist the process".

In a website, webpages contain content and the content is summarised by a list of keywords described as text. According to Ricca et al (2004:204), keywords have different scores in a webpage that is; more specific keywords weigh more and have a better score than other keywords. Ricca et al., further states that if a webpage contains a large number of a set of certain keywords then the webpages are regarded as similar, hence the proportionality to keyword weight is measured by its similarity contribution to the webpage (Ricca et al).

Thus, the most important reason for designing a website is to provide content that is relevant to what searchers or end-users are looking for (Weiss and Weideman 2008:2). High-quality content enables a website to be reputable and referenced by other websites and ultimately attain a better positioning on the SERP.

Visser and Weideman (2011) mentioned that the existence of a website can be determined by the page on which it ranks, thus ranking on the first page of the SERP for a given keyword is an indication of the presence of a website. SEO becomes a prerequisite for this reason.

2.3.1 Website visibility

According to Weideman (2009:14), visibility is defined as the ease with which a search engine crawler can find a webpage. After finding the information, it is then defined by the degree of the success the crawler has in indexing the page. Borchardt and Weideman (2008:2) stated that website owners should concentrate on visibility since there is a logarithmic increase in the number of webpages on the World Wide Web. A webpage with high visibility can be easily found and has been designed in

such a way that a crawler will discover a substantial amount of relevant, easy-to-index information on the page.

The most basic purpose of a website is to provide relevant, valuable content and enable users to locate it on a SERP. When a user performs a search for information via a search engine, the interface directs the query to the index where matches are made to the content of the index. The results from the index are presented to the user on a SERP (Weideman 2009:30). If it is a business site, used to promote or sell products or services, the success of the business may depend on whether the site is displayed on the first search page or two when users conduct Web searches through Google and other search sites. Every Internet Web publisher desires a high webpage visibility in SERP, in order to increase accessibility of their webpages. It is widely recognised that many variables contribute to a successful and highly visible website on SERP.

A well-designed, content-rich, and easily navigated site is an aim for Web designers. However, if a website is not crawled and indexed by search engines or if it is crawled and indexed but not well optimised, only a limited number of Internet searchers will access the site. All of the effort made with regards to webpage content would be wasted (Zhang and Dimitroff 2005b:692). Iler (2006) indicated that search engines typically provide in excess of 87% of the traffic to a website. Companies that do not optimise their sites for search are losing opportunities to reach customers. Furthermore, an information-rich, well published website sells products and services to potential clients around the globe, 365 days a year, 24 hours a day. In other words, there are large quantities of webpages being submitted to search engines every day for indexing.

Success in SEO ranking comes down to the survival of the fittest; therefore, visibility is not a right to all who submit their site but a privilege to those who submit suitable and relevant webpage content. With this in mind, SEO practitioners spend more time in trying to incorporate high ranking strategies in their work in order to be visible and highly ranked on SERPs. Studies have shown visibility on search engines as one of the best ways to promote online content. Nonetheless, numerous companies lose potential customers by failing to effectively promote their website through search engines.

According to Iler (2006), a 2003 Forrester Survey established that 87% of Internet users rely on search engines to locate information on the Web. Therefore, website

advertising is a major opportunity that should be taken advantage of as a marketing tool since users are spending more time on the Internet than ever. The introduction of Web 3.0 has seen a drastic increase in the number of Internet users. According to Cho (2008), Web 2.0 is about social networking and mass collaboration with the blurring of lines between content creator and user, whereas Web 3.0 is based on “intelligent” Web applications using:

- natural language processing,
- machine-based learning and reasoning and
- intelligent applications.

The goal is to tailor online searching and requests specifically to users’ preferences and needs. Although the intelligent Web sounds similar to artificial intelligence, it is not the same. Web 3.0 attempts to bring order and allow users to be more accurate in searching and finding precisely what they require. This demonstrates how much effort is being put into webpage optimisation in order to enhance visibility. Each day is a tug of war on the implementation of strategies that results in high ranking (WordStream 2009:3). The first step to success is to have a website included in the indices. Research has proven that at least 67% of users will only read the first page of results, while only 9% will read further than the third page. The implication is that if a website is not listed on the top half of the first page of results, it is virtually invisible as far as the average user is concerned (Chen 2010; Weideman 2008:10).

2.3.2 Webpage usability

Usability measures the extent at which a visitor can easily and quickly use webpage resources. Usability includes the following factors: easiness of learning, subject satisfaction, easiness of use, efficiency of use, memorability and error frequency and severity. The objective of usability is to eliminate any hindrances impeding the experience and process of online communication (Eisenberg et al 2008:158).

Users are generally sensitive to information that they obtain from the Web; what they search for is what they want to be displayed by the search engines. If a user navigates through a webpage and does not attain satisfactory results, they promptly abandon the search or close the site and visit another site. Usability deals with the feeling of being able to easily utilise a website, as it satisfies fundamental principles of user-friendliness. From the three user interfaces for Google, Yahoo! and Bing, it is clear that Google and Bing are simpler to navigate than Yahoo!.



Figure 2.6: Google central home page (Source: Google 2010a).



Figure 2.7: Yahoo! central home page (Source: Yahoo! 2010a).



Figure 2.8: Bing central home page (Source: Bing 2010a).

According to Visser and Weideman (2011), the inclusion of usability attributes will enhance conversion; therefore, effective website design should incorporate usability as a prerequisite. They further state that there is a need for weighing in terms of importance of usability and SEO towards search engines and visitors, since these two practices occasionally contradict each other.

Therefore, websites should be decomposed into smaller segments that are easier to manage and this will enable usability to be a factor considered at each level. Ricca et al. (2004:204) cited that as the website grows, its navigation structure and content tends to grow. This usually results in an increased level of complexity, hence opposing the aspect of usability.

2.3.3 Sitemaps

The sitemaps protocol (Sitemaps 2009) allows a webmaster to inform search engines about URLs on a website that are available for crawling. A sitemap is an XML file that lists the URLs for a website. It allows webmasters to include additional information on each URL, such as when it was last updated, how often it changes and how important it is in relation to other URLs in the website. According to George (2005:29), a sitemap would be a collection of static links, with relevant anchor text, pointing to the dynamic content. This allows search engines to crawl the site more intelligently. Sitemaps are a URL inclusion protocol and complement robots.txt, a URL exclusion protocol. Sitemaps are particularly advantageous on websites where:

- some contents of the website are not linked with public pages and webmasters use rich Ajax or Flash contents that are not normally processed by a search engine and
- sitemaps assist in finding the hidden contents when submitted to crawlers and do not replace the existing crawl-based mechanisms that search engines already use to discover URLs.

2.3.4 Keyword placement

Previous studies conducted by Jupiter research (Iler 2006), indicated that 72% of Internet searchers are not satisfied with their results. The dissatisfaction may be due to the keyword usage in the websites or other Web related matters. Keyword placement may also contribute to the website visibility and/or spamdexing.

Appleton (2010), pointed out a common mistake made by several small website developers, is believing that they have to place primary keywords everywhere on the page (e.g. titles, headers, ALT tags, meta descriptions, body text and anchor text), as often as possible, for visibility purposes. Whilst a certain level of keyword density is imperative, overuse of keywords with little or no synonym consideration can be identified as spamdexing.

Kritzinger and Weideman (2007) indicated that designers of e-commerce based websites should pay close attention to the use of keywords on webpages. For each separate HTML page the relevant keywords must be identified and placed inside the top text areas. However, care must be taken to ensure that the keyword density does not lead to spamdexing penalties by search engine algorithms.

According to Lim (2010), adding countless keywords on a landing page could generate an alert from the search engine crawler, reporting an exploitation of a technique to make a page relevant. This would, however, cause the search engine to remove the URL from the SERP. Lim further warns website developers that this penalty destroys the site's existence, as no traffic to the site would be generated.

Weideman (2009:55) graphically presented, in order of importance, a list of elements to be considered when designing a website for search engine crawler visibility. The ultimate result of the author's model of relative magnitude of positive elements' scores, showed inlinks at the first position, body keywords (second), anchor text (third), meta-tags (fourth), title tags (fifth), H1 (sixth) tags, etc. The keyword structuring in the website is usually placed in all of the above specified areas.

- Title tags - Page title tags carry weight in search engine algorithms; therefore, the title should include the keywords being targeted.
- Meta Keyword description - The Meta Keyword description tag was originally designed to offer the opportunity to influence the SERP description. Website coders and SEO practitioners abused the technique by stuffing the tag with keywords. As a result, the search engine algorithms do not consider the meta keywords. Weideman (2009) indicates that due to the extensive abuse by designers, only a few search engines recognise the keywords metatag. However, it is recommended that this metatag be used as a library to record the important keywords for a webpage.
- H1 and H2 heading tags - Headings have a high level of importance to search engine crawlers and they assume that the information is important and gives an indication of the content of the website (Weideman 2009:90).
- Body text - Body text, such as H1 and H2 tags, should have the keywords appear in the first paragraph of the page.
- Link text - Link text is often referred to as anchor text. Including the keywords in the link text is one of most important elements of a search algorithm.

Kritzinger (2005) did not find any empirical evidence that the placement of keywords, more specifically, the placement of keywords within the body text of a webpage, had a measurable effect on a website's visibility to search engines. Crowell's (2004) opinion reflected that the title tag and the visible body text are the two most important places to insert keywords, due to the fact that search engines index and place significant 'weight' on the text.

In their conclusion, Kritzinger and Weideman (2007) indicated that a website's ranking in search engines decreases when the keywords are placed at the bottom of the webpage. Their results did not include the percentage distribution of the keyword density in the respective areas they mentioned. For this reason the researcher investigated the keyword density in website text and its corresponding effects on spamdexing.

2.4 SEARCH ENGINE MARKETING

According to East (2007:20), SEM is a combination of SEO and PPC (Pay-Per-Click).

Generally, SEO can be defined as the activity of optimising a webpage or a whole site in order to make them more search engine friendly, thus obtaining higher positions in search results. SEO assists in increasing the traffic to one's site and can be integrated in paid search results for given keywords. However, the motive for making use of SEO techniques is to acquire top placement, as a site is relevant to a particular search term, not because of payments made. SEO offers a number of advantages to businesses, such as improved brand exposure and awareness, economical, scalable advertising, and an opportunity for sales and competent market leads (Wordstream 2009:1-2).

According to Stamoulis (2009), SEO is 80% marketing and 20% technical website factors and time should be taken to understand the targeted audience, the demographic and psychographic elements of the potential customers and the realistic consideration of the competitors.

Raisinghani (2005) and Curran (2004:202-205) point out that there are a number of steps a webmaster can take to raise a website's search engine ranking, which include fine-tuning the site's title tag and ensuring that the search engines can easily index the site. Curran (2004:202-205) also suggests conducting extensive keyword or key-term research, in order to determine what users are searching for. In addition, Curran stresses the importance of choosing a good site title, as well as the appropriate and correct use of metatags and site content.

Due to the deterioration of content quality on the Web, preventing spamdexing is a top priority for the search engine industry (Abernethy, Chapelle and Castillo 2009:2; Henzinger, Motwani and Silverstein 2002:11-22). This is particularly true for the high commercial value in top ranked search results.

Even though more complex searches are possible, the majority of Web users conduct simple searches on a keyword or key phrase. Search engines return the results of a search based on a number of factors. All of the major search engines consider the relevance of the search term to sites in their indexes when generating search results. Thus, a search for the word "car" would return webpages that pertain to automobiles.

The exact algorithms used to determine relevance are constantly changed and are confidential and undisclosed (Malaga 2007:68-82).

2.4.1 The growth of SEM

As far as Internet marketing or SEM is concerned, SEO considers what end-users search for and how search engines work. Optimising a website mainly comprises of tweaking its content, HTML and related coding to increase its relevance to specific keywords and eliminating obstacles to the indexing activities of the search engines' filtering process.

The Web is characterised by relentless growth - VeriSign estimates that over one million new domain names are being acquired in the dot-com domain each month (Featherstone, Adam and Borstorff 2009). However, there are many questions still asked that are attributed to website growth such as:

- How much of the growth of the Web may be attributed to business?
- What types and proportions of businesses populate the Web?
- Is the Web more amenable to large businesses or to small businesses?
- Does the Web consist mostly of entrepreneurial start-ups or companies who have adapted their pre-existing business models to this new environment?
- How "entrepreneurial" is the Web?

Throughout (or because of) the frenzy of the dot-com obsession in 2001, various fundamental questions, such as the ones outlined earlier, about business on the Web have remained unanswered (Featherstone, Adam and Borstorff 2009).

It is possible to have a visible website without effective navigation, architecture and other visibility factors. Website marketers can pay for search engine visibility through "paid inclusion" or PPC programs. In addition, MarketingSherpa (2005) shows that websites that are optimised in order to appear higher in SERPs have a higher conversion rate (sales per visit) (4.2%) than PPC ads (3.6%). Clearly, it is important for a site to rank well on the SERPs, which would be within the first three pages, but preferably on the first page. Therefore, several website owners attempt to manipulate where their sites appear on the SERPs through SEO.

Malaga (2007:68-82) noted that site profit is dependent, in part, on site traffic; meaning, that as traffic increases, revenue should also increase. This author further mentioned that search engines are a major traffic driver for numerous sites and search engine traffic may be improved through the use of PPC advertising or SEO. Finally, the same author added that SEO is influenced by indexing, backlinks, and onsite optimisation.

It is more important that e-commerce website designers develop them in accordance to specific guidelines, which will culminate in more website hits, more customers, and a potential higher return on investment.

2.4.2 Search engine optimisation

SEO or search engine positioning (SEP) was introduced in 1997 (Yung 2011:4). It is the process of identifying factors which will improve the amount or quality of traffic to a webpage from search engines via “natural” or unpaid (“organic” or “algorithmic”) search results, which would impact the search engine accessibility of the website.

The abovementioned will ensure that the webpage achieves the highest possible visibility when a search engine responds to a relevant query (Zhang and Dimitroff 2005a:666). Search engine optimisation has grown into a busy industry, with specialists analysing and employing sets of techniques in an attempt to produce relevant human-readable content. Search engine algorithms differ from one search engine to the other, and this makes SEO a complex practice as careful attention is required when optimising a website. SEO is required to please one particular audience, namely search engine crawlers, whilst website usability is aimed at the human user audience. These two audiences are highly dependent on each other, but occasionally they function in a manner where one disregards the other (Visser and Weideman 2011). Considering this phenomenon, visitors’ needs should be satisfied or else the visitor will opt for other sites that address their specified requirements (Kritzinger and Weideman 2008).

Optimising a domain or website for an improved search engine ranking primarily involves regular editing of its content in order to both increase its relevance to specific keywords being used in search queries and to encourage more frequent indexing by search engines.

Traditional websites lack a standardised structure and an emphasis on strong visual presentation of content to capture the attention of the viewer/researcher, not

necessarily focusing on data. As a result, search engines need to apply complex algorithms to determine whether or not those websites might be beneficial to visitors, evaluating a number of variable characteristics (Telnic.com 2009).

Matching one particular search algorithm in trying to cheat the system at a particular time will not be a viable long-term marketing strategy. Real SEO is not about exact domain name vs. number of keywords vs. outbound links. Search engine companies employ human methods to evaluate and improve their search results. Specially trained “ratters” inform the search engine if the top results are relevant for the query, while a monitored number of clicks show whether actual users follow the links at the top of the search results table (Telnic.com 2009).

SEO requires devotion of significant periods of time and effort to ensure that significant results are attained. It has become a very complex, sophisticated practice that requires constant research, practice and re-evaluation in order to be effective. Axandra (2008), mentioned that SEO is a long-term beneficial marketing plan and if well-articulated it realises profitable results. However, its success relies on patience and a particular way of paying attention to details and policies that govern its operability.

2.4.2.1 Search engine spamdexing

Spamdexing has existed for over a decade and content credibility problems have received a fair share of research attention over the past few years. The trustworthiness of content and the assessment of information accuracy has been a major challenge that resulted in many requests for technology to facilitate proper measurement to address the problems (Castillo, Gyongyi, Jatowt and Tanaka 2011:313-314). A recent workshop of the Joint WICOW/AIRWeb Workshop on Web Quality¹ at WebQuality 2011 was conducted with the main objective being to provide the research communities working on spamdexing, abuse, credibility and reputation topics with a survey of current problems and potential solutions.

The term spam is most closely associated with batches of unsolicited e-mail (Hayati and Potdar 2009:3-7; Weideman 2009:28); this is referred to as e-mail spam. According to Weideman (2009:28): “Search engine spam, also known as spamdexing (SPAMming the inDEXes of search engines), is a deliberate attempt by website developers to bypass the intelligence of computer programs and the intention of this is to deceive search engines”.

The fundamental difference between e-mail spam and search engine spam (spamdexing) is that e-mail spam is centred on tricking the users whilst spamdexing is targeting the search engine crawlers in order to rank high on the search engine results. Jerkovic (2009), described spamdexing as the same as search engine spam and Web spam which involves the repeated use of certain keywords in webpages. The author further highlights the fact that Google removes a website from its index if the website is suspected of spamdexing. Clay and Esparza (2009) reviewed spamdexing, as viewed by a search engine, as a violation and misdirection of search engine policies. Most interestingly, they added that spamdexing is measured by the intent and extent to which a SEO tactic is used, which does not clarify the degree and measurement of the extent to which a particular tactic is used, which became the basis of this research. In this research, the researcher will make use of the terms spamdexing and spam (unsolicited e-mails), with their respective meanings, as defined above.

According to Gyongyi and Garcia-Molina (2005), there is some progress regarding the area of spamdexing filtering; areas of formulating ways to ascertain useless Web content that have the sole purpose of manipulating search engine results. Four years later, after Gyongyi and Garcia-Molina's (2005) comment on the earlier work on spamdexing and workshops held, Dai, Davison and Qi (2009) cited that the Workshop on Adversarial Information Retrieval held in 2009 also did not present marginal improvement over spamdexing results and did not attempt to compare their performance. However, several papers focus on providing temporal structures on improving classifications, but significant progress was not achieved. However, Hotho, Benz, Jaschke and Krause (2008) mentioned that the introduction of social networks have caused spamdexing to draw the attention of researchers.

Despite the fact that search engine guidelines are not clear on the subject of spamdexing, Ralf, Pettijohn and James (2008:133) are in agreement that Google made another change in the rules of the search engine when it introduced an extra set of factors (in addition to keyword density) into its ranking algorithm, hereby affecting the quantity, quality and context of incoming links to a website. The result of this change is that linkages to a particular website have a significant impact on how high it is ranked with respect to a particular keyword.

However, according to Kimura, Saito and Sato (2005:723-729) there are often SEO spamdexing websites that contain little or no relevant content and whose sole aim is to increase their position in the search engine rankings. Such spamdexing involves

obtaining more exposure for a website than it deserves for a given search term, leading to an unsatisfactory search experience. Hence, it is an important research topic to investigate SEO spamdexing on websites.

2.4.2.2 Keywords in body text

Keywords are focal descriptive natural language words that are utilised as search queries by Internet users when searching for information on search engines. Keywords are important SEO items for every search engine since they are what search strings are matched against. As a result, it becomes very important that websites are optimised for the right keywords. Website keywords must be capable of describing the content of the site and should generally be utilised by users to locate the site if searched for. Website developers need to consider not only the relevance of keywords of the site but also the statistical and demographic distribution of users' monthly keyword search patterns on the SERP.

Too much text is viewed as "bad writing" and this occasionally occurs on websites. However, text on each page should be halved and the remaining text should then be halved again to reduce congesting content to the users (Johnson 2007:169). Measures must be taken so that webpages are not overloaded with similar targeted keywords, hence reducing readability.

Body text is described as the textual content of a website. Different views are presented by many scholars in respect of body text and keyword density. Various SEO experts consider the optimum keyword density to be 1 - 3 %. Using a keyword more than that could be considered search spamdexing (Wikipedia 2010). Todaro (2007) recommends that keyword density be kept to 3 - 4% of the body text, per page. The author further reiterates that the overuse of keywords in the body text raises a red flag with Web spiders and might disqualify the site.

Charlesworth (2009) is of the opinion that keywords which appear twice in 50 words have a better ratio than the use of four keywords in 400 words. Appleton (2010) states that a benchmark of 3 - 7% keyword density is acceptable and points out that anything more than this strays into spamdexing territory. This prompts the question - what is the best keyword density in body text?

Certain 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, in Google’s PageRank algorithm (Weideman 2004b:904-907).

Keyword stuffing poses a problem in terms of content relevancy; knowing the right amount of keywords that does not compromise relevancy is important. However, Kassotis (2009) stated that for every 100 words in a webpage, one keyword should be used. Kassotis (2009) further states that the general rule is a keyword density of 2 - 2.5%.

Keyword research is one of the foundations and cornerstones of a search engine’s marketing effort. A site that makes use of popular keywords can easily be discovered, while one that does not may languish in obscurity. Web developers ensure that their keywords are the terms which the searchers use when conducting searches on the Internet. Those keywords have to match the search queries used by searchers when querying the search engines. The on-page factors relate to the effective use of keywords. Keywords mark the fundamental strategy, be it SEO or PPC.

One of the foundations of an effective search engine marketing campaign is to select the best keywords that potential customers might use to find the site (Weideman and Kritzinger 2003:231-232). 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 as well. No diminishing returns were found with full-text keywords (Zhang and Dimitroff 2005a:688). If the wrong keywords are initially chosen, all the time and effort that is devoted in attempting to get the site a high ranking with search engines will be pointless.

2.4.3 SEO white hat techniques

According to Weideman (2009:181), a white hat technique is an ethical and above board SEO method that is used by SEO practitioners and website designers on websites to achieve a higher ranking. Among others, the use of correct metatags, correct use of keywords and inlinks helps to achieve positive desired website goals.

2.4.3.1 Links

According to Goldsborough (2005:40-42), the Web is heavily populated with interlinked pages; sites linking to each other increases ranking chances on search pages. Search engine crawlers track these links to index the website. Links are one

of the most important SEO strategies that guarantee a high ranking of a website. The more links to a site, the better the chances of ranking higher, as the quality and quantity of links form part of the algorithms of search engines for calculating relevancy (Mbikiwa and Weideman 2006).

Even though an increased number of sites linking to a website has a positive impact on the website, “link farms”, which are websites that are not genuine directories but rather schemes merely designed to increase site's rankings, should be avoided. Goldsborough (2005:40-42) mentioned that websites that make use of these practices risk penalisation by search sites.

Paulussen (2004:449-466) concludes that it becomes very easy for search engine spiders to navigate through a well-constructed and coded linking structure and this could ultimately assist visitors in easily locating website content via search engines.

2.4.3.2 Inbound and outbound links

Outbound links are any hypertext links that start from a site and lead to another site. Inbound links, or backlinks, on the other hand, come from an external site to one's site. The popularity and most importance of a site is measured by the backlinks and most importantly are the links originating from reputable sites.

Having inbound links from reputable sites boosts rankings (Zhao 2004:108-119), whilst bad links risk website blacklisting. Inbound links are important for the following reasons:

- they can increase website ranking in SERPs for queries using the keywords in the link anchor text,
- inbound links from a site already indexed will enable robots to find a site during the normal indexing process,
- they can increase a website's PageRank or ranging with similar algorithms and
- they bring extra traffic to the site.

Links can also be from page to page and join the documents of the same site; these links are called internal links. Internal links are not as important as the backlinks, but are also fundamental to SEO.

2.4.3.3 Cross-linking

Cross-linking is a method of simply obtaining inbound links from other sites.

2.4.4 SEO black hat techniques

George (2005:18) defined 'Black Hat SEO' as a practice that causes a site to rank higher than its content would otherwise justify or that any changes made specifically for search engines that do not improve the user's experience but the site. Ralf, Pettijohn and James (2008:133) noted that in the early days of the Internet, a site's ranking was entirely based on the words and HTML code found on its webpages. Thus, online marketers were only required to repeat a keyword dozens of times on their webpages, and the pages would automatically rise to the top of the rankings. The world of search is different these days and such practices are viewed as a type of spamdexing. This practice may result in a website being penalised in respect of its rank in search results or being completely removed from a search engine's index. With current search engines, Romow Web Directory ([Peter] 2007) listed the following tactics as the widely exploited ones:

- cloaking,
- duplicating content,
- keyword metatag abuse,
- hidden text,
- bad links and
- extra title tags.

Krause, Schmitz, Hotho and Stummer (2008) advocated for a spamdexing-fighting mechanism that could result in the spamdexers refraining from their practices, which would further increase the benefit of sharing Web content as responsibly and truthfully as possible. Erdelyi, Benczur, Masanes and Siklosi (2009), state that it is difficult to define the boundary between spamdexing and honest search engine optimisation; this makes spamdexing filtering important in Web archives. The spamdexing concern is increasing with different measurement and estimates, thus roughly 10% of the websites and 20% of distinct HTML pages comprises of spamdexing (Erdelyi et al. 2009). The aforementioned affects resource storage wastage and further results in an increase of processing and bandwidth, hence economic sustainability is compromised.

2.4.4.1 Keyword metatag abuse

The keyword metatag is no longer very useful in website ranking in the search results, but there is a need to avoid overstuffing this tag with hundreds of keywords. Weideman (2009:76) indicates that due to the extensive abuse by designers, only a few search engines recognise the keywords metatag. Nevertheless, it is recommended that this metatag be used as a library to record the important keywords for a webpage. The most important and relevant keywords to a site should be listed.

Example of a metatag

```
<META NAME="KEYWORDS" CONTENT="Laptops, laptop, getlaptops1, computers, notebook">
```

According to Weideman (2009:76), George (2005), and Ramos and Cota (2004:50) abusing a metatag by listing hundreds of keywords, repeating keywords or listing keywords that are not relevant to a site's content essentially results in a site's ranking in Google worsening.

2.4.4.2 Hidden text

Henzinger, Montwani and Krause (2002) described text spamdexing techniques being used to modify the text in such a way that the search engine rates the page as being particularly relevant even though the modifications do not increase perceived relevance to a human reader of a document. The author also noted that there are two ways to attempt to improve ranking:

- by concentrating small sets of keywords (repeating these keywords often at the bottom of the webpage) and
- increasing the number of keywords of the document.

Style sheets (CSS spamdexing) can be used in an effort to hide these manipulations from search engines' anti-spamdexing filters (George 2005:19). Search engines crawl the Web, looking at particular site items (mainly text) to get an idea of what a site is about. All search engines evaluate the content of a document to determine its ranking for a search query. Hidden text is one of the oldest keyword stuffing techniques available, but hiding text on website is easily detected by Google and the other major search engines and will result in a lower search ranking or removal from their index.

Hiding text can be achieved by filling the webpage background with keywords and then matching the colour of the text with the colour of the background, so that it cannot be seen by humans (Yung 2011:40). A variation of matching the colour to hide text is to make the text small, so that it will not be visible to human website visitors.

2.4.4.3 Cloaking

This is a form of a spamdexing technique where content or URLs presented to search robots are entirely different to the one presented to human visitors. For example, the displaying of information that would lead search engines to think that a website deals with the selling of second hand laptops in Cape Town, but when a searcher clicks on the website, it displays advertisements for Viagra. Webpages or content will be delivered differently depending on the IP address and/or agent who is requesting it.

2.4.4.4 Keyword stuffing

Keyword stuffing is regarded as one of the black hat techniques that attempt to achieve a highly ranked site by tricking search engine algorithms. However, a problem pertaining to the ratio of the keyword density to content count has not been clarified. Keywords are fundamental to websites and result in a positive effect on search engine ranking. Therefore, there should be an acceptable ratio of keywords and the content density of a site. A line that demarcates the crossover point for a site to be regarded as spamdexing by search engines should also be included. For this reason, the researcher performed an experiment to determine the acceptable point at which frequency and density of a keyword can cause a website to be blacklisted, regarded as spamdexing or removed from the index.

Dejarnette's (2010) recent article describes the importance of keywords and key phrases as help that direct searchers to content they wish to see on the Internet and keywords which leads a searcher down multiple paths to many matching and relevant websites. It is a filtering process that leads the holder to the destination which they wish to visit. In the view of supporting the importance of keywords in content, Ricca et al. (2004:206) provided a simple approach to weighing the importance of a word in a webpage. The authors stated that more frequency or repeated referral of word occurrence in a webpage reflects its importance.

An overriding motivation for all website designers and website content providers is to have highly ranked websites, well indexed positions, maintain their website

competitively and at the same time complement the related markets (Iler 2006). Much has been done to ensure that SEO is considered at all times. Among the main strategies that are paramount for a website to be highly ranked are keywords. Numerous studies have been conducted to determine the importance of keywords to the success of a highly ranked website. Bowman (2004) reflected on the following factors that are important to take note of when using a keyword:

- how many times they are repeated,
- where they appear,
- how they are positioned relative to each other and
- which tags surround the keywords.

On the other hand, Weideman (2009:64-68) indicates that there should be a balance between the use of a keyword or phrase to both the user and the search engine crawlers. This is necessary in order to achieve the website goals. Nevertheless, search engines are aware that other websites are designed in exclusion of properly laid down procedures and regulations governing website visibility. Web developers, SEO practitioners and content providers are required to utilise keywords in such a manner that they avoid keyword stuffing, which could lead to the website being blacklisted. However, at the same time enough keyword density needs to be maintained in order to earn a high ranking.

Past studies indicate that many scholars and various keyword tools place emphasis on the importance of frequent use of keywords in website content. It should be ensured that the site will not be suspected from spamdexing, which can result in the website being removed from the index. A big risk is time wastage. Time might be wasted in attempting to add additional keywords to the content whilst there may be a point and keyword count that is acceptable by the search engines.

One of the worst mistakes which should be avoided is densely distributing keywords, to the extent that it results in websites being blacklisted or regarded as spamdexing. In the conclusion to their study, Kritzinger and Weideman (2007) stated that 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 must be identified and placed inside the top text areas. However, care ought to be taken to ensure that the key word density does not lead to spamdexing penalties by search engine algorithms.

2.4.4.5 Link farm

It is a black hat technique whereby an array of webpages have a large number of hyperlinks interchanged to each other or other pages. It is done with the main motive of promoting artificially boosting link popularity and ultimately enhancing search engine rankings. The sole intention of the linking of the sites to one another is to boost popularity. This collection of the links can point to every other page in the site. Furthermore, a link farm can be added at the bottom of the webpage. This artificially increases the number of links to the webpages, with the main aim of increasing rankings in the SERP. According to Beel and Gipp (2010b:297-298) link spamdexing can be created by using dummy websites that link to the website intended to be pushed, exchanging links with other webmasters, purchasing links on third party webpages and posting links to one's website.

2.4.4.6 Content duplication

Content duplication is identical or very similar to pages that can be accessed from different URLs. An example would be copies of the Open Directory Project listings or online books taken from projects. An individual might steal the content of the website (George 2005:20).

George also identified the following as content duplication:

- documents served in different formats, HTML, PDF, text for different audiences,
- mirrored documents to avoid delays or to provide fault tolerance,
- content syndicated and re-branded for different audiences and markets,
- duplication of press releases by media outlets and
- registering of different versions of names pointing to the same content in order to protect business trademarks.

2.4.4.7 Extra title tags

Adding extra HTML codes to include multiple title tags is known as "Title Stacking" and can simply be found by search engine robots. Web designers or content providers attempt to stuff more keywords into the very important title tag by duplicating the tag. The effects of this are not seen by the website visitor even

though this may have helped some webpages to rank better for those particular keywords (Romow Web Directory [Peter] 2007).

2.4.5 Commercial websites and spamdexing

For commercially-oriented websites, whose income depends on their traffic, it is in their interest to be ranked within the top 10 of the SERP for a query relevant to the content of the website. The purpose is to boost those websites' rankings on search engines such as Google when users search for keywords. Higher rankings mean more click-through and, often, more money (Zahorsky 2010:32-33). Henzinger, Motwani and Krause (2002) also agreed that to achieve a high ranking, website developers deliberately try to manipulate their placement and to achieve this they use either a text-based approach, a link-based approach, a cloaking approach or a combination of either or all of the techniques.

Search engines are constantly reviewing their algorithms in order to become resistant to these spamdexing techniques. On the other hand, the search engines do not publish or review their anti-spamdexing doctoring techniques to avoid aiding the spamdexers to circumvent them. Bartow (2010:1079-1080) reflected on the need for good laws, norms and codes of conduct to be regulated over the Internet usage; the author further noted that failure to put these in place will result in the Internet being compromised.

However, the Discovery Challenge 2010 recorded best spamdexing classification results (Nikulin 2010) and Guang-Gang Geng, Jin and Zhang (2011) provided a number of classifications quantities of quality components that would help to increase the Web content quality.

2.4.6 Keyword density

Keyword density relates to the percentage of target keywords relative to the total text on a page. Keyword density is calculated as follows:

$$KD = (K/TWC) * 100$$

KD is the Keyword density.

K is the number of times a specific keyword is repeated in a page and

TWC is the total number of words analysed on the page.

The density of a keyword phrase is calculated as follows:

$$KPD = (K * WPC / TWC) * 100$$

KPD is the Keyword Phrase density.

K is the number of times a specific keyword is repeated in a page

WPC is the number of words in the phrase and

TWC is the total number of words analysed on the page.

Generally, the idea is that the higher the keyword density, the more relevant to the search string a page is; however, the quality of a keyword does not go hand in hand with its quantity. As the keyword quantity increases so does the probability of keyword stuffing (Webconfs.com 2010).

2.5 Pay-Per-Click

To achieve a highly visible website without effective navigation architecture and other visibility factors, website marketers are forced to pay for search engine visibility through “paid inclusion” or “pay-per-click” programs. That is, the site owner only pays when a user clicks on the advertisement and visits the target site (CyberWyre 2006).

Ralf, Pettijohn and James (2008:133) pointed out that PPC advertisements on search engines can be relatively effective, yet it can also be rather costly. Given the potential of a high cost, it is important for online marketers to understand how they can increase their site's ranking in “natural” or “organic”, unpaid search results. In other words, it is imperative that they understand the importance of SEO. SEO can result in substantial savings, and at the same time it can be used to enhance PPC campaign results.

According to Wordstream (2009:7-8), PPC and SEO, if implemented together, support each other, hence they are considered as complementary. Furthermore brand trust and increased conversions are easily attained if ranking is both natural and paid results. Wordstream also stated that pages that are well optimised for organic search likewise make strong destination URLs for paid advertisements and have a positive impression on quality scores; hence, better results can be achieved if PPC and SEO are used simultaneously.

2.6 SEARCH ENGINES AND KEYWORD STUFFING

Many scholars mentioned the relationship between search engines and the correct and respectable use of keywords. Different opinions revolved on how search engine algorithms perceive the certain levels of keyword density as spamdexing. Some of these mixed ideas formed the basis of other SEO practitioners' company strategies in respect to keyword stuffing control.

2.6.1. Google and keyword stuffing

Do websites developers really understand Google's viewpoint on spamdexing or is it merely a perpetuating confusion and fear of the unknown? Google employs a complex, proprietary algorithm based on 200 factors to decide which sites rank high for any given keyword search. Google explains the process in the following manner:

"PageRank (e.g. the software behind Google's ranking technology) performs an objective measurement of the importance of webpages by solving an equation of more than 500 million variables and 2 billion terms. Instead of counting direct links, PageRank interprets a link from Page A to Page B as a vote for Page B by Page A. PageRank then assesses a page's importance by the number of votes it receives" (Google 2008b).

According to Google (2008a), "PageRank also considers the importance of each page that casts a vote, as votes from some pages are considered to have greater value, thus giving the linked page greater value. Important pages receive a higher PageRank and appear at the top of the search results. Google's technology uses the collective intelligence of the Web to determine a page's importance. There is no human involvement or manipulation of results ..." "Google's search engine also analyses page content (through a process Google calls "Hypertext Matching Analysis"). However, instead of simply scanning for page-based text (which can be manipulated by site publishers through metatag), Google's technology analyses the full content of a page and factors in fonts, subdivisions and the precise location of each word. Google also analyses the content of neighbouring webpages to ensure the results returned are the most relevant to a user's query".

Google, in their Webmaster guidelines (Google 2010e), states that webpages should not be loaded with irrelevant keywords, otherwise known as "keyword stuffing". This practice is an attempt to manipulate a site's ranking in Google's search results.

According to George (2005:52), filling pages with keywords result in a negative user experience and can harm a website's ranking. The focus should be set on creating useful, information-rich content that appropriately utilises keywords and use them within the correct context. The aforementioned raises various questions, such as the level of keyword density Google deems suitable. This raises several of the fundamental questions that form the basis of this research.

2.6.2. Yahoo! and keyword stuffing

Some search engines (such as Yahoo!) have a large number of human site editors, and a 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. Yahoo! defines search engine spam (spamdexing) as pages that are considered unwanted and appears in search results with the intent to deceive or attract traffic and with little regard to the relevance or overall quality of the user experience (Yahoo! 2010c). Yahoo! excludes these webpages in its index and regards them as unasked for sites. Similar to Google, Yahoo! does not clearly specify its interpretation of keyword density or spamdexing (keyword stuffing). Crawlers' spamdexing interpretation details are hidden and unclear.

2.6.3 Bing and keyword stuffing

Likewise, Clay (2009) states that the Bing search engine, under its content guidelines, explicitly states that the MSNBot identifies webpages filled with irrelevant keywords. Canel (2010) reflected on the impending change from MSNBot to Bingbot as the new crawler. Bing crawls a variety of content forms found on the Web, index the content, apply appropriate algorithms, and finally directs relevant content to user queries on SERPs (DeJarnette 2009).

It is clear that these search engines do not quantify their definition of spamdexing, which enhances this research on crossover point of keyword density to spamdexing. Fundamentally, DeJarnette (2009) defined SEO to organising website content and using other non-paying methods to help improve a site's placement in search engine results. In addition, SEO is a subset of search engine marketing (SEM), which also includes buying keywords or sponsoring search results to artificially elevate a website's placement in search engine results.

The author further wrote that websites which naturally appear high in search engine results typically have maximised their SEO implementation, by using:

- relevant and unique title tags on all webpages,
- succinct metatag content descriptions,
- header tags (specifically h1 tags) within the content,
- text navigation links,
- XML Sitemaps and
- robots.txt files.

Contrary to other authors, DeJarnette (2009) did not list keywords as one of the important SEO methods to be maximised for high ranking.

DeJarnette (2009), highlighted in his article that SEO is fundamentally about creating websites that are beneficial to users. The author's most basic advice for achieving optimum rank for a website in Bing was to do the following:

- develop original content (including well-implemented keywords) directed towards the intended audience,
- use well-configured code in the webpages (including images and sitemaps) so that users' Web browsers and search engine crawlers can read the wanted indexed content and
- earn several, high-quality, authoritative inbound links.

2.7 CONCLUSION

Keywords are a fundamental element in website design as they contribute a great deal in the content structuring in websites. They form the basis of webpage visibility and enable webpages to be searched on the WWW. Regardless of which strategy a company uses, keywords form the epicentre of a well-designed website and contribute to the success of good websites. However, caution should be taken when using them to avoid annoying search users rather than enhancing satisfaction. Likewise, search engines rely on keywords to display information on their search pages in response to search queries. It is clear that website designers and online marketers have different goals and do not necessarily cooperate when a website is being constructed, hence some website risk penalisation if black hat techniques are used.

Some scholars agree that keywords are not as important as they used to be, even though there is no way one can use any ranking technique without considering them, thus keyword

research perpetuates in current studies. In the current Web marketing, SEO have dominated profoundly especially among small businesses, due to the reduced cost factor involved. Google, Bing and Yahoo!, as the current search engine giants, have a large impact on the way in which SEO practitioners design their websites. They should further adhere to the principles that govern the existence of their website of the index of these respective search engines although not all of them respect this.

Based on the academic literature explored, it is clear that there are a number of questions that are left unanswered as far as the spamdexing problem solving is concerned, as similar problems have been evolving since the last decade (Yung 2011:15), when SEO was first introduced. The difference between black hat, white hat and gray hat techniques and where a SEO technique overlaps to spamdexing (Castillo, Chellapilla and Davison) has formed part of debates, even though the problems remain.

Spamdexing is one of the main challenges Web search engines need to address, since it weakens the user's trust with the search engine rather than only the deterioration of the quality of results (Henzinger, Motwani and Silverstein 2002). According to Abernethy, Chapelle and Castillo (2010), a non-spamdexing website rarely links to a spamdexing website, even though spamdexing websites regularly link to non-spamdexing ones. This should be noted in order to avoid unnecessary waste of a substantial amount of computational resources in the search engine.

Conclusively, adequate content is one that is characterised by enough keywords that satisfy both the targeted end-user and the content value acceptable by search engine crawlers. The problem of attracting visitors has become more challenging due to the fast growth of websites. New techniques are always being implemented by website designers and content providers to enable their websites to be found, especially through popular search engines like Google, Yahoo! and Bing. This is particularly true with malicious website developers whose techniques are to achieve undeserved, high rankings by exploiting algorithms used by search engines (Shin, Gupta and Myers 2011).

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1. INTRODUCTION

It is not sufficient to simply add keywords and submit a website to search engine indices and directories and expect major improvements to a website. The previous chapter emphasised the importance of identifying keywords and the ability to use them correctly and appropriately in a search engine friendly manner, as well as them being acceptable by human readers.

Different scholars have opposing opinions about keyword density in the body text of a website, but most of them agreed that keyword stuffing is a waste of time and effort. The researcher, however, did not find any empirical results to define the crossover point of keyword rich website text to spamdexing and has further investigated and determined that different search engines have different indexing strategies.

In articulating the research project, the researcher used a mixed methodology with an experimental design as the backbone of the study, and it takes an objective and detached epistemological stance. The researcher implemented triangulation and gathered results from the following:

- search engines,
- SEO practitioners and website designers and
- experiments.

The researcher compared the results gathered from search engines and SEO practitioners with the outcome of the experiments and then based a conclusion on the experimental outcome.

During the interviews, the researcher supplied comprehensive information with a layout of the research question and sub-questions, as well as the details of the experimental design. The responses provided by the SEO practitioners and other factors relevant to the research methodology were also explored.

3.2. RESEARCH QUESTIONS

The author identified the core research problem to be:

No consistent guidance exists on search engine interpretation of the difference between keyword rich website text and spamdexing, which could result in legitimate websites being blacklisted.

This research is based on the following research question:

How do search engines interpret natural language text as being keyword rich or spamdexing?

In order to answer the above research question, the following sub-questions were also identified and investigated:

- What is the view of search engines on the definition of spamdexing?
- What is the view of academic experts on the definition and understanding of spamdexing?
- How can search engine interpretation of webpage content be measured?
- How can keyword density and subsequently “keyword richness” and/or “keyword poorness” be measured?
- How does a Web developer know if webpage content was interpreted as search engine spamdexing?

3.3. ONTOLOGICAL STANCE

Ontology is the theory of objects and their ties. Ontology provides criteria for distinguishing various types of objects (concrete and abstract, existent and non-existent, real and ideal, independent and dependent) and their ties (relations, dependences and predictions) (Corazzon 2009).

Nel and Com (2007) define ontology as the precedence of epistemology, and Knowzit (2009) further states that it is the characteristic of matter concerned with the ‘how’ rather than the ‘what’, as is the case with epistemology. In this case, the researcher will concentrate on the target prior to gathering information on it and this will include SEO practitioners, SEO scholars and search engines.

SEO practitioners: these are the people involved in the process of improving website visibility that includes designing or modifying websites in order to improve SERP ranking.

SEO scholars: these are people who research and publish literature on search engines, keywords density and spamdexing.

Search engines: these are services that allow a user to enter a keyword or phrase to search for information, and display results on its SERP.

3.4. EPISTEMOLOGICAL STANCE

Epistemology is the study of knowledge and justified belief. As the study of knowledge, epistemology is concerned with the following questions: What are the necessary and sufficient conditions of knowledge? What are its sources? What is its structure, and what are its limits? (Steup 2005).

Epistemology is a characteristic of knowledge (Knowzit 2009) that is related to the study of knowledge, how to obtain it and how to reason (Nel and Com 2007). With epistemology, evidence is required to substantiate findings in order to show that they are more than an opinion. On the other hand, ontology does not have to be proven as it is evident: it is what already exists (e.g. the Internet).

In this research project the study consists of elements that are established, indicating that the researcher is able to take an objective and detached epistemological stance.

3.5. CONCEPTUAL FRAMEWORK

A conceptual framework is defined by Perez and Anthony (1995) as a map of concepts and their relationships. More specifically, it describes the factors of significance of the research (entities) and characteristics of and associations between pairs of those factors of significance (relationships). Figure 3.1 indicates the conceptual framework for this study.

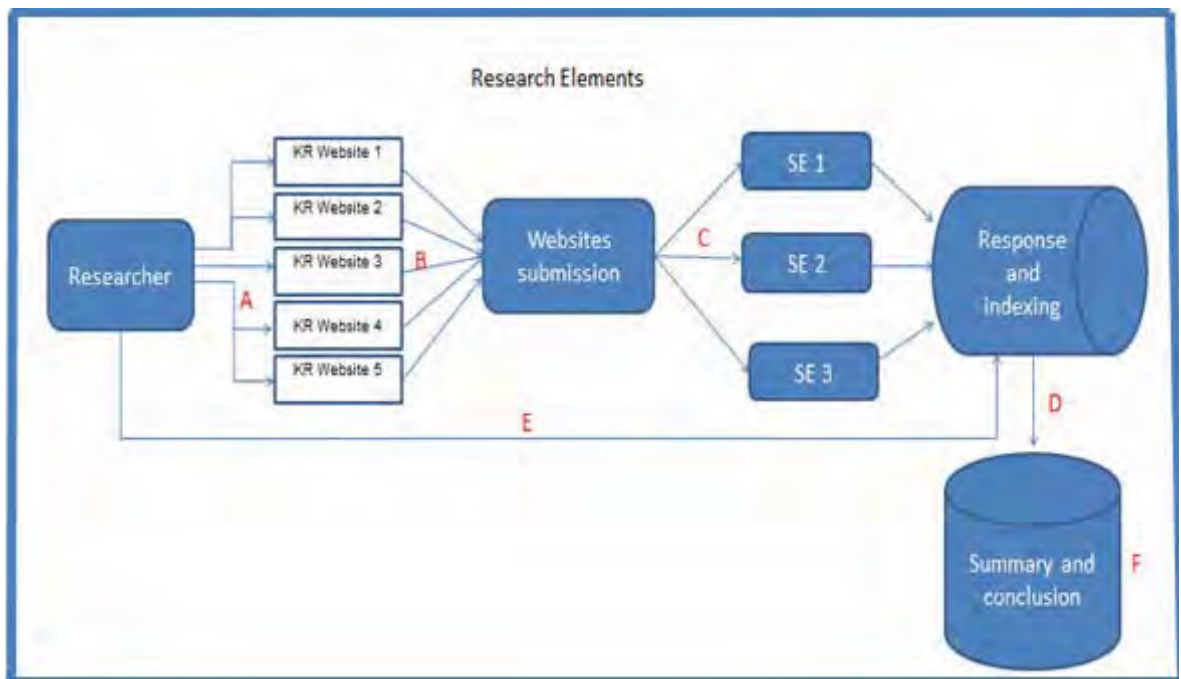


Figure 3.1: Relationship between various research elements.

Figure 3.1 shows different entities in the research study and how these elements relate to each other. There are five sections, A - E, that represent different activities at each level.

Key

A – The researcher designs, modifies and/or maintains a keyword rich website text.

B - Five test websites exist with similar content but varying keyword densities.

C - All five websites to be submitted to three search engines (Google, Yahoo! and Bing).

D - Depending on the SE algorithms, webpages are either indexed or blacklisted as spamdexing.

E - The researcher inspects the indexing of the webpages.

F - The researcher analyses findings, summarises conclusions and make recommendations.

KR Website - Keyword rich website text

3.6. RESEARCH METHODOLOGY

Research is a process of collecting, analysing and interpreting information to answer questions. However, to meet the requirements of research work, the process must have certain features. These include that it must be controlled, rigorous, systematic, valid and verifiable, empirical and critical (Kumar 2005).

In order to answer various questions that constitute the study, the researcher explored multiple methods, procedures and models of research methodology which would help to achieve best results to the research objectives. After studying various methods, the researcher chose the best method that would allow effective evaluation of data collected leading to a well-informed, reliable and validated conclusion.

3.6.1 Quantitative research

Quantitative research is where the researcher explores relationships using numeric data. The survey is generally considered a form of quantitative research. Results can often be generalised, however, this is not always the case.

Quantitative research seeks explanations and predictions that will generalise to other persons and places. The intent is to establish, confirm or validate relationships and to develop generalisations that contribute to existing theories (Leedy and Ormrod 2010). The quantitative research method works with large, representative samples and it uses structured data collection to obtain a general conclusion to the phenomena (Leedy and Ormrod 2005:183; Thomas 2004:22; Struwig and Stead 2004:4). However, this study does not involve a large sample and the researcher could not fully use quantitative methods exclusively.

3.6.2 Qualitative research

Qualitative research is a free-form research technique that is used to gain insight into the underlying issues surrounding a research problem by gathering non-statistical feedback and opinions rooted in people's feelings, attitudes, motivations, values, and perceptions, often from small samples, also known as soft data. Qualitative research

leads to answers that are typical for a specific context, and explains what makes the phenomenon different from others (Stenbacka 2001:551).

Leedy and Ormord (2010:136), state that qualitative researchers seek a better understanding of complex situations and the work is sometimes exploratory in nature, and they may use observations to build theories from the ground up. Qualitative research allows open mindedness, so as to interact with the participants but categories emerge from the data leading to information patterns and theories that assist in explaining the phenomenon of the study.

According to Peshkin (1993:23-29), qualitative research studies typically serve one or more of the following purposes:

Description: Demonstrating the nature of certain situations, settings, processes, relationships, systems, or people.

Interpretation: They enable a researcher to:

- attain new insight about a certain phenomenon,
- improve new concepts or theoretical perspectives about the phenomenon and/or
- discover the problem that exists around the phenomenon.

Verification: Enables the researcher to test the validity of some assumptions, claims, theories, or generalisations inside the real-world contexts.

Evaluation: They provide the capacity through which a researcher can evaluate the effectiveness of certain policies, practices, or innovations.

The aims of qualitative research include establishing the socially constructed nature of reality, to stress the relationship between the researcher and the object of study, and to emphasise the value-laden nature of the inquiry. The advantage of qualitative research is that it uncovers the underlying motivations for people's behaviours, attitudes, opinions and perceptions. The disadvantage of qualitative research is that the results cannot be generalised to the wider population of interest but should be used as a guide only.

3.6.3 Triangulation

According to Leedy and Ormord (2010), numerous sources of data are gathered with the expectation that they will all merge to support a particular hypothesis or theory. This approach is especially common in qualitative research. For instance, the researcher looked for common themes and conducted in-depth interviews with SEO practitioners to determine their viewpoint on keyword stuffing as perceived by search engines.

According to Olsen (2004:3), triangulation in research is defined as the collaboration of data or methods in order that different viewpoints or standpoints can shed light on a topic. The author further states that the mixing of data types, known as data-triangulation, is often thought to assist in validating the claims that might arise from an initial study. The researcher selected triangulation as this would enable the researcher's results interpretation to gain an increased reliability and to provide more in depth picture gathered from unified sources of the questionnaire, interviews, literature and experiments. This would further derive a concrete and validated conclusion. With triangulation, both quantitative and qualitative data was collected to answer the research questions.

3.6.4 Sampling

A sample is any subset of the elements of the population that is obtained for the purpose of being studied. The process by which elements are drawn from the population is known as sampling (Fox and Bayat 2007:54). The researcher identified three entities where sampling was utilised in the study, namely:

- SEO practitioners in Cape Town,
- websites and
- search engines.

The size of a population usually makes it impractical and uneconomical to involve all the members of the population in a search project (Mouton 2001:83). In this regard, the researcher depended on the data obtained from a sample of the population. The researcher considered the following in determining the size of the sample:

- the level of certainty of collecting the required information from the representing group,

- the margin of relevance of the information and the accuracy of the data to be collected from the sample size,
- the analysis to be made in respect of the data to be collected and
- the total size of population for which the sample had to be drawn.

3.6.5 Types of sampling

The fundamental significance of sampling is the representativeness of a population the researcher aims to include. Blanche, Durrheim and Painter (2006) identified the following types of sampling:

- Convenience Sampling
 - selecting participants who are available, without any prior rationale,
 - non-representative (cannot generalise) and
 - used in experiments, where (universal) processes are supposedly examined.
- Random Sampling
 - every case in the population has an equal chance of being (randomly) selected,
 - representative (generalise) and
 - used in surveys.
- Purposive Sampling
 - cases selected for theoretical reasons (good examples of the phenomenon) and
 - used in qualitative research.

In this regard, the researcher utilised the purposive sampling as listed in Table 3.1.

Entity	Sample Size
Search Engines	3
Websites	5
Search Engine Practitioners (Interviews)	5

Table 3.1: Sample size for identified entities.

3.6.6 Interviews

According to Angrosino (2007:42), interviewing is a process of guiding a conversation so as to collect information. The same author further claims that the hallmark of the observational research is to record details in as nearly a descriptive manner as possible, avoiding interpretations and inferences, and setting aside one's own preconceptions. Interviewing grows logically out of observation. In this sense, the researcher embarked on this form of open-mindedness nature of data gathering with the intention of collecting meaningful information and being able to explore the results in the questionnaire (see Appendix B for the questionnaire).

The researcher telephonically requested an interview with the selected individuals. Upon acceptance of the offer, the researcher sent confirmation e-mails see; Appendix A.

The researcher conducted interviews with five SEO practitioners, based on a questionnaire, through a personal interview. The structured interviews focused on three Web search engines, Google, Yahoo! and Bing. The purpose of the interview was to determine if the abovementioned professionals could identify the keyword density and distribution, and when a search engine regards a website as a spamdexing. Also included in the questionnaire was a section that required the interviewees to comment on the five experimental websites and how they thought the search engines would interpret them. The information gathered was summarised in relation to questions asked and the answers fall under the following headings:

- SEO facts,
- the SEO practitioner's beliefs and perspectives about keywords and spamdexing and
- present and past information.

All the interviews conducted were recorded upon interviewees' acceptance for the sessions to be captured. The researcher chose to use the interview method due to the following reasons:

- it provides instant feedback from the respondent,
- the opportunity to obtain clarity and explicit expansion on certain facts and ideas and

- it had the advantage of a visual aid provided by the structure of the five websites.

3.7. INTERVIEWS WITH SEO PRACTITIONERS

The five selected interviewee companies were selected due to the fact that they were trusted representations of the SEO industry. The researcher drew this sample to represent the entire population of the body of SEO because of:

- their ability to conduct business in a trusted manner and
- their experience with the application of SEO principles.

For the sake of confidentiality purposes, the company and individual names will not be divulged in this study; however, the researcher has assigned random IDs as follows:

Company A (Interviewee A),
Company B (Interviewee B),
Company C (Interviewee C),
Company D (Interviewee D) and
Company E (Interviewee E).

3.8. SEARCH ENGINES' POLICIES

The researcher collected information from search engine Webmaster guidelines in order to determine how the search engines, as described in their policies, interpret the keyword rich text website and how it defines spamdexing. The researcher further sent e-mails to three individuals working for Google, Yahoo! and Bing with similar questions to the ones addressed to the SEO practitioners. This was intended to obtain clarity on information provided on respective Webmaster guidelines.

3.9. EXPERIMENTAL RESEARCH

According to Fox and Bayat (2007:10), an experimental research approach aims at forecasting what may occur or, otherwise, intends to bring together changes or new approaches within the prevailing situation in order to determine the outcome. The researcher decided to include an experiment as a means to evaluate the unclearly presented elements of keyword richness in website text and where keyword density will have an effect on keyword stuffing. The researcher controlled other influential factors except for keyword density.

3.9.1 Phase 1 website structure

The five experimental websites were designed in simple HTML (Hyper Text Mark-up Language) with no Flash files so that the crawlers would easily visit them (see Appendix G for the sample code). Consistent, relevant and similar content was placed on every website to meet all the appropriate white hat regulations with the exception of keyword density. The five websites are commercial sites that provide information about various second hand laptops sales as well as laptop accessories. The research is centred on the homepage whilst the other pages were intended to provide an increased site content that would help during crawler visitation. The website consisted of three pages, namely:

- the home page,
- the catalogue page and
- the contact page.

The catalogue page and the contact page had similar but different content and were designed to increase the website textual content that would assist in terms of importance of the site to the user. This would further improve the webpage's chances of being indexed and favoured by crawlers. The home page for each website was the page with varying keyword densities and distribution, whereas the word count and content was similar. There were no underhand techniques that could lead to any of the sites being regarded as spamdexing. See Appendix D for the layout of the five websites.

3.9.2 Homepage keyword and content structure

The keyword density was carefully chosen and varied between low (well written English text) to high, keyword stuffed text. Table 3.2 shows the various keyword densities.

WEBSITE	WORD COUNT	TARGETTED KEYWORD	KEYWORD COUNT	KEYWORD DENSITY
www.getlaptops1.co.za	330	laptops	13	3.94%
www.getlaptops2.co.za	330	laptops	20	6.06%
www.getlaptops3.co.za	330	laptops	28	8.48%
www.getlaptops4.co.za	330	laptops	40	12.12%
www.getlaptops5.co.za	330	laptops	90	27.30%

Table 3.2: A summary of Phase 1 websites content and keyword densities.

3.9.3 Websites domain registration

All five experimental websites were registered with UniforumSA and hosted by Hetzner South Africa. They were all registered on 12 September 2010 and the files were uploaded on 06 October 2010.

3.9.4 Websites submission to search engines

The websites were simultaneously submitted to Google, Bing and Yahoo! and the researcher investigated how these search engines indexed each site. A daily check was then carried out to establish when, if at all, these sites were indexed. The submission was done using the consoles in Figure 3.2, Figure 3.3 and Figure 3.4, respectively.



Google Add your URL to Google

[Home](#)
[About Google](#)
[Advertising Programs](#)
[Business Solutions](#)
[Webmaster Info](#)
Submit Your Site
Find on this site
Search

Share your place on the net with us.

We add and update new sites to our index each time we crawl the web, and we invite you to submit your URL here. We do not add all submitted URLs to our index, and we cannot make any predictions or guarantees about when or if they will appear.

Please enter your full URL, including the `http://` prefix. For example: `http://www.google.com/`. You may also add comments or keywords that describe the content of your page. These are used only for our information and do not affect how your page is indexed or used by Google.

Please note: Only the top-level page from a host is necessary; you do not need to submit each individual page. Our crawler, Googlebot, will be able to find the rest. Google updates its index on a regular basis, so updated or outdated link submissions are not necessary. Dead links will fade out of our index on our next crawl when we update our entire index.

URL:

Comments:

Optional. To help us distinguish between sites submitted by individuals and those automatically entered by software robots, please type the squiggly letters shown here into the box below.

Add URL

Figure 3.2: Google site submission console (Source: Google 2010c).



Figure 3.3: Bing site submission console (Source: Bing 2010b).



Figure 3.4: Yahoo! site submission console (Source: Yahoo! 2010b).

The researcher opted to use different search engines due to the difference in the algorithms they use. Consequently, a site ranked highly on Yahoo! may not necessarily be ranked as high or indexed in Google's search results.

Keywords were placed in the body text and headers. The researcher decided to select body text, but not title tags and anchor tags, as a result of the importance of

body text in websites (Weideman 2009:55). The limitation was also done to reduce the effects of other variables that would make the research unmanageable and uncontrolled. The researcher utilised the body text in order to reduce the complexity of the research, as well as to provide a control element in this research.

3.10 PHASE 2 EXPERIMENT

Figure 3.5 is an adaption of Bunge's model (Bless, Smith and Kagee 2006) of the testability of scientific statements. The researcher applied Bunge's model to the scientific claims to prove the validity of procedures and theoretical assumptions that led to the support of this empirical research.

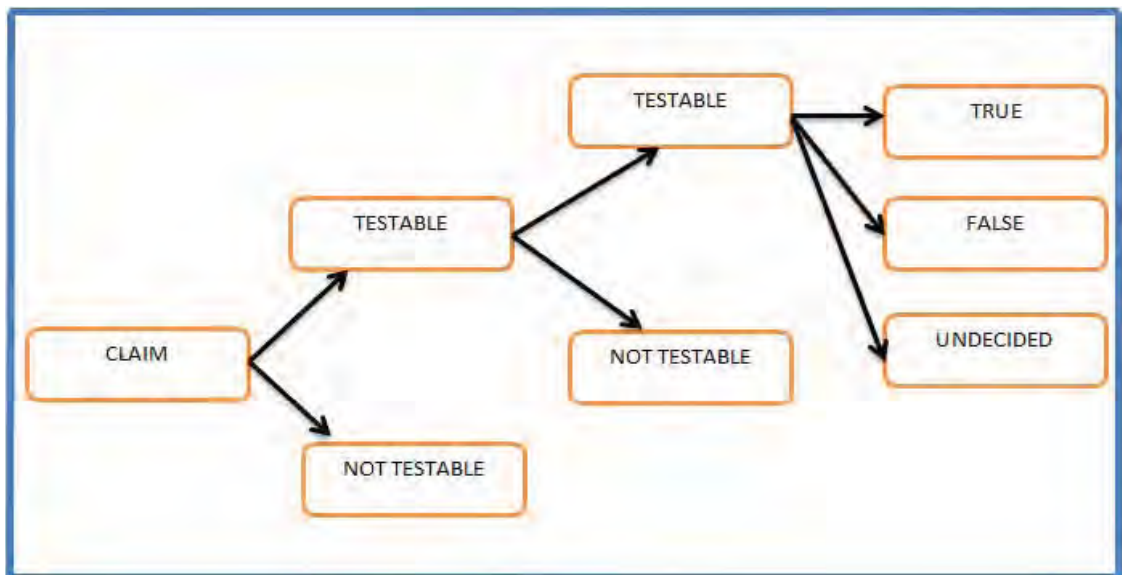


Figure 3.5: Model of partitioning of scientific and non-scientific claims (Bless, Smith and Kagee 2006:8).

After collecting the experimental results certain anomalies were noted. As a result, the researcher decided that the first experiment was to be considered as a first phase (Phase 1) and to extend the work to include a second experiment (Phase 2). This enabled the researcher to design a second phase of the experiment with intentionally highly keyword densities. The researcher made this decision as the theoretical claim was tested and the result was false. The following were changed in order to reinvestigate the response of the search engine crawlers:

- the keyword density of all the websites, with the fifth website carrying close to the maximum keyword density with “laptops” as the keyword and
- a fourth page per website was added to attract crawlers’ attention to the sites.

The websites were resubmitted simultaneously to Google, Bing and Yahoo! Table 3.4 shows the Phase 2 summary of the website keyword density and other related information.

WEBSITE	WORD COUNT	KEYWORD	KEYWORD COUNT	Phase2 Keyword Density
www.getlaptops1.co.za	330	laptops	100	30.3%
www.getlaptops2.co.za	330	laptops	132	40%
www.getlaptops3.co.za	330	laptops	170	51.52%
www.getlaptops4.co.za	330	laptops	232	70.3%
www.getlaptops5.co.za	330	laptops	321	97.27%

Table 3.3: A summary of Phase 2 websites content and keyword densities.

To avoid content duplication, all Phase 1 website files were deleted from the hosting servers and Phase 2 files were uploaded. The submission to Google, Yahoo! and Bing followed the same procedure as Phase 1. The submission was done on 13 December 2010 and the recording process commenced on the second day of submission (see Appendix J). The Phase 2 experiment spanned 67 days of monitoring, as was the case with Phase 1.

3.11 SUMMARY

The significance of this chapter is in identifying the research question and its sub-questions and providing a clear indication of how the research was designed and articulated. In this chapter the researcher discussed the methodology used and provided an analysis of why certain methods were used. The triangulation method formed the basis of the study and this was selected in order to prove the validity of the claims identified in Chapter 2 regarding keyword stuffing and penalisation of websites. The results and analysis of the research is covered in Chapter 4.

CHAPTER FOUR

RESULTS AND ANALYSIS

4.1 INTRODUCTION

Having assessed the measurable facts on keyword density and spamdexing interpretation in the preceding chapters, this chapter aims to collectively gather the responses and review the findings from the sources used. In this chapter, data will be analysed according to the categories specified below:

- SEO practitioners,
- e-mails sent to SEO experts,
- search engine policies and
- data collected from five websites under the experiment.

4.2 RESULTS FROM SEO PRACTITIONERS

Five participants representing five well-known companies in Cape Town were involved in the interviews conducted. A structured questionnaire (See Appendix B) was used during the interview and the researcher summarised the responses. The participants' responses were recorded here as reported speech in a simplified and summarised manner and the response meanings were maintained. Responses were recorded with the questions asked and interviewees' responses thereafter as shown by the section below. Each question is represented by the letter "Q" and interviewees' responses with the letters "A-E".

4.3 INTERVIEW QUESTIONS AND RESPONSES

The section below shows 15 questions and the subsequent responses received from the interviewees. A brief analysis based on the information obtained has also been included.

4.3.1 Q: How many years have you been in the SEO industry and why did you consider being in this field?

A: The interviewee has 11 years experience in the industry. He noticed a gap in the market and was interested in the ranking of websites.

- B:** The interviewee was in an intensive research field on keywords for seven years and has a year of practical experience in the industry. The interviewee is still in the learning process and has found that there is a big difference between research and the practical implementation of ideas. He unintentionally entered the field.

- C:** The interviewee has two years experience in the SEO industry and at least six years in the search marketing industry. She started with PPC and then SEO and was interested in the online industry.

- D:** The interviewee has been in the industry for at least seven years; he started with graphic development, then moved to software development and analysis and is now stationed in SEO. He was very interested in business requirements analysis and meeting users' needs.

- E:** The interviewee has been in the industry for eight years. His entry point was a BTech with a journal paper that examined the keywords demographics and how it affects the search. For example, how many keywords an end-user can use for a search, more keywords, fewer keywords and based on the keyword what sort of results a user will obtain. The initial interest point was artificial intelligence/robots.

The above question was designed with the motive of gathering information about the experience of the interviewee, as well as determining if the interviewee had time to explore various SEO techniques. The researcher expected that the higher the number of years of experience an individual had in the industry, the better the quality of information they would share. This was evidenced by the way that they answered questions and examples were given in support of their facts. They provided practical knowledge, coupled with theoretical background information. This question also assisted the researcher in obtaining an overview of Web developers and SEO practitioners' understanding of their industry. The question further supported the sub-questions of this research. See Figure 4.1 for a summary of the interviewees' years of experience.

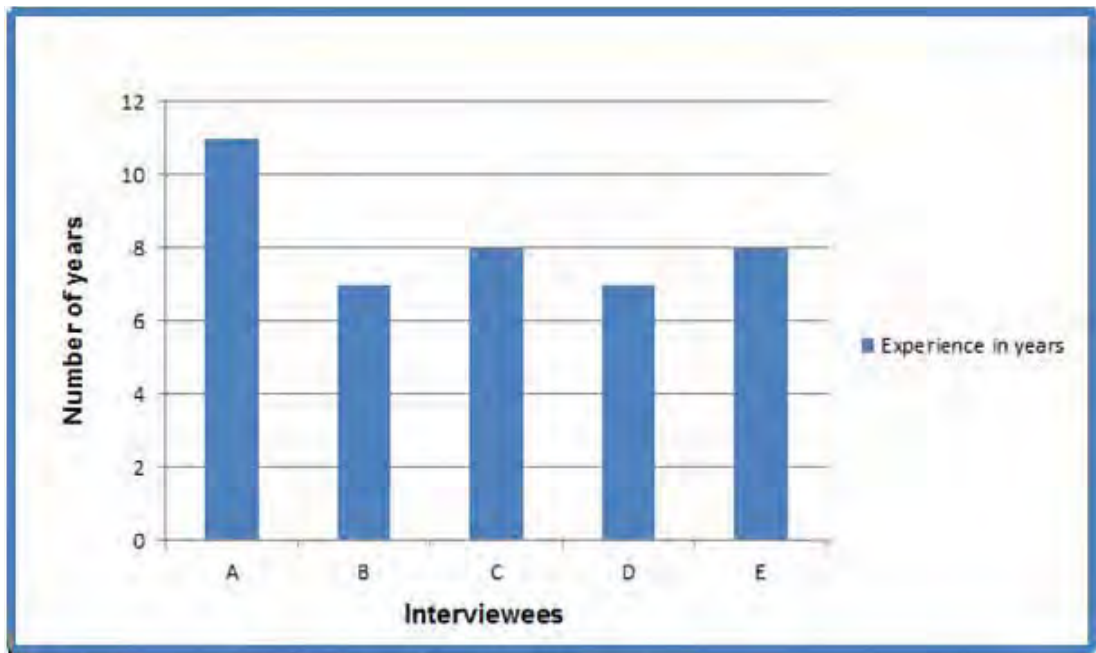


Figure 4.1: Interviewee SEO industry experience in years.

4.3.2 Q: What SEO tools do you regularly use and how do they work?

A: They use a variation of tools such as IBP SEO software. They also utilise commercial and non-commercial software, as well as experience.

B: There are three tools that they use:

- Google Analytics to track all the websites that they implement as they consider it very useful to track individual pages rather than the whole website, as well as optimising every page.
- Google Adwords (which provides suggestions based on the keywords that one enters) to locate the keywords to focus on and use it for decision making processes.
- Internet Business Promoter - to perform ranking checks (e.g. 25 keywords). They normally use 6 search engines, namely, Google, Yahoo!, Bing, Ananzi, Google SA and ASK.

C: They utilise Open source tools, Website Grader, Spider Claws, Webmaster tools, Site Explorer, Driving Tools, and Web SEO.

D: They make use of both commercial and open source software.

E: They utilise a number of tools and most of them are free (e.g. Google Adwords and Trends Site). Their company prioritises the use of tools supplied by search engines, rather than many other tools which are not relevant to the search engine as SEs are not third parties.

The above question was aimed at identifying whether the strength, judgement and decision making processes of SEO practitioners was centred on tools, literature facts and/or experience. The researcher classified the tools used as Commercial software, Open Source software and Individual experience. The interviewees' rated each software category according to how they use it. The responses were summarised (as percentage tools utilisation) in a graphical form in Figure 4.2.

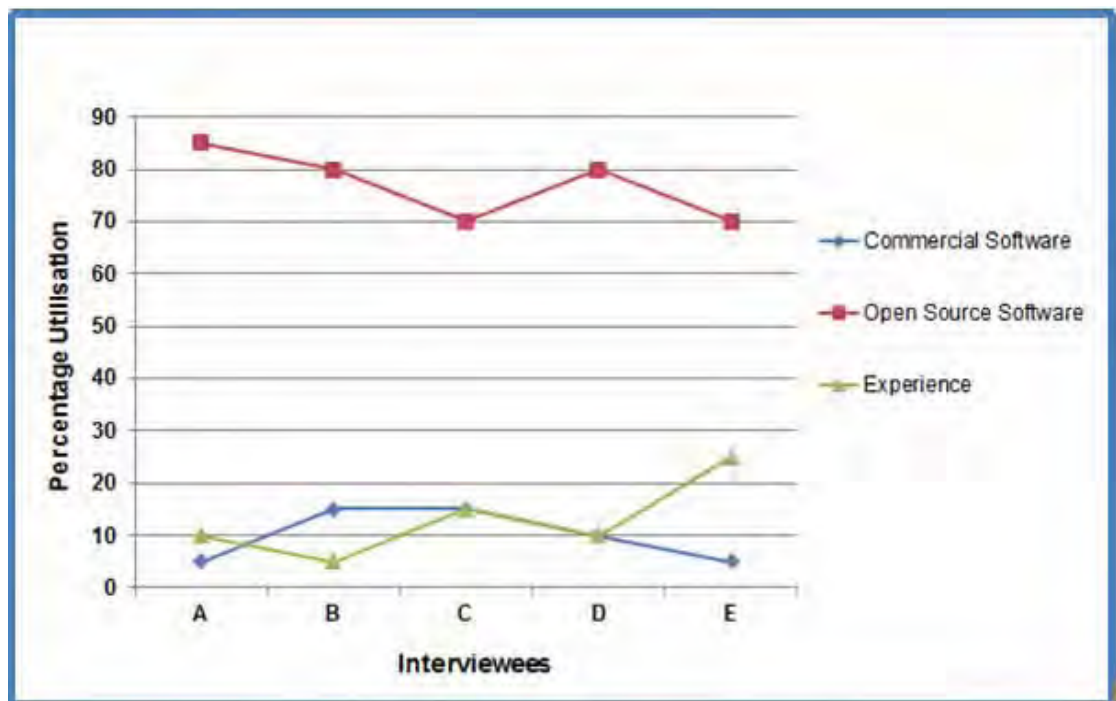


Figure 4.2: SEO tools used by interviewees for decision making process.

4.3.3 Q: What areas do you think are currently the most important in organically ranking a site?

A: The interviewee considers a combination of factors, ranging from relevant keywords, Meta titles and many others. Their biggest aim is to understand a product, the objective of the product and what the customer is trying to achieve and whether they require exposure, cost back position or other factors. Once they understand what they want to achieve with the campaign, they look at consumer behaviour, as well as what people are searching for.

They believe that marketing should be more focused on:

- I. On-Page (Is the metadata correct, is the content focused on the targeted market?).
- II. Off-Page (Are there any social campaigns on the site; is there communication of the brand on the Internet?).

B: The interviewee thinks that it is the domain name.

C: The interviewee considers keywords to no longer be high ranking for the search engine, but links as well as social media have become more important with community pages such as Facebook and Twitter. However, they are of the opinion that keywords are important in terms of relevance and that there are other ranking factors such as the age of the domain and the number of links to a site. Furthermore, the interviewee stated that keywords are no longer ranking factors; it is imperative to run a set of keyword analysis tools to verify the relevance of the keywords to the content.

D: The interviewee prioritises what the developer wants to achieve with the site, whether it is to sell something, make people contact them, run an e-commerce business or trying to determine if the business is in line with the strategy. Once the aforesaid is established, they use a set of tools to recommend and reach an objective conclusion. They also consider measures that mitigate bounce rate, load time and user experience; for example, a fast loading page achieves better ranking than a slow loading page.

Two other aspects that they consider to be important are:

- semantic relevance (is one using singular or plural words or are they using other words that describe their keywords; e.g. dog, dogs, canine, etc.) Use of a lot of variation along the keyword is recommended.
- architectural links: Is one using text links from other websites linking with a keyword? Out page content should be humanly understandable (links to one's site need to portray the same thing;

e.g. if selling a BMW and someone links the site as Mercedes).
The user should link a site with the same keywords.

Also, in their opinion there is more traffic to a page if there are more links to that page, which is link strategy (e.g. they can have either book quality or high quality). They also believe that collective site authorities are crucial and the page must be semantically correct and should get the right word to repeat in the correct syntax.

E: The interviewee also believes that content comes first and keywords second. In terms of organic listing, Google is governed by an artificial intelligence, a program that attempts to learn from what it sees and how people are responding based on statistics (e.g. how does it interpret an apple being a computer?). This is achieved by content and by keywords.

They mentioned that external linking was important but there is a need to focus on basics or well-structured good foundations based on artificial intelligence interpretation.

They have witnessed several other SEO practices that do not consider keywords as important, but do not necessarily disagree with that in terms of their methodology as they might be using a different approach. Their approach might be of such a nature that they base it on a link environment. They further believe that there might be other factors they could be considering, based on their methodology on a specific strategy.

This question was intended to provide the researcher with information leading to the identification of the importance of keywords in websites. The question also formed part of the research questions and led to some information relating to SEs to be reviewed (e.g. search engines' interpretation of keywords in a website text). The researcher's intention with the rating of keywords was supported by all the interviewees, even though there were a few differences on placement. Content was mentioned by all the interviewees as the key to indexing and ranking.

4.3.4 Q: If a submitted site is not showing up on SERP, does this mean it has been banned and how long does it take for a site to be indexed and show up on the SERP?

A: The interviewee stated that there are many determining factors that will result in the site ranking high, for example, the uniqueness of the keyword or key phrase used (e.g. blue sun glasses Cape town can rank higher than sun glasses). They recommend that developers also consider link generation on the site and the age of the site. According to them, indexing takes 14 days, sometimes one day, but the longest should be less than two months.

B: The interviewee thinks that if a site does not appear on the SERP, it does not necessarily mean it has been banned. They believe that there can be other contributing factors that may be of a technical nature. Their longest waiting period was three months, but they pointed out that it may take between a few weeks to three months.

C: The interviewee is of the opinion that a site not appearing on the SERP will not have been banned, but this may be caused by other technical reasons that caused the SE not to look at a developer's webpage (e.g. robot.txt file can make a developer ban their own site). Their advice to website developers is to make the architectural structure reflect whether or not the site is inaccessible by robots (e.g. the site might contain unfamiliar JavaScript, Flash or might have meta refresh or redirects where the site loops. This will result in the domain jumping to a different loop and they will resolve to a different page that the SE cannot see). They also advised Web developers to ensure that sites have verification codes for the Webmaster tools so that they can immediately see what SEs can see.

They believe that from the moment a site is launched it should take a day.

D: The interviewee cited that numerous developers make use of the robot.txt file to notify the SE that it is not ready to be indexed (and the site does not get indexed until the file is removed) and they often forget this piece of code. However, one will be instructing Google not to index the site and this even happens with big clients. When the abovementioned occurs, the site might not appear in the top 10 or first or second page, but it does not mean that it is not indexed; it could appear on the tenth page. The interviewee also pointed out

that there are operator tools that can be used to fully examine whether the site is not indexed, is not there or is not yet ranked. The interviewee also mentioned that after the submission of a sitemap it takes no longer than 24 hours for a site to be indexed.

E: The interviewee thinks that an incorrect submission can result in a site not appearing on a SERP and if a site is submitted correctly it can take two to three days. The robot.txt file communicates directly with the search engine as the user agent. The interviewee is of the opinion that the objective should be to satisfy the user in terms of usability and that most methodologies rank the homepage first and then the other pages.

This question provided the researcher with information regarding the indexing of websites by SEs. There were different views in respect of the time it could take for a website to be indexed. This was characterised by time ranging from a day to three months if the correct procedure of submission was followed. Various reasons for a site to be penalised were also explored.

Figure 4.3 indicates the opinions of the five experts on the time it takes to get a website indexed.

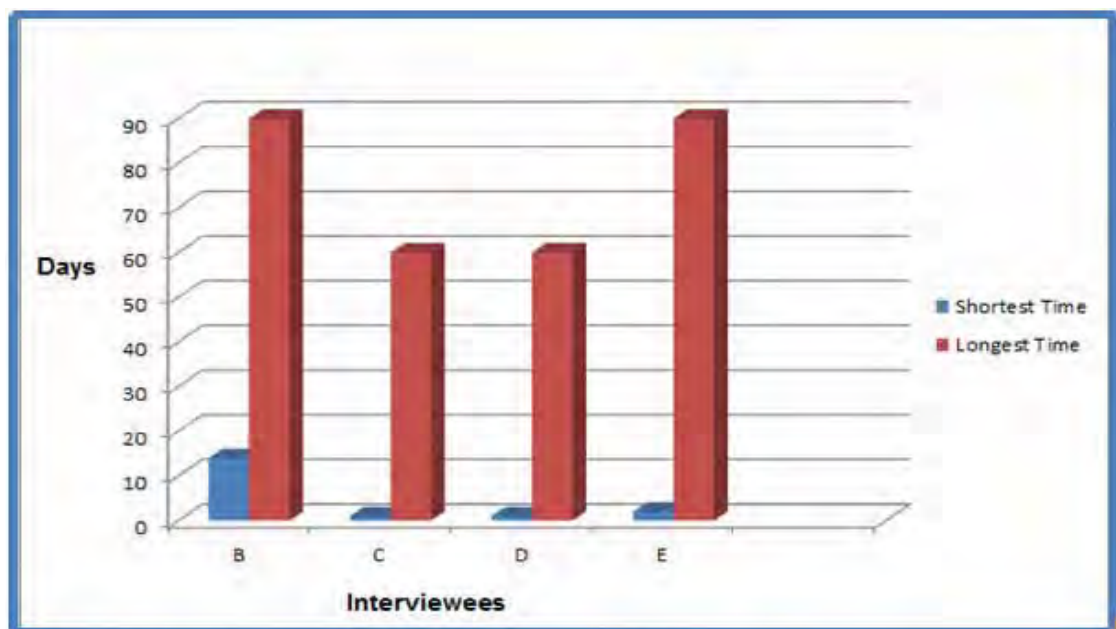


Figure 4.3: Minimum and maximum indexing time as stated by each interviewee.

4.3.5 Q: Should one optimise for the singular or long tail form of keywords and why?

A: The interviewee regards the keyword phrase as the most suitable, as it explicitly defines a specific market whilst a singular keyword has the following limitations:

- high bounce rate and
- irrelevant traffic being driven to the site.

B: The interviewee recommends the use of a key phrase rather than a single keyword, since single keywords are highly competitive.

C: The interviewee thinks that key phrases are better than a singular keyword.

D: The interviewee suggests the use of a key phrase due to the fact that singular keywords lack understandable descriptions and may easily be interpreted wrongly by SEs and results in a high bounce rate.

E: In the interviewee's opinion, a one term keyword has become more generic and competitive and very difficult to rank. The interviewee also cited that to obtain more appropriate results, users are now searching using long tail ranks thus 2-3-4 terms per search hence SEO practitioners are pushed to try to rank for long tail terms/ geo focus. Generic terms are not only very complicated to rank, they are also incredibly competitive.

The main focus of this question was to determine the differences and/or advantages of utilising a singular keyword or a key phrase as a ranking keyword strategy. The researcher noted that all the interviewees agreed that utilising a single keyword is highly competitive and more difficult to rank than key phrases. Furthermore, a key phrase provides a more detailed description and is sound in terms of providing a clear understanding of the search to the SE.

4.3.6 Q: What do you understand by the phrase keyword stuffing?

A: According to the interviewee keyword stuffing is a form of cloaking, which is practiced by many website developers. The interviewee believes that a product should be attractive from a reader's perspective and that anyone

practising cloaking is foolish and cannot get away with it although it is a common practise.

- B:** The interviewee indicates that there are no guidelines, but articles define it as repeating keywords and if it is done the website is banned.
- C:** The interviewee defined it as a form of spamdexing whereby an individual repeats keywords in a website in an attempt to cheat the SE.
- D:** The interviewee described it as a deliberate procedure to deceive SE, achieved through stuffing keywords in a website in order to earn a higher ranking on SERP.
- E:** The interviewee knows that if it is carried out a website will be penalised by SE.

The researcher included this question in order to find out if the SEO practitioners understand what keyword stuffing involves and how it impacts on the indexing and ranking of a website. Various definitions were supplied and the focus was on Web designers attempting to trick the SEs with the intention of obtaining a better ranking; they all believed that penalisation from the SEs is inevitable.

4.3.7 Q: How often do you carry out experiments on search engines to check if their search and indexing algorithms have changed?

- A:** They do not do experiments (They are in the business sector and time means money; they learn from others' experiences).
- B:** They do not do experiments.
- C:** They do not carry out experiments, as they deal with development, but find it easy to recognise the algorithm change.

However, the interviewee is certain that Google makes between 350-500 algorithm changes per year, which is equivalent to at least one change a day. An individual will never know what they changed as they constantly tweak the algorithm.

So how do you respond to that?

According to the interviewee, SEO is a guessing game and by checking with the different pages, developers can see some of the factors that have become important to Google and follow the webmaster forums where discussions are continuously taking place.

- D:** According to them they do not carry out experiments, but instead monitor traffic trends on the site. The dropping of traffic trends from the previous week or increase in traffic can be an indication of Google rating the site differently or it will drop for a specific reason (e.g. if orders drop from 100 to 50 orders a day then it's an indication that the activity on the site or the ranking has changed).

So how do you respond to that?

The interviewee recommends forums and sites that they lead to an agreement to certain practices on how to follow it; recommendations from forums means that one is 80-90% correct, bearing in mind that they are in a competitive space.

- E:** Yes, they are constantly busy with experiments.

The researcher designed this question in order to establish how often SEO practitioners carry out experiments in order to check the validity of certain facts and opinions that revolve in the industry. Also, the question was mainly concentrating on the reaction thereof when a change is assumed by a SE. Noted with great concern was that 80% of the interviewees do not experiment at all; they base their facts on other people's opinions. A great deal of focus is placed on mass production and profit making rather than validating and quality testing of strategies in use. Figure 4.4 shows a summary of the percentage of the interviewees that do experiments.



Figure 4.4: Experimenting SE algorithm change.

4.3.8 Q: Do you follow what forums and bloggers say regarding the SEO industry?

A: They do not listen to them because Google wants developers to ensure that their website loads faster in order to reduce users' waiting time. They also mentioned that SEO is becoming more specialised and that it is not easy to be familiar with every angle of it.

B: The interviewee follows and advised developers to visit blogger sites and listen to what other people are saying (e.g. Quick Company, SEO MOZ.org and synergizeit.co.za).

C: The interviewee follows them because webmaster forums have a lot of discussions going on all the time.

D: The interviewee follows forums and bloggers lead to an agreement to certain practices on how to follow it and recommendations from forums means developers are 80-90% correct, bearing in mind that they are in a competitive space.

E: The interviewee follows them due to the fact that there is a lot of information sharing in forums and blogs.

The main emphasis of the question was to find out exactly where the SEO practitioners obtain their information and how reliable the information is. The researcher determined that the main source of information is forums and blogs, as this was characterised with 80% of the interviewees being in favour of them.

4.3.9 Q: Have you ever seen a blacklisted or banned website and what does it look like?

A: They have not seen a blacklisted or banned site.

B: The interviewee believes that if one visits Google and search for top spamdexing websites or banned sites they will provide snapshots of the banned sites or the code that resulting in them being banned. The one that he observed was the hiding of keywords on the bottom of the page with the font colour being the same as the background colour. This makes it impossible for human users to see; only the crawlers notice it.

C: They have not seen a blacklisted or banned site.

D: The interviewee knows that there are blacklisted sites, although they cannot be seen on SERP, but the browser will generate a warning and mention something pertaining to the spamdexing. The warning will be visible and the site is accessible, however, visiting it will pose numerous risks.

E: The interviewee has seen a few blacklisted sites, for example, BMW Germany, who was blacklisted for cloaking and also some ISPs blacklist the URL based on other factors. The interviewee's advice to website developers who want to have authority in an industry is to have a good name with other players in the sector like ISPs.

The question provided an indication of what to expect from the experiment in this study. There was no clear and comprehensive answer as to what to expect from the search engines. All the interviewees knew something about blacklisted websites, but their knowledge was not factual, which is why it is not possible to clearly comment on what to expect from the experiment.

4.3.10 Q: How do you measure the richness or poorness of a keyword in the body of a webpage?

- A:** They do not normally go higher than 5%, because above that the page becomes unreadable and will not make sense.
- B:** They use the old rule called the 3% rule, for example, a word should appear three times in every 100 words. This is so that it does not sound strange when being read by a user. Likewise, they regard the poorness as less than 3%.
- C:** The interviewee uses a Keyword Analysis tool to analyse the poorness or richness of a keyword and avoid longer phrases or stop words and/or single words but three word phrases.
- D:** The interviewee thinks that when content on a webpage stands out to a human, it will do the same to a search engine. The interviewee further believes that anything above 5% of the keyword appearing on the page (e.g. 50 times in 1000 words) should be in order and there ought to be a variation.
- E:** The interviewee thinks that the keyword density should not exceed 5%.

Like many scholars, such as Appleton (2010), Charlesworth (2009), Todaro (2007) and several others, there was inconsistency in defining the keyword richness or poorness in a body of a website text. All the interviewees' opinions of good keyword percentage in a body text ranged from 3% to 5%. This question formed the foundation of this study.

4.3.11 Q: How do Google, Yahoo! and Bing interpret keyword stuffing and do their algorithms stick with their respective document guidelines and procedures?

- A:** According to them, they are also a Google partner and they regularly look at guidelines and read Google information and blogs. The interviewee mentioned that Google values social network, and it is much easier to rank on Yahoo! and Bing on keywords than on Google, as the aforementioned is very competitive.

- B:** The interviewee is aware that Google, Yahoo! and Bing penalise those who do not stick to their guidelines.
- C:** The interviewee stated that too many website developers game it (trick SE algorithm), which will result in keyword stuffing.
- D:** The interviewee believes that guidelines are the recipe to success as they are intended for Google to be beneficial to the developer and at the same time the developer useful to them; however, they do not provide a guarantee of ranking.
- E:** The interviewee is aware that continuous repetition of keywords in a body text will result in the SE penalising the site.

The motive of the question was to ascertain if the SEOs continuously revisit the SE guidelines and procedures and how often they adhere to them. This proved to be a difficult question in terms of Yahoo! and Bing, as all of the interviewees' answers were referring to Google rather than the other two. This meant that Google is considered and strictly followed as compared to the others, even though Yahoo! and Bing are also big SEs.

4.3.12 Q: Do you think SEO practitioners and website developers understand spamdexing?

- A:** They believe that a lot of website developers have ideas but are unable to apply them. They do not understand marketing. According to them, a website will rank highly when it has constructive reporting and is marketing oriented.
- B:** The interviewee believes that Web developers do not know what SEO is and what it does; they are only concerned about the site's looks and do not stop to consider how users get to the website. This is why they use JavaScript and Flash when designing sites.
- C:** In the interviewee's point of view, SEO has become a guessing game whereby designers pretend to know what they are doing but in actual fact wasting clients' time by promising them what they cannot deliver. They think that ranking comes from following what a SE wants rather than what designers want and a lot of designers do not know what they are doing.

- D:** In none of the websites that the interviewee accessed had SEO strategies been implemented. The interviewee believes that whilst some practitioners understand the SEO concept, they often fail to include the marketing aspect of the business; this results in clients' requirements not being met. However, different developers use different methods, hence achieving varying results.
- E:** The interviewee found the situation to be very tricky; for example, Google's algorithm consists of approximately 200 statement variables that impact on ranking. The interviewee further stated that these are randomly controlled by artificial intelligence so there is no specific time that the algorithm will change; for example, every month it dynamically takes place since it is controlled by a system.

In their company, when they speak to or interview website owners, most of them have a general idea what these variables are, but do not fully understand how to put all this together.

Typically, what they have in their company is quite unique and their theories have enabled them to maintain their clients' base and it has worked for them. Whenever they discover something they are typically 6 months behind the technology which makes it very difficult to keep up to date.

The researcher designed a question that enabled the SEOs to rate each other in terms of how much they know and understand about the field; unfortunately they unanimously agreed that many experts pretend to know what they are doing but in actual fact do not deliver according to the customer specifications. It was further agreed that no one can guarantee website rankings; only the SE can make that decision.

4.3.13 Q: Do you think there is a crossover point from keyword rich website text to spamdexing and how do you interpret it?

A: The interviewee thinks it is 12%; however, they use 5% and it is very high so he estimates 1 - 3%. They consider the end result, which is content that is rich and readable, logical and places a strong focus on the message.

- B:** The interviewee does not understand the concept or what happens if the point is crossed over; however, the interviewee believes that it is highly probable that the point exists.
- C:** The interviewee thinks that it is black hat - a lot of repetition.
- D:** The interviewee thinks that there are a lot of black hat techniques/practices and that Google is not obliged to give a developer a ranking. They believe that one can be reconsidered after being regarded as a spamdexer, should they be able to justify their actions.
- E:** The interviewee is aware that there is one but unsure of the exact point.

The basis of this research was centred on this question and likewise the answers were not clear enough to be justified. Nonetheless, they posed a clear picture of the inconsistency of the actual percentage of the crossover point of keyword rich website text to spamdexing. All the interviewees agreed that this point exists, however, 80% of them did not commit to a specific percentage, whilst only 20% said 12% should be the point.

4.3.14 Q: Are sitemaps helpful for website indexing?

- A:** The interviewee believes that they are helpful as they increase productivity and aids a developer in being proactive in updating their site.
- B:** The interviewee is of the opinion that they are helpful, since they speed up the indexing process.
- C:** The interviewee thinks that they are very helpful.
- D:** The interviewee confirmed that they are helpful and enables the SE to easily visit a site's pages. The interviewee cited that when a site map is submitted for the first time, the crawler crawls and refreshes the content. If it returns again and finds there is new content, it will revisit more frequently and will develop a rhythm where it will come back once or twice every week. However, if it finds the same content the subsequent visits will be less frequent and if there are no changes in five consecutive visits the crawler never returns. The

interviewee believes that the trick to success is to launch the sitemap and then keeping the crawler busy.

E: The interviewee thinks that sitemaps should be utilised, especially on big websites that have regular changes.

The motive of this question was to determine if sitemaps are important for site indexing. The interviewees agreed that they were helpful and enables the site to be visited numerous times; it was also the most appropriate for large sites that are regularly updated.

4.3.15 Q: Where do you think the SEO industry is heading?

A: The interviewee is of the opinion that SEO is becoming bigger and more specialised, with more companies coming online.

B: The interviewee believes that the relevance of information on search is still a problem. For this they think that there is still a lot to be improved on SEO, although they could not think of an alternative way which users may utilise in order to search for information on the Web. The interviewee thinks that it is still growing and is in its early stages.

C: The interviewee advised that there is future in SEO and that the sky is the limit.

D: The interviewee trusts that there is a future for SEO, but that would diversify; for example, there are going to be practitioners that are social media experts, video experts, focus on video production for websites, excellent in architectural and content link building matters and outsourcing consolidated products.

E: The interviewee was sceptical about SEO, as it is still developing; he mentioned that whilst a lot of people have a great deal of information they lack the understanding to make it function as one.

In order to identify how much more research needs to be done in relation to the SEO industry, the researcher designed this question. The question further focussed on future developments and areas of specialisation that need to be concentrated on. In

general, the industry is growing, which can be substantiated by considering the countless applications that are becoming Web enabled.

4.3.16 Interviews analysis

All of the interviewees had at least seven years of experience in the SEO industry. The researcher found that their careers were built up and developed through applying different strategies and coming from different disciplines. Hence, the information gathered was based on experience rather than academic literature. Approximately 80% of the interviewees have previously been involved in website design and marketing and the diversity and exponential growth of the industry led them to venture into SEO. Their responses showed a high level of tactical maturity and understanding of SEO methodologies and the evolving of the industry.

Based on their responses, it is clear that SEO tools take precedence over literature and their experience supports the tools in decision making processes. The reasons for this may be to save time and increase productivity, but it takes a while for a software tool to be updated after an algorithm change, which is why some practitioners are not on par with the technology. Figure 4.2 shows a high average percentage utilisation of 77% of open source software, compared to commercial software (10%) and experience (15%). Open software requires free licensing, hence the decrease in costs and increase in productivity.

All five interviewees claimed that content was the most important element in organically ranking a website. However, they all held different views about keyword rating. One out of the five interviewees placed keywords as the second element and they all agreed on keywords playing an important role on content relevance and the ability to show the importance of content published or the product on sale. Their differences might be as a result of the exponential changes in the focus of websites' strategies, whereby different methodologies are being implemented to meet clients' requirements. All the interviewees recognised the importance of keywords and thereby utilising them in various strategies. However, they were no longer using them as a stand-alone ranking strategy for fear of this being interpreted as keyword stuffing. Furthermore, the search engines had reduced their importance among other strategies, which is why the interviewees are not concentrating on optimising keywords. As noted above, the use of different tools in determining the optimum level of keyword density have resulted in the overlooking of keywords' density by SEO practitioners. From the researcher's point of view, the use of keyword decision tools

enable SEO practitioners to refrain from concentrating on keyword stuffing as they perceive themselves to be on the safe side of penalisation.

At one point, 80% of the interviewees indicated that they have waited for up to a period of 90 days to have a website indexed; various reasons, which were beyond their control, were stated. The main reason was based on crawler visitation that is not known by anyone and the failure to locate the site being ranked lowly for some specific competitive words. However, site search assists in eliminating this problem. Since competitive keywords pose a high probability of low ranking, the researcher found that all the interviewees support the use of a key phrase rather than one keyword. Singular keywords are prone to a higher bounce rate due to occasional wrong interpretations by search engines. For example, when optimising the keyword “spam”, it is difficult to be sure whether the search results displayed would be correct in terms of relevance, as the word “spam” can either refer to a type of canned luncheon meat, e-mail or Web spam (Zuze and Weideman 2010:58). Several studies were done by scholars such as Zing, Kritzinger, Weideman and others to determine the users’ search behaviour and how the length of a search string affects user results. The interviewees have confirmed that they are fully utilising this SEO technique.

The research further found that 60% of the interviewees are aware of the existence of keyword stuffing and that they were informed by different sources. However, there was no evidence that the interviewees got to the bottom of the tactic to see if it really exists and how it affects the website. The responses gathered by the researcher showed that information shared was based on scholars’ understanding rather than the interviewees’ knowledge on the subject. This was further depicted by differences in definitions and similar examples presented in order to show what keyword stuffing involves, for example, the penalisation of the Germany BMW website. In this regard only 20% of the interviewees partake in regular experiments whenever they hear about algorithm changes; 80%, on the other hand, are business centred, which is why they devote their time to business processes.

The 80% showed that they mostly base decisions on blogs and discussions, as well as analysis of the changing trends of a site, to determine the effects of an algorithm change. Nevertheless, the researcher noted that this may be a common business practice although it has some drawbacks in terms of factual decision making processes.

The research established that a keyword density of 5% was supported by 60% of the interviewees as the maximum keyword density level that is acceptable by both humans and crawlers. Nonetheless, 20% of the interviewees agreed that 3% is the best keyword density and above this mark they considered the level to be unacceptable by the end-user. These two figures became interviewees' measurement of a keyword rich website whilst a website with a keyword density below 3% was considered poorly optimised. A website with a keyword density above 5% was considered to be in the bracket of keyword stuffing. However, 20% of them acknowledged 12% as the crossover point to spamdexing, whilst 80% of the interviewees clearly desist from figuratively responding to what they considered to be the crossover point to keyword stuffing (spamdexing).

The position of Bing and Yahoo!, with respect to keyword stuffing, was not well articulated by all interviewees as concentration was centred more on Google. The research, however, concluded that the obsessive emphasis on Google was as a result of Google controlling a large search market share (Sterling 2010). All the interviewees were able to share their understanding of spamdexing on Google's perspective, which showed that Google's guidelines were often visited and the interviewees were all alert on following them.

The fact that the SEO industry is so diverse and many secrets are being kept has contributed greatly to the industry's growth. A division of expertise has subsequently been brought forward. In this respect, all the interviewees admitted that the industry is still growing and that many areas of specialisation continue to emerge.

4.4 INTERVIEWEES' WEBSITES INDEXING ANALYSIS

After the interview was concluded, the interviewees were furnished with copies of websites to enable them to determine whether or not the webpages of the website will be indexed. See Figure 4.5 for a summary of their responses. The Figure indicates that all the interviewees expected the Getlaptops1, Getlaptops2 and Getlaptops3 sites to be indexed. Among the interviewees, 40% were unsure if Getlaptops4 would be indexed whilst 60% agreed that this website would be indexed. None of the interviewees differed about the state of Getlaptops5 in terms of indexing; they all agreed that the site would not be indexed.

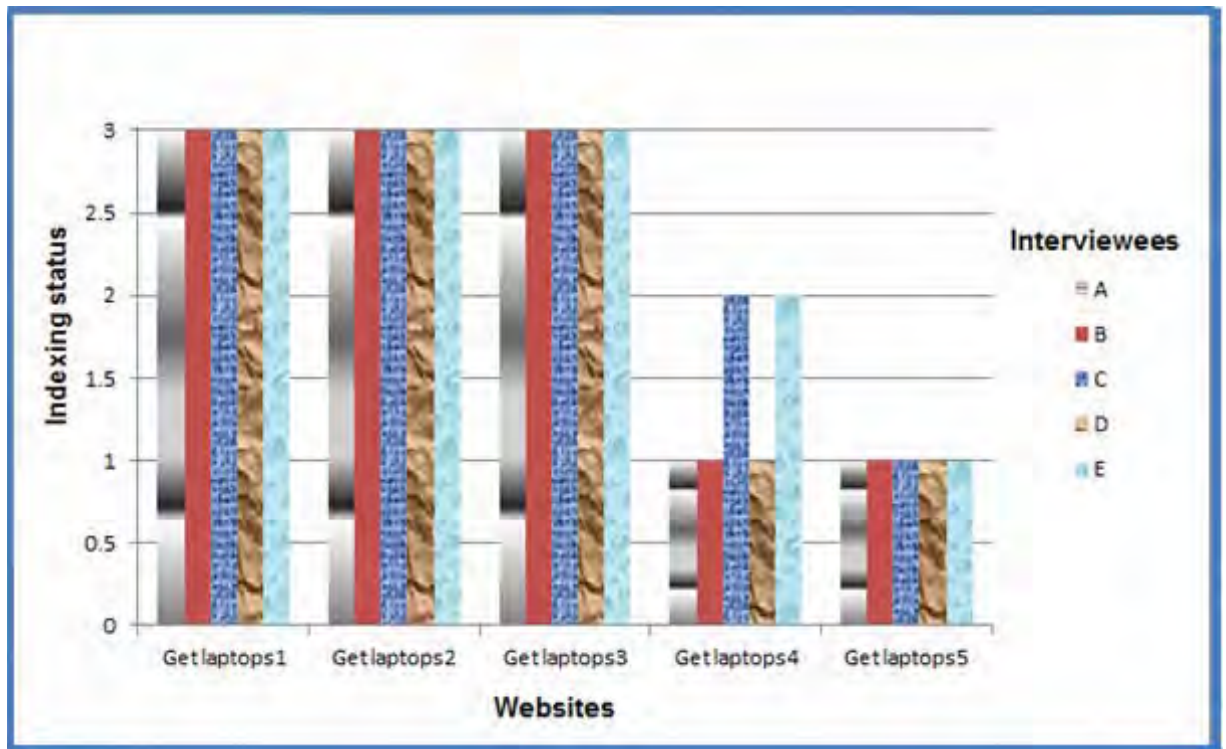


Figure 4.5: Webpages indexing predictions by interviewees.

Key

- 1 = Will not be indexed
- 2 = Not Sure
- 3 = Will be indexed

Although there is no official rule that stipulates the time taken for a website to be indexed, Apexacific (2010) and Zaslavsky (2010) pointed out that major search engines take in excess of three weeks and even up to six weeks to index a webpage. However, Pay-Per-Inclusion engines usually index within two to seven days, depending on the payment plan.

4.5 RESULTS OF E-MAILS SENT TO GOOGLE, YAHOO! AND BING PROFESSIONAL EMPLOYEES

In a bid to obtain inside information about the search engines' interpretation of keyword stuffing, e-mails were sent to three professionals working with Google, Yahoo! and Bing. The e-mails were sent by the supervisor, on behalf of the researcher, in order to increase the authenticity of the e-mail as well as an attempt to prove to the addressee that the content requested was solely for academic purposes. Appendix C shows the initial e-mail and the follow-up e-mail. The addressees did not

respond on the first e-mail, where after the researcher sent a second follow-up e-mail; however, unfortunately no response was forthcoming.

The researcher decided to stop pursuing the e-mail method due to the lack of response. The researcher considered the possibility that the addressees were unable to provide answers to the questions because:

- they were too busy with commercial oriented duties and could not find time to respond to the questions,
- the questions posed a conflict of interest and could result in confidential information being revealed,
- they might not have satisfying answers to the questions and/or
- this type of information is possibly only given to search engine partners.

Therefore, due to non-response the researcher was unable to continue with this part of data gathering, and concluded that the way forward was to utilise the information collected through the other methodologies.

4.6 RESULTS AND ANALYSIS OF SEARCH ENGINES' GUIDELINES

The researcher gathered guideline results from Google, Yahoo! and Bing. An analysis of the three search engines was conducted; although the information was limited to indexing and penalisation of websites found abusing any of the search engine guidelines.

4.6.1 Google

Google's Webmaster guidelines are used by Google to find, index, and rank a site. Failure to adhere to the guidelines might lead to a site being removed entirely from the Google index or alternately, being penalised. If a site has been penalised, it may no longer show up on the results on Google.com or on any of Google's partner sites (Google 2010d). These guidelines are divided into three categories, namely:

- design and content guidelines,
- technical guidelines and
- quality guidelines.

4.6.1.1 Design and content guidelines

Amongst other things, Google expect from webmasters under the design and content of websites:

- a created website should be beneficial, full of information, and the pages must clearly and accurately describe the content,
- users ought to be the main priority and the key concern should be about the word or words they would type to find the webpages, and the website should include the words within it and
- priority should be given to text rather than images to display important names, content, or links. Text should not be put into images because the Google crawler does not recognise it; therefore, in circumstances where images are utilised for textual content, the "ALT" attribute should be made use of and some descriptive words ought to be included in the text.

4.6.1.2 Technical guidelines

These guidelines are mainly concerned with how the crawlers crawl sites and the indexing of websites. It further looks into monitoring of websites to evaluate performance and optimise load times.

4.6.1.3 Quality guidelines

According to Google (2010d), the quality guidelines cover the most common forms of deceptive or manipulative behaviour. It further states that Google may respond negatively to other misleading practices not listed here (e.g. tricking users by registering misspellings of well known websites). Google offers a reporting site where anyone believing that another site is abusing Google's quality guidelines can report this fact at

<https://www.google.com/Webmasters/tools/spamreport>.

Examples of quality guidelines are as follows:

- stay away from using hidden text or hidden links,
- cloaking or sneaky redirects should not be used,

- directing automated queries to Google is prohibited,
- avoid loading pages with irrelevant keywords and
- avoid generating multiple pages, subdomains, or domains with substantially duplicate content.

Google (2010d) does not give a clear interpretation of what abusive keywords look like, but only provided the following information in reference to keyword stuffing:

“‘Keyword stuffing’ refers to the practice of loading a webpage with keywords in an attempt to manipulate a site's ranking in Google's search results”.

Overloading pages with keywords results in a negative user experience and can damage a site's ranking. Typically, keyword stuffing can be identified in the form of lists or a paragraph of keywords, often randomly repeated. It can often be in the form of hidden text, or hidden in title tags or alt attributes. However, emphasis should be placed on creating useful, information-rich content that uses keywords appropriately and in context.

4.6.2 Yahoo! search content quality guidelines

Yahoo! guidelines are used in order to provide the best search experience on the Web by guiding users to high-quality and relevant Web content in response to a search query (Yahoo! 2011).

4.6.2.1 Pages Yahoo! wants included in the index

Yahoo! does not provide absolute information regarding keyword stuffing and how it interprets it. However, according to Yahoo! (2011), some pages are created deliberately to trick the search engine into offering inappropriate, redundant or poor-quality search results. It states that Yahoo! does not want these pages in the index. Below are a few of the pages Yahoo! wants included in the index:

- original and unique content of genuine value,
- pages designed primarily for humans, with search engine considerations a secondary concern,
- hyperlinks intended to help people find interesting, related content, when applicable and

- metadata (including title and description) that accurately describes the contents of a webpage.

4.6.2.2 What Yahoo! considers unwanted

Yahoo! Search Content Quality Guidelines are designed to ensure that poor-quality pages do not degrade the user experience in any way. As with other Yahoo! guidelines, Yahoo! reserves the right, at its sole discretion, to take any and all action it deems appropriate to ensure the quality of its search index (Yahoo! 2011).

Some, but not all, examples of the types of content that Yahoo! does not want included:

- pages that harm the accuracy, diversity or relevance of search results,
- pages dedicated to redirecting the user to another page (doorway pages),
- multiple sites or pages offering substantially the same content,
- pages using methods to artificially inflate search engine ranking,
- the use of text or links that are hidden from the user,
- cloaking,
- pages built primarily for the search engines or pages with excessive or off-topic keywords and
- pages that seem deceptive, fraudulent, or provide a poor user experience.

4.6.2.3 Bing Webmaster center results

Like other search engine Webmaster centres, Bing Webmaster center is a hub of information, where resources regarding indexing and ranking on Bing are explored. The center is one of the places where SEO practitioners and website designers obtain what Bing expects from them and what subsequently results in penalisation.

4.6.3.1 Guidelines for successful indexing

The Bing Webmaster Center provides access to all the information webmasters need regarding using Bing, including the way in which MSNBot works, guidelines for getting a website indexed successfully by Bing, and usage information on Bing Webmaster Center tools.

4.6.3.2 Technical recommendations for a website

Bing provides techniques that can be used to ensure that websites are technically optimised for MSNBot and other Web crawlers. Below are some of the techniques provided by Bing (2011):

- only well-formed, HTML code should be used in webpages. All paired tags must be closed, and all links ought to lead to the correct webpage,
- broken links in websites should be fixed, as MSNBot may not be able to index the website effectively, thus preventing users from reaching all of the webpages,
- the design of the website should enable the MSNBot to crawl the site without encountering any difficulties. Furthermore, MSNBot must not be on the list of Web crawlers that are prohibited from indexing the website,
- URLs have to be kept simple and static, as it is difficult to index URLs that are complicated or that change frequently as link destinations and
- malicious software (malware) should be investigated closely. Links to webpages on a website that lead to malware on third-party websites or contain malicious content, such as a maliciously corrupted image or document file, or a harmful ActiveX control or JavaScript, will be disabled and highlighted as Malware in Bing results webpages (Bing 2011).

4.6.3.3 Content guidelines for a website

Having valuable content in which the target audience is interested is the best method to attract users to a website, and to keep them coming back. The following guidelines can assist in creating a more effective and popular webpage:

- always include words that users make use of when constructing search query terms to locate information on a website,
- all webpages should have a reasonable size which does not annoy users when navigating. An HTML webpage with no images must be under 150 Kilobytes (kB) and Bing advises covering one topic per webpage,
- text intended for indexing should not be placed within images,
- a Sitemap may be added, which assists MSNBot in locating all webpages. Crawlers have difficulties in accessing links that are embedded in menus, list boxes, and similar elements unless they appear in a sitemap and
- follow a fairly flat website hierarchy structure that is, each webpage should only be one to three clicks away from the default webpage (Bing 2011).

4.6.3.4 Techniques that may prevent a website from appearing in Bing results

Bing Webmaster Center listed some techniques that are not appropriate to use when attempting to gain higher ranking with the Bing index. They further noted that the use of the following techniques might adversely affect how a website is ranked within Bing, and could even result in a website being removed from the index:

- If an attempt is made to increase a webpage's keyword density by increasingly stacking a lot of irrelevant words. Also included in this technique is stuffing of ALT tags that users are unlikely to recognise,
- use of hidden text or links, nevertheless, text and links that are visible to users should be used and
- use of techniques, such as link farms, to artificially increase the number of links to a webpage.

4.6.4 Google, Yahoo! and Bing Webmaster centers

The researcher noted with great concern that the argument behind the penalisation of a site or removal from the index of the respective SE was shown by all the three search engines; unfortunately, the extent to which the penalty is implemented is not justified. Each search engine has its own guidelines but their results or their main focus converge to not tricking the SE or the user. Harsh rules upon failure to adhere to the guidelines are depicted by all search engines, with the worst penalty being removal of a website from the index. All three search engines brushed through the penalisation issue and addressed it in an umbrella scenario without providing specific examples. Google went on to describe keyword stuffing under its quality guidelines, however, nothing was mentioned about the point at which a site is blacklisted or removed from its index after repeating keywords. Yahoo! and Bing did not go into detail about keyword stuffing even though they stated some points in passing as part of the techniques that must be prevented. The failure by all three search engines to address this issue might be a reflection of the possibility that their algorithms could fail to counteract the keyword stuffing problem.

4.6.5 Websites' experimental results

After the submission of five websites to Google, Yahoo! and Bing, the researcher recorded each day's indexing results. They were checked using the following methods:

- a string search,



Figure 4.6: Getlaptops1 results using a search string on Google (Source: Google 2010a).



Figure 4.7: Getlaptops2 results using a search string on Yahoo! (Source: Yahoo! 2010a).

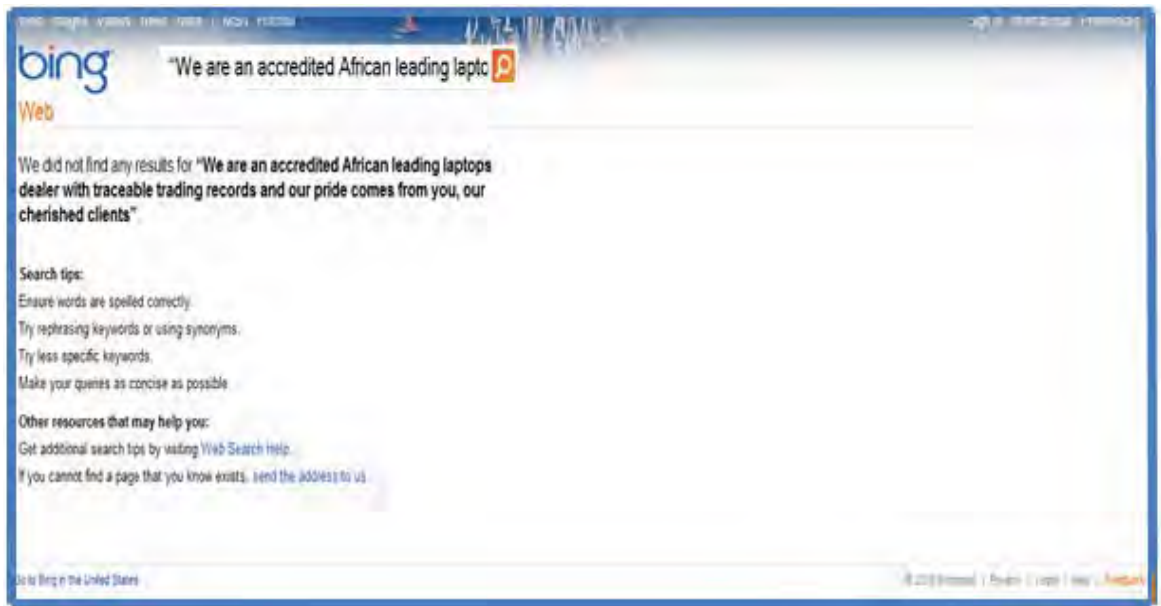


Figure 4.8: Getlaptops3 results using a search string on Bing (Source: Bing 2010a).

4.6.5.2 Site search

The researcher used the site search method to check the website indexing status; for example, by typing **site:url** on the search page for each SE.

4.6.5.3 Webmaster tools

The researcher registered all five websites with Google, Yahoo! and Bing Webmaster accounts and the result is shown in Appendix K (Webmaster Accounts).

4.7 PHASE 1 RESULTS AND ANALYSIS

Initially, the researcher had not planned to do two phases of the experiment but due to the variance in the initial results it was decided to record the first results as Phase 1. The results depicted in Appendix I could not prove that a keyword density of even as high as 27.30% would result in either the website being penalised, blacklisted or banned. After recording the indexing results for 67 days, all the website pages were successfully indexed with the exception of Getlaptops1, which was not indexed by Google. The research showed the fifth website as being the most favoured one and was indexed first by Yahoo! and Bing four days after submission. After five more days Yahoo! and Bing registered the remainder of the websites with Getlaptops5, having the highest keyword density of 27.30% (see Table 3.2 for a summary of keyword density).

Table 4.2 shows the homepage indexing time in days, recorded over a period of 67 days.

	GLPS1	GLPS 2	GLPS 3	GLPS 4	GLPS 5
GOOGLE	NI	28	29	33	33
YAHOO!	10	9	10	10	5
BING	10	9	10	10	5

Table 4.2: Phase 1 website homepage indexing time.

Key

GLPS - Getlaptops

NI - Not indexed

Google took longer to index the first website; the researcher assumed that the main reason for this slow pace of indexing may be due to the Google “Sandbox Effect”. The “sandbox” denotes the fact that Google applies an aging filter to its index; simply put, it prefers older sites to newer sites. However, 16 days after submission of the website Getlatops1, Getlatops2 and Getlatops3 were scraped by the “Magical Iranian” website (see Figure 4.9).

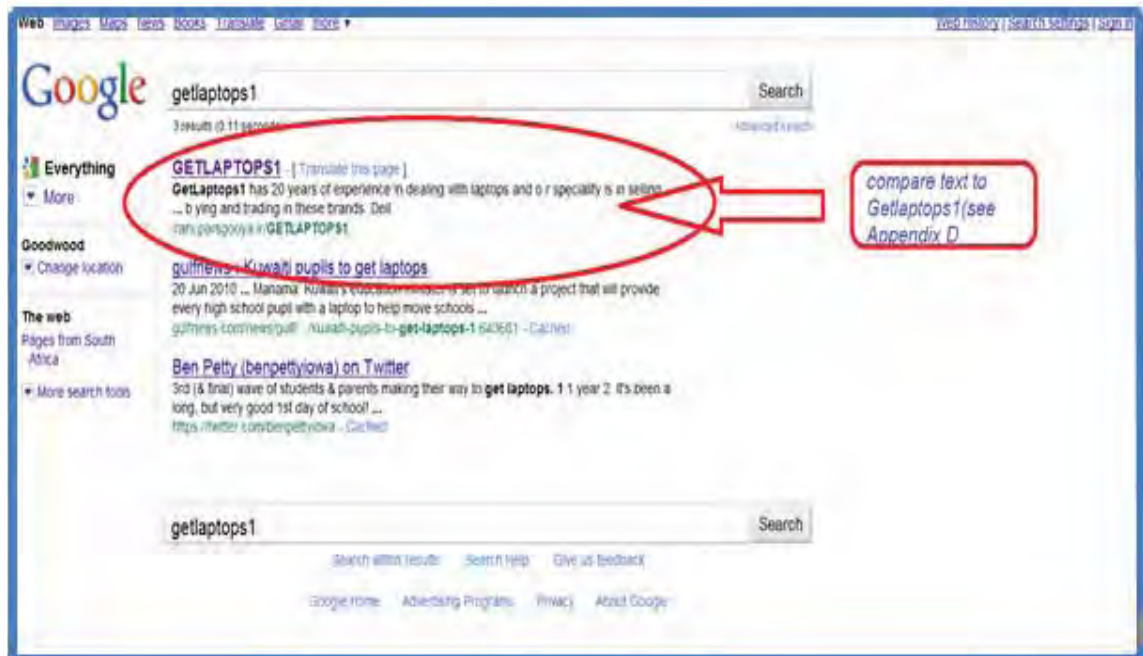


Figure 4.9: Screenshot of Getlaptops1 scraped by an Iranian website (Source: Google 2010a).

The text displayed by Google is exactly the same as the one for Getlaptops1; however, if the user visits the site it presented completely different information from

computer technology. In relation to the definition of cloaking in section 2.4.4.3, the research identified the Iranian website tactic as cloaking since content presented to the crawler is different from the content presented to the browser of the human visitor (see Appendix F). In this instance the researcher found that two unethical tactics were used by the Iranian site – content scraping and cloaking.

4.7.1 Analysis of scraped websites

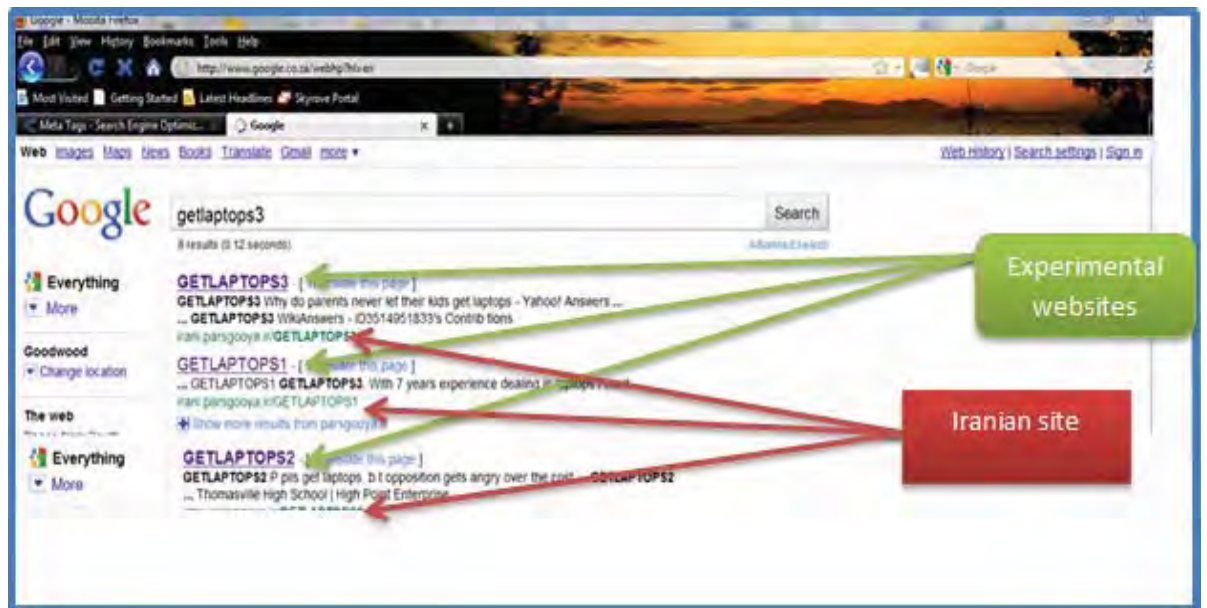


Figure 4.10: Screenshot of Getlaptops1, Getlaptops2 and Getlaptops3 scraped by an Iranian website (Source: Google 2010a).

Based on the identical text displayed on Getlaptops1, having all 3 domains' information displayed by one domain (irani.parsgooya.ir), and the researcher deduced that the Iranian website had nothing to do with technology, it was conclusively noted that these 3 websites were scraped. The relevancy of the content displayed by the Iranian website was accessed through the use of the translation tool offered by Google on its SERP. See Figure 4.11: Translating a webpage to another language).

The website was translated to the English language as follows:

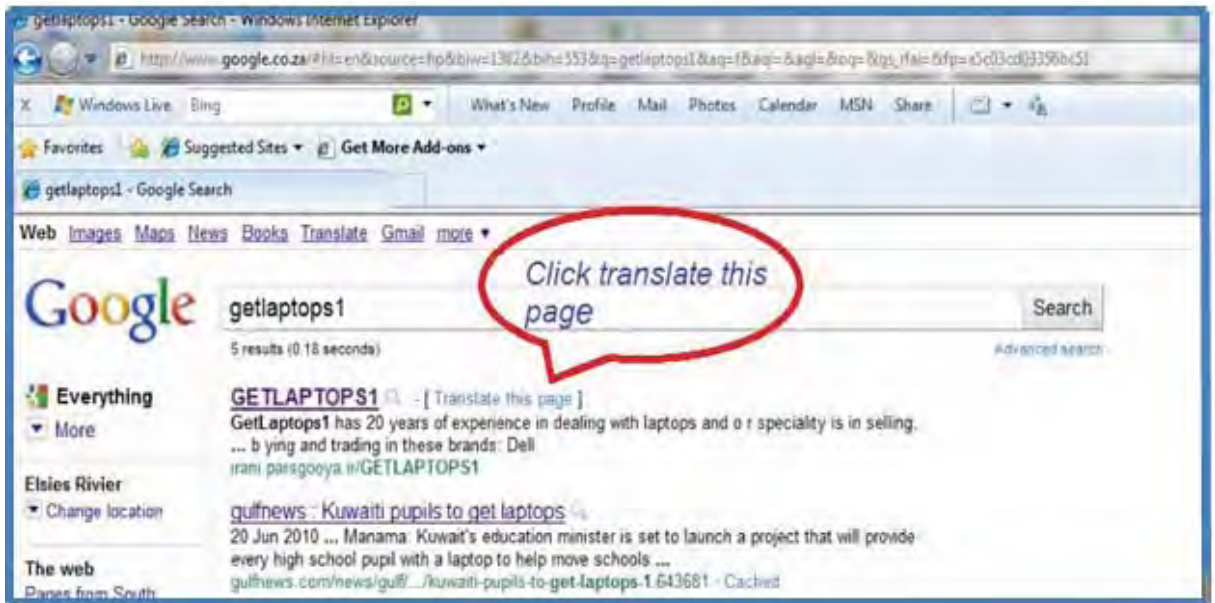


Figure 4.11: Translating a webpage to another language (Source: Google 2010a).

The results of the webpage translation are shown in Appendix L.

Getlaptops1 was the most affected as its text was taken as it was, whilst in the other two websites no text was directed to the site, the domain names only. The scraping lasted for 39 days (See Figure 4.12) shows the duration of the scraping per website.

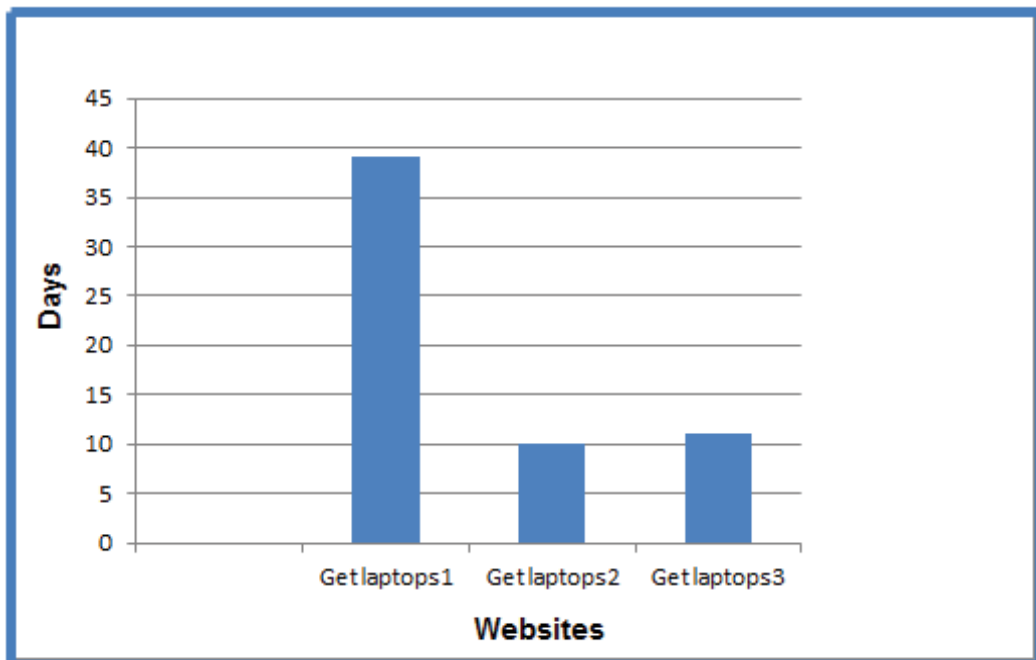


Figure 4.12: Duration of scraping per website.

Google indexed Getlaptops2 first and a day after it indexed Getlaptops3; however, Getlaptops1 could not be indexed for unknown reasons. After Getlaptops 2 and

Getlaptops 3 were indexed, both sites' content were removed from the Iranian website. The three websites could not be removed from the Iranian site. After 39 days of Getlaptops1 being scraped it was also dropped by the Iranian website. The researcher decided to proceed with the experiment for 12 more days to determine if Google was going to index the website. Unfortunately, the result stayed the same and the researcher terminated the experiment to pave way for the second phase. It was further concluded that Getlaptops1 may not have been indexed due to the damage it acquired through the scraping. The indexing result was contrary to what was predicted by the interviewees, as well as what literature states in terms of favourable keyword density.

4.8 PHASE 2 RESULTS AND ANALYSIS

Phase 2 of the experiment was carried out under the same conditions as the Phase 1 experiment; nevertheless, the notable difference was on an extra page that was included to attract the crawler visitation. The main difference was the keyword density (see Table 3.2 and Table 3.3 for the keyword density values). After 67 days, the experiment was officially closed and results were recorded as shown by Appendix J.

Table 4.3 shows the homepage indexing time in days recorded over a period of 67 days.

	GLPS1	GLPS 2	GLPS 3	GLPS 4	GLPS 5
GOOGLE	NI	11	NI	NI	NI
YAHOO!	24	23	23	20	19
BING	24	23	23	29	19

Table 4.3: Phase 2 website homepage indexing time

Key

GLPS - Getlaptops

NI - Not indexed

4.9 INDEXING STATISTICAL ANALYSIS

After the collection of the data from Phase 1 and Phase 2 the researcher found that the best way of statistically analysing the data was by using survival analysis. Survival analysis is based on the time an event takes to occur. There are occasionally instances when the event does not take place at all for the duration of the study and these cases are labelled "Censored" cases. Applying the concept to this study, the

researcher took a case where a webpage did not get indexed during the period of the study, disregarding the event that it might be indexed after the study. The researcher implemented the Kaplan-Meier procedure, which is a method of estimating time-to-time-event in the presence of censored cases.

The SPSS Manual (2007) describes the Kaplan-Meier model as being founded on estimating conditional probabilities at each time point when an event occurs and using the product limit of those probabilities to estimate the survival rate at each point in time. The Kaplan-Meier Survival Analysis assumes that the probabilities for the event depend only on time after the initial event. The researcher used this model to determine if the time for a webpage to be indexed (e.g. Time to event) was significantly different between the three search engines.

The data had to be transformed into survival format data so that for each situation the number of days it took for the event to happen (SE = Google, Keyword Count = 13, Phase 1) could be calculated. However, 30 records of data from Phase 1 and Phase 2 were produced (see Appendix M). The survival analysis was done on three different situations, namely:

- comparing indexing time between the three search engines,
- comparing the indexing time between the two groups of keyword situations and
- a Cox regression, which is a survival analysis, where one can include another co-variant (literally another independent variable that may have an effect on the outcome).

4.9.1 Analysis 1: comparing indexing time between Google, Yahoo! and Bing

Case Processing Summary

SE	Total N	N of Events (Indexed)	Censored	
			N	%
Google	10	5	5	50.0%
Bing	10	10	0	.0%
Yahoo!	10	10	0	.0%
Overall	30	25	5	16.7%

Table 4.4: Case processing summary.

Appendix N shows the survival table with the cumulative proportional survival per time. The data for each of the search engines is ordered by the number of days a webpage took to be indexed (time-to-event or survival time). For the search engine Google, there were five records that show censored values (e.g. webpages were not indexed for the duration of the study). This did not happen for the other two search engines. The fifth column (“Cumulative Proportion Surviving at the Time: Estimate”) shows that after 10 days the cumulative survival value is 0.9. Thus, the estimated probability of not being indexed beyond 10 days is 90.0%. The estimated probability of not being indexed beyond 32 days is 50%.

4.9.1.1 Analysis of Google indexing time

The mean values in Table 4.5 are not the arithmetic average, but an estimated value from the survival curve. The results showed webpages taking longer to be indexed with Google than with Yahoo! and Bing.

Means and Medians for Survival Time

SE	Mean ^a				Median			
	Mean	Std. Error	95% Confidence Interval		Median	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Google	46.400	6.765	33.141	59.659	32.000	.	.	.
Bing	14.000	2.226	9.637	18.363	9.000	2.767	3.577	14.423
Yahoo!	15.100	2.718	9.773	20.427	9.000	2.767	3.577	14.423
Overall	25.167	3.720	17.875	32.458	22.000	3.757	14.637	29.363

a. Estimation is limited to the largest survival time if it is censored.

Table 4.5: Means and medians for survival time on webpage indexing

Overall Comparisons

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	19.072	2	.000
Breslow (Generalised Wilcoxon)	14.735	2	.001
Tarone-Ware	16.911	2	.000

Test of equality of survival distributions for the different levels of SE.

Table 4.6: Overall comparison using Chi-square.

The distribution of indexing time is significantly different for the three SE populations.

4.9.1.2 Survival functions for Google, Yahoo! and Bing

The plot below shows the cumulative survival function over time. There is a more rapid drop-off in the cumulative survival function for Bing and Yahoo! than for Google; there are no censored values for either Bing or Yahoo!. The cumulative hazard plot reflects the same as the survival plot. It indicates that the “risk” of being indexed increases more rapidly over time for Bing and Yahoo! than for Google.

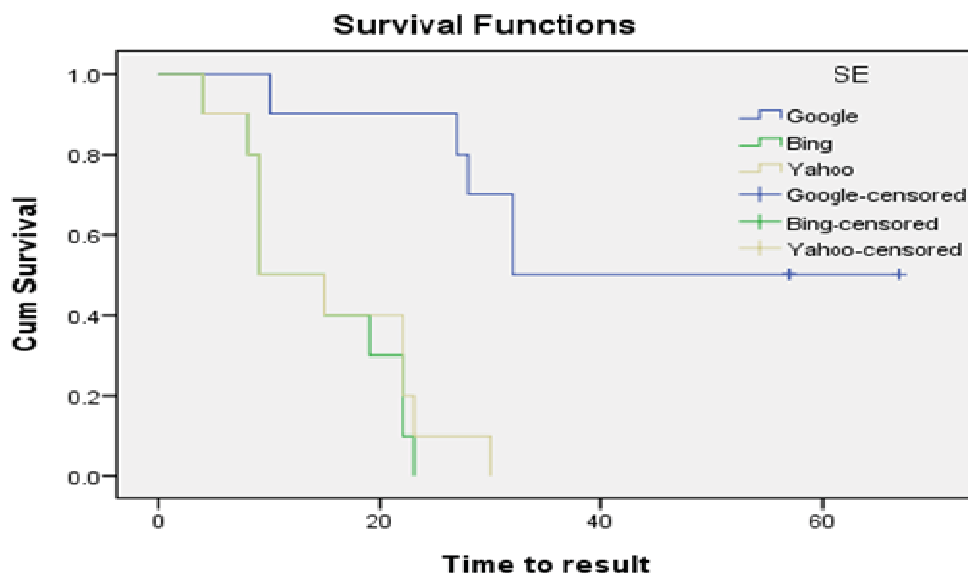


Figure 4.13: Webpage indexing survival function.

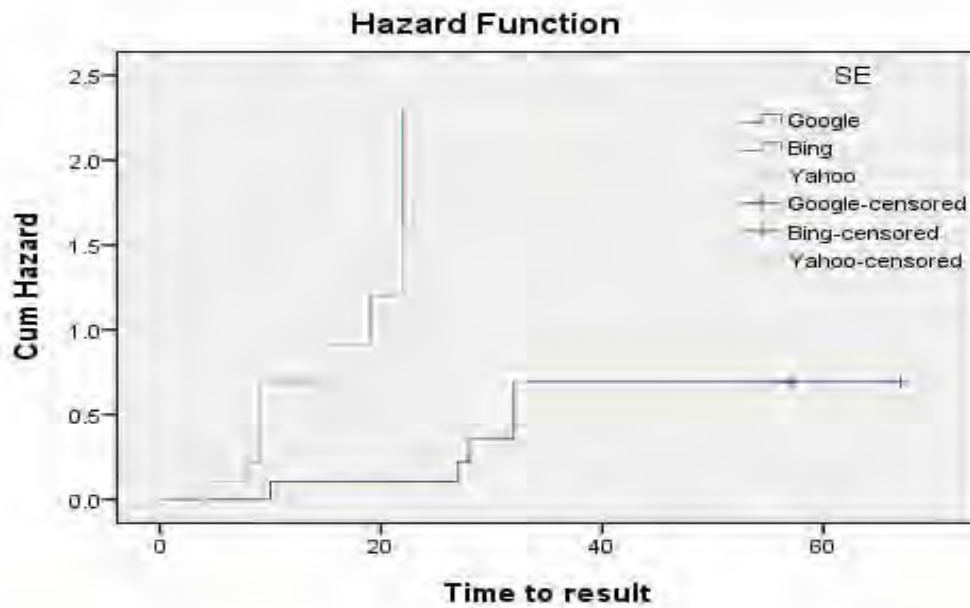


Figure 4.14: Webpage indexing hazard function.

4.9.2 Analysis 2: comparing indexing time between two keyword groups

The researcher created a new variable, with two new keyword groups that would assist in explaining keyword count indexing period. This variable splits the data into the websites where 40 or fewer keywords were used versus the situation where more than 40 keywords were used. The variable was named "KeywordGroup1". Table 4.7 shows the keyword case processing summary as discussed above.

Keyword Case Processing Summary

KeywordGroup1	Total N	N of Events (Indexed)	Censored	
			N	%
40 or fewer keywords	12	11	1	8.3%
More than 40 keywords	18	14	4	22.2%
Overall	30	25	5	16.7%

Table 4.7: Keyword case processing summary.

The data (See Appendix O) for each of the two groups is ordered by indexing time (time-to-event, or survival time). For the first group, there is only one censored record, and for the second group (having more than 40 keywords) there were four censored values (e.g. webpages were not indexed at all for the duration of the study).

The fifth column of Appendix O (“Cumulative Proportion Surviving at the Time: Estimate”) shows that after 8 days the cumulative survival value is 0.833. If the website had 40 or fewer keywords, the estimated probability of not being indexed beyond 8 days is 83.3%. For this situation the estimated probability of not being indexed beyond 32 days is 8.33%. However, if the website had more than 40 keywords then the estimated probability of not being indexed beyond 10 days is 83.3% and the estimated probability of not being indexed beyond 32 days is 22.2%.

4.9.2.1 Mean and medians for survival time

KeywordGroup1	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
40 or fewer keywords	18.667	4.901	9.061	28.272	9.000	.272	8.467	9.533
More than 40 keywords	27.278	4.106	19.229	35.326	22.000	1.581	18.901	25.099
Overall	25.167	3.720	17.875	32.458	22.000	3.757	14.637	29.363

a. Estimation is limited to the largest survival time if it is censored.

Table 4.8: Means and medians for survival time on KeywordGroup1.

The mean above is not the arithmetic average, but an estimated value from the survival curve. The results indicate that a webpage with at least 40 keyword count takes longer to index than one with less than 40. It is not clear whether or not this difference between the means is significant.

4.9.2.2 Overall comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	1.883	1	.173
Breslow (Generalised Wilcoxon)	2.869	1	.090
Tarone-Ware	2.394	1	.122

Table 4.9: Overall comparison of survival distributions for the different levels of KeywordGroup1.

Test of equality of survival distributions for the different levels of KeywordGroup1.

The distribution of indexing time is not significantly different for the two situations. For example, there is no significant difference in the indexing time between the websites that had 40 or fewer keywords, or the websites that had more than 40 keywords.

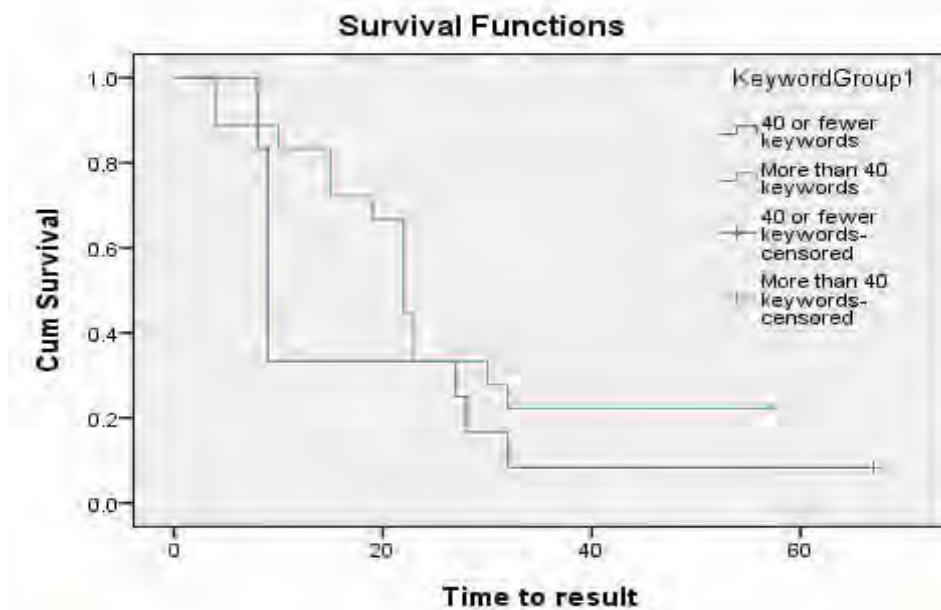


Figure 4.15: KeywordGroup1 survival function.

The plot above shows the cumulative survival function over time. There is a more rapid drop-off in the cumulative survival function for 40 or fewer keywords, than for the group of websites having more than 40 keywords. Note that both situations have censored values towards the end. The cumulative hazard plot shows the same as the survival plot. It indicates that the “risk” of not being indexed increases more rapidly over time for more than 40 keywords than for the websites having fewer than 40 keywords.

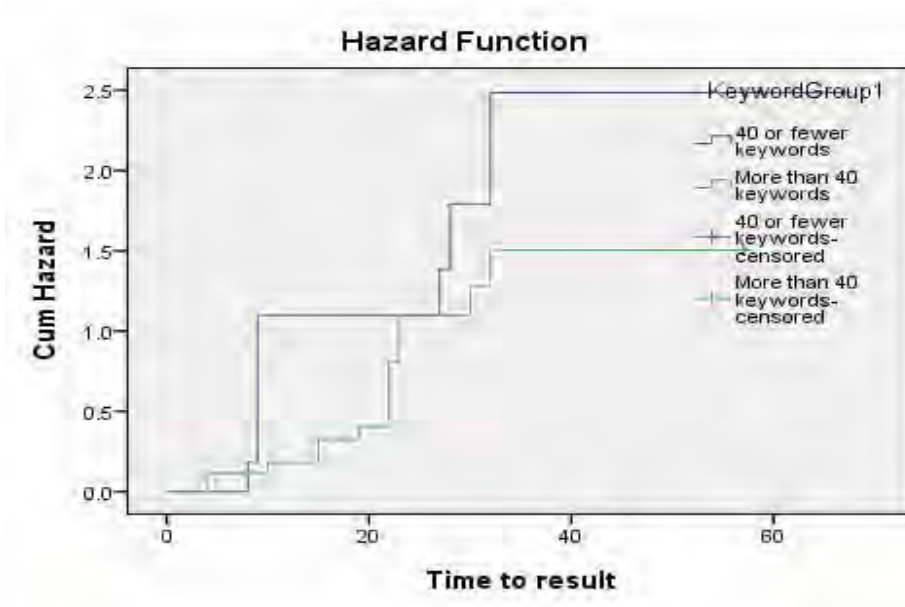


Figure 4.16: KeywordGroup1 hazard function.

4.9.2.3 Cox regression

In the following analysis the researcher did a regression using the indexed time as the dependent variable (also having censored values) and used the search engine and the variable, created for the previous analysis, as the independent variables.

According to the SPSS Manual (2007), “Cox Regression builds a predictive model for time-to-event data. The model produces a survival function that predicts the probability that the event of interest has occurred at a given time t for given values of the predictor variables. The shape of the survival function and the regression coefficients for the predictors are estimated from observed subjects; the model can then be applied to new cases that have measurements for the predictor variables. Note that information from censored subjects, that is, those that do not experience the event of interest during the time of observation, contributes usefully to the estimation of the model.”

Cox Regression is also a survival model that represents hazard (hazard = the probability that the event occurred at time t) as a function of time and predictor variables that can be continuous or categorical. Because it allows for multiple predictors, it is more general than the Kaplan-Meier method. It is considered a nonparametric, or perhaps more accurately, semi-parametric model, as it does not require a particular functional form to the hazard or survival curves. The model does assume that the ration of the hazard rate between two individuals or groups remains constant over time. If this assumption is not met, the Cox model has been extended

to incorporate time-varying predictors, which are interactions terms between the predictors and time.

The researcher decided to group Yahoo! and Bing under one category since other analyses indicated that they have a similar analytical behaviour. Therefore, the researcher had a new variable called “NewSE” having 0 = Not Google, and 1 = Google.

4.9.2.4 Case processing summary

		N	%
Cases available in Event^a		25	83.3%
analysis	Censored	5	16.7%
	Total	30	100.0%
Cases dropped	Cases with missing values	0	.0%
	Cases with negative time	0	.0%
	Censored cases before the earliest event in a stratum	0	.0%
	Total	0	.0%
Total		30	100.0%

a. Dependent Variable: Time to result

Table 4.10: The case processing summary for the indexing time.

Categorical Variable Codings^{c,d}

		Frequency	(1) ^b
NewSE^a	0=Not Google	20	1
	1=Google	10	0
KeywordGroup1^a	1=40 or fewer keywords	12	1
	2=More than 40 keywords	18	0

Table 4.11: The categorical variable coding.

a. Indicator Parameter Coding

b. The (0,1) variable has been recoded, so its coefficients will not be the same as for indicator (0,1) coding.

- c. Category variable: NewSE
- d. Category variable: KeywordGroup1

Note that the categories have been changed by SPSS: for KeywordGroup1, a “1” now indicates 40 or fewer keywords, and a “0” indicates more than 40 keywords and for NewSE a “1” indicates “Not Google” and a “0” indicates Google.

4.9.2.5 Omnibus tests of model coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-Square	df	Sig.	Chi-Square	df	Sig.	Chi-Square	df	Sig.
115.393	21.657	2	0.000	27.136	2	0.000	27.136	2	0.000

a. Beginning Block Number 1. Method = Enter

Table 4.12: Omnibus tests of model coefficients.

All the variables were entered at once (e.g. this is not a stepwise regression), so the values for “Change from previous step”, and “Change from previous block” are identical. The researcher’s objective was to test whether the effect of one or more of the predictor variables were considerably different from zero in the population. The results in the table above are analogous to the overall F-test used in regression analysis. The results indicate that at least one predictor is significantly related to the hazard.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
NewSE	3.165	.791	15.997	1	.000	23.681
Keyword Group1	1.573	.587	7.169	1	.007	4.819

Table 4.13: KeywordGroup1 equation variables.

In Table 4.14, the B-coefficient estimates relate the change in natural log of the hazard per unit change in the predictor. For this reason, the “Exp(B)” column is used when interpreting results. The significance of each predictor is tested using the *Wald*

statistic and the associated probability values that are recorded in the “Sig” column. In this case, both SE and KeywordGroup1 are significant (p-values < 0.05).

The “Exp(B)” column presents the estimated change in risk (hazard) associated with a one unit change in a predictor, controlling for the other predictors. In this case, where categorical predictors are used, Exp(B) represents change in hazard when changing from the reference category to another category. The Exp(B) for KeywordGroup1 is 4.819; this means that other things being equal, the hazard in KeywordGroup1 = 0 is 4.819 times greater than the hazard in KeywordGroup1 = 1. That indicates that the indexing time is shorter for KeywordGroup1 = 0 (more than 40 keywords).

The Exp(B) for SE is 23.681. This means that the hazard in SE = 0 is 23.681 times greater than the hazard for SE = 1 (“Google” to “Not Google”). This means that the indexing time for the “Not Google” sites was shorter.

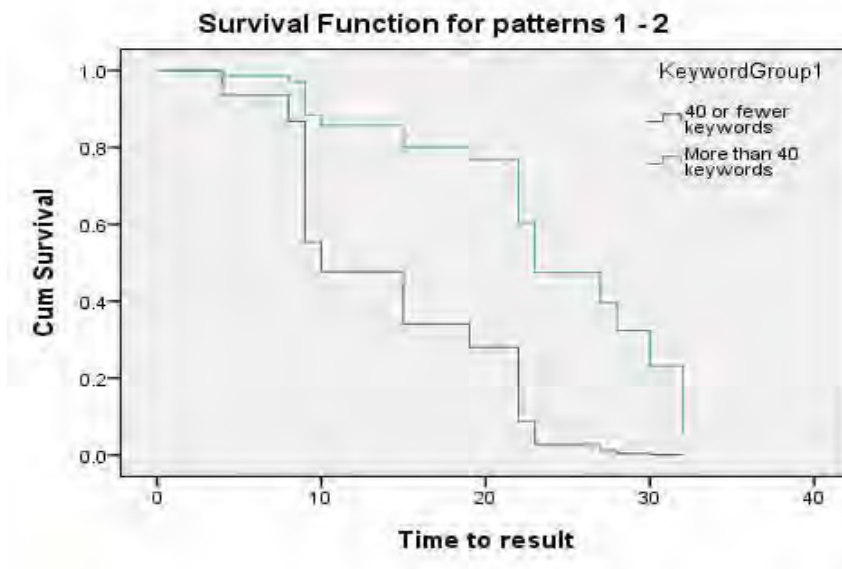


Figure 4.17: The survival function for patterns 1-2.

The plot above shows the cumulative survival function over time. There was a more rapid drop-off in the cumulative survival function for 40 or fewer keywords, than for the group of websites that had more than 40 keywords. Note that both situations have censored values toward the end. The cumulative hazard plot shows the same as the survival plot. It indicates that the “risk” of being indexed increases more rapidly over time for 40 or fewer keywords than for the websites that had more than 40 keywords.

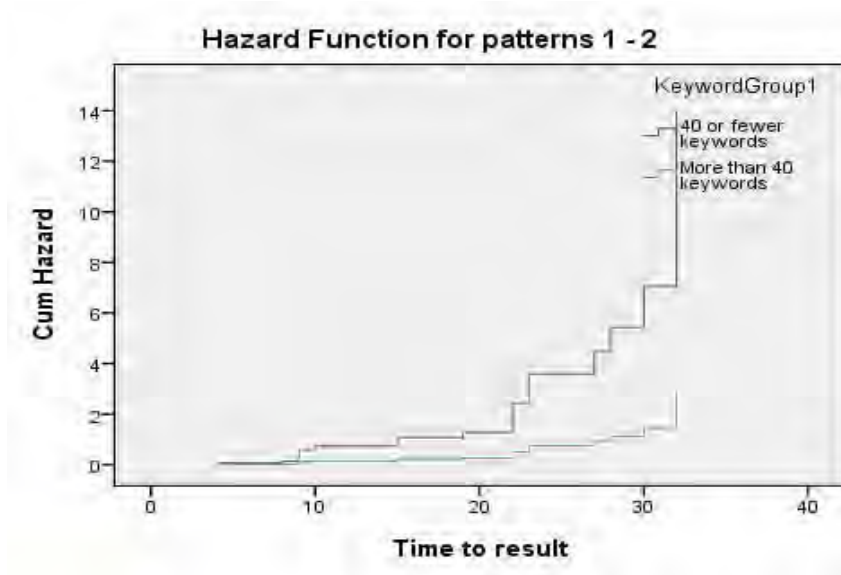


Figure 4.18: The hazard function for patterns 1-2.

4.10 SUMMARY

The results from academic literature demonstrated that the optimum keyword density is 12% and approximately 80% of the practitioners affirm a lower keyword density below 7%. However, this result differed from SEO practitioners whose acceptable keyword density was centred at 3% - 5%, with 20% marking 12% as the maximum keyword density acceptable by both the end-user and the search engine crawlers. Both the scholars and the SEO practitioners agreed that if a website keyword density exceeds the desired optimum keyword density acceptable by the search engine, the website risks penalisation (thus being removed from the search engine index). SEO practitioners and various scholars mentioned that if a website is blacklisted, notification is provided by the respective search engine reflecting the action taken and the reason therefore.

Efforts to acquire information from search engine experts working directly with the search engines were fruitless as none of the e-mails sent were responded to. However, this was not a major setback, as an experiment was conducted to test the validity of the facts and opinions gathered. Likewise, the information gathered from the search engine guidelines did not clearly outline the extent to which keyword density is regarded as spamdexing. A few sentences were noted just mentioning site penalisation in response to keyword stuffing. This was exactly the same result gathered from academic scholars and SEO practitioners regarding how secret the algorithm is.

The results gathered from academic literature, the interviews and the search engine guidelines were triangulated against results gathered from the experiment conducted. Both Phase 1 and Phase 2's experiments took 67 days. The Phase 1 experiment showed Bing and Yahoo! indexing all the five websites, whilst Google indexed four. Google did not register Getlaptops1, which was expected by all the interviewees to be indexed instead of Getlaptops4 and Getlaptops5. Getlaptops1 was scraped by an Iranian website from the 17th day of submission up to the 60th day of the experiment.

A Phase 2 experiment was conducted with the fifth website having a keyword density of more than 97% is an extreme excess. Likewise, Bing and Yahoo! indexed all five websites; however, Google exceptionally indexed Getlaptops2 leaving the other four websites. There were no notifications from search engines to inform the researcher about the indexing status of the four websites that were not indexed by Google. However, the researcher is of the opinion that Getlaptops1 was not indexed due to the damage caused by the scraping site, but no evidence was found to support this claim.

Apart from this, no empirical evidence based on this research was found in the form of notifications for blacklisted sites or banned sites as indicated by the interviewees and the literature. Getlaptops1 was not indexed by Google even though it had the lowest and most favourable keyword density of 3.94%, supported by the interviewees and the scholars. However, the Google algorithm was able to handle a webpage keyword density of 40% as shown by Table 3.4. Figure 4.19 and Figure 4.20 depicts the Phase 1 and Phase 2 indexing time recorded during the experiment.

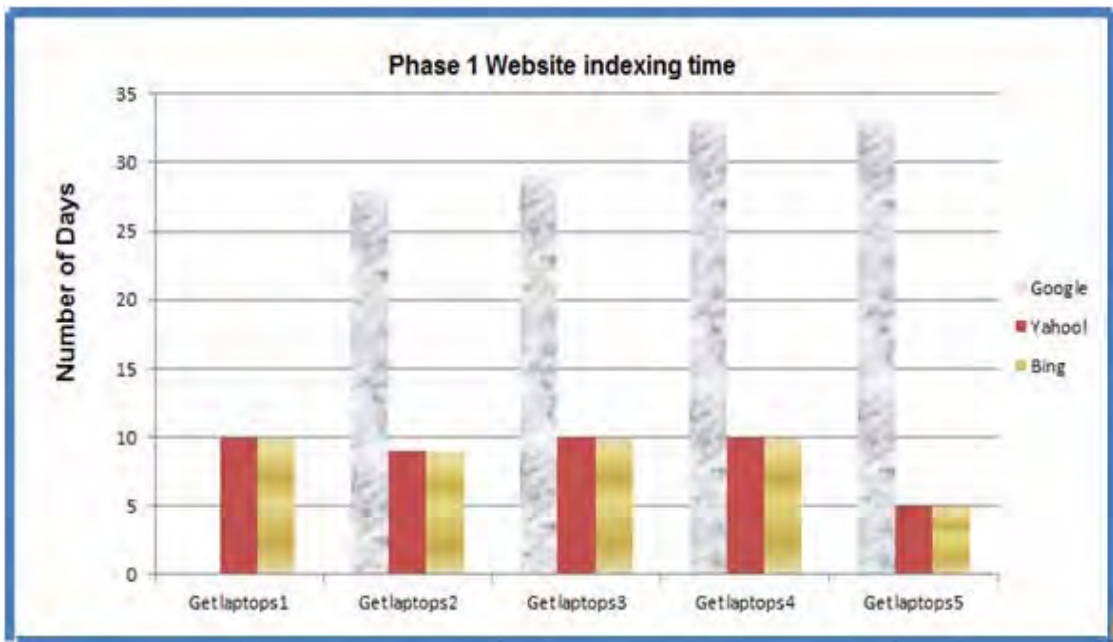


Figure 4.19: Phase 1 website indexing time recorded over a period of 35 days.

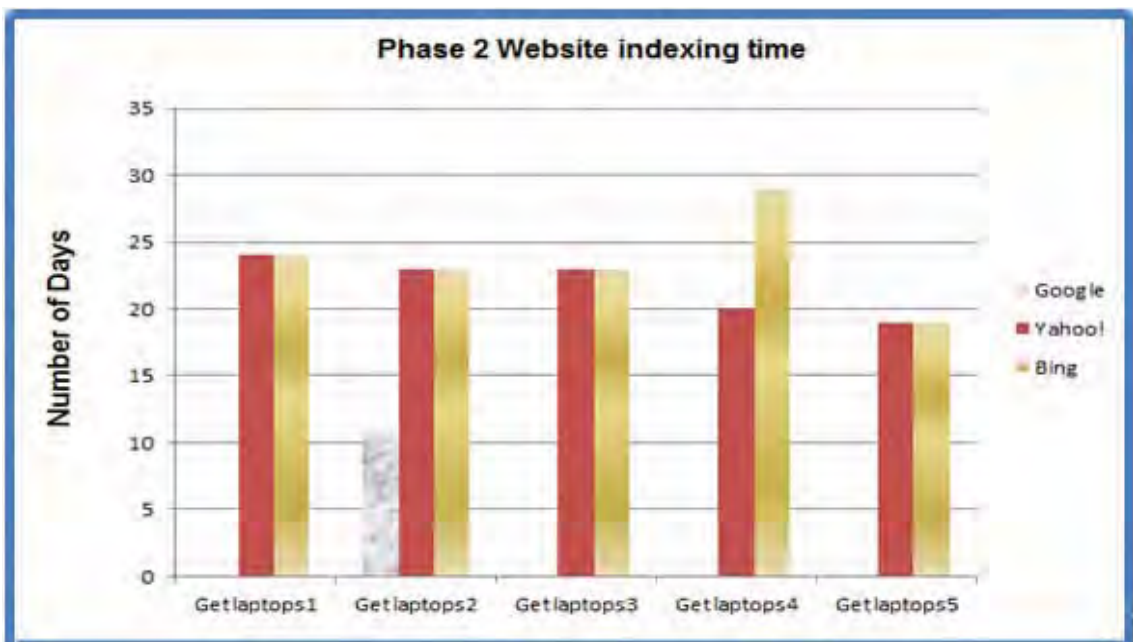


Figure 4.20: Phase 2 website indexing time recorded over a period of 35 days.

Both experimental phases recorded a maximum indexing time of 33 days for the 67 days of the experimental period; that is to say the maximum indexing time according to this study was 33 days and the minimum indexing period was five days. However, many scholars differ in terms of indexing time since the time is relative to crawlers' visitation. Thus a site can even be indexed in one day. It is ideal to use the long tail string search, Webmaster Tools and the site search methods in order to determine if the homepage or any of the subpages have been indexed. Using at least two of the search methods will provide a clear picture of the status of the website indexing, but

above all Webmaster Tools form the standard check as it provided analytical data regarding the indexing of the website pages.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 INTRODUCTION

This chapter is intended to provide a summary of the research and provide necessary recommendations as well as a conclusion. The conclusion is based on the facts gathered from the experiment, the literature reviewed and the researcher's own interpretation of various parts of the research exercises. The chapter further explores the understanding of relativity of keyword density in terms of penalisation of webpages, competitiveness and making informed decisions during the process of addressing the customers' particular needs.

5.2 WEBPAGE INDEXING

The researcher, as well as the interviewees and several other scholars, including Weideman and Zhang and Dimitroff, found empirical evidence indicating that the indexing period for webpages is not fixed but varies according to the crawlers' visitation. Borglum (2009:30) stated that after the submission of a website, one has to wait for a period of up to one month to evaluate indexing results. The author further cited that a search engine often only updates websites on a monthly basis. They all acknowledged a period of a day to three months if appropriate procedures are engaged during the design and submission of the webpages. However, according to this study a period of approximately 15 days is a reasonable average waiting time.

Phase 1's shortest indexing waiting time was five days and the longest was 33 days. Phase 2's shortest waiting time was 19 days and the longest waiting time was 29 days. During Phase 1, Bing and Yahoo! indexed all websites whilst Google indexed four sites, with the exception of Getlaptops1. Getlaptops1 had a low keyword density of 3.94%, compared to Getlaptops5 which had a keyword density of 27.3%. During the Phase 2 experiment, Bing and Yahoo! indexed all five sites, including the one with highest keyword density of 97.3%. Google only indexed one of the five sites, with a keyword density of 40%. In conclusion, the indexing percentage of the 15 homepages during Phase 1's experiment was 93%, whilst Phase 2 was 73%. Therefore, the average homepages' indexing for both phases is 83%.

5.2.1 Scraping

Even though scraping and the resultant duplication had been denounced by the search engines, following the evidence shown by this study this practice still exists and the search engines, such as Google, continue to display webpages of such nature on their results page. The researcher has evidence that search engine algorithms are still failing to fully address these practices and some developers are implementing them unnoticed. This research established that the waiting time for indexing can be prolonged by such practises; this may result in some websites not being indexed at all.

In this research, three websites were scraped on Google SERP, namely Getlaptops1, Getlaptops2 and Getlaptops3. Getlaptops1 was unable to get indexed by Google for both phases of the experiments. This is despite the fact that it had a good keyword density, as supported by scholars and the respective interviewees. The researcher concluded that the scraping that was executed by the Iranian site (see Appendix L) might have negatively affected the website on Google's index. Also, the Iranian webmasters might have opted to scrape the site, expecting a possibility of a higher number of visitors to the site looking for laptops rather than the information that was included on their site. The researcher also believes that Getlaptops1's homepage was the most preferred webpage for indexing, as compared to the other four webpages. This could be the reason why the site was scraped by the Iranian site.

The scraping result had a negative impact on the predicted results as the outcome deviated from the expectation. However, the researcher used the opportunity to address a second black hat technique that was also regarded illegal by search engines. The researcher further concluded that as a result of the differences in the way that the three websites were scraped, Getlaptop1 was not indexed whatsoever after the Iranian site discontinued the scraping of the sites.

Google did not provide any clear information or messages concerning the indexing of the websites, but generated an error message, as shown by Figure 5.1, if an attempt to check the cached result was made.

Note: The three scraped websites displayed the same results as the ones not indexed.



Figure 5.1: A snapshot of cached result error for Getlaptops1 (Source: Google 2010a).

However, the scraping effect would not have been noticed if the three indexing searching methods were not used.

5.2.2 Search engine algorithm

This research proved that the SE algorithms can never be predicted and change on a regular basis. This was verified through the recording of a very high webpage keyword density of 97%.

5.2.3 Keyword stuffing

Phase 2's websites contained extreme examples of keyword stuffing; however, in this research the webpages with the highest keyword densities, for both phases, were the first to be indexed by Yahoo! and Bing (see Figure 5.2).

Phase 1						
	Getlaptops1	Getlaptops2	Getlaptops3	Getlaptops4	Getlaptops5	
Google	0	28	29	33	33	
Yahoo!	10	9	10	10	5	Days
Bing	10	9	10	10	5	
Phase 2						
	Getlaptops1	Getlaptops2	Getlaptops3	Getlaptops4	Getlaptops5	
Google	0	11	0	0	0	
Yahoo!	24	23	23	20	19	Days
Bing	24	23	23	29	19	

Figure 5.2: Website indexing time.

Keyword stuffing should not be practised and a great deal of consideration should be given to the webpage content. A clear strategy in respect of the content and keyword density should be applied and must be in line with business requirements and policies. Instead of simply repeatedly placing keywords in the body text of the webpage, proper attention should be paid to the value of the content and if it makes sense to human readers, who are potential customers.

The research has proven that a webpage can have a high keyword density and be indexed; however, the adverse effects can be the increase of bounce rate due to lack of content relevance. This could further destroy the brand image, as well as provide a negative impact on the online marketing strategy.

Although crawlers might overlook stuffing, as proved by this study, stuffed content does not retain human readers. Time wasted on black hat techniques could instead be spent on creating decent content, moving away from vulnerabilities to the safe side, considered relevant by both the crawler and human readers. Furthermore, the time can be spent in designing websites that can be effective in adding importance to the user experience by its ability to present new, valuable, relevant, and comprehensive information.

The researcher noted that for all reported experimental results, website developers and content providers should fear “blacklisting” their own site to end-users rather than search engines blacklisting the site. The search engines in this study did not reject

indexing the site considered worst from a human point of view. Proper consideration should also be given to users not wanting to visit irrelevant webpages in order to locate the accurate information they are looking for.

5.2.4 Keyword density

The research has shown that keyword stuffing is not identified by Yahoo! and Bing. However, proper attention should be given to identifying the client's needs and a well informed decision must be taken in adherence of the objective. High keyword density in webpages lacks relativity and chases away potential customers as it lacks integrity and focus.

Each webpage ought to contain an acceptable keyword density in order to maintain the content quality required by users and furthermore regarded as helpful by search engines (Kassotis 2009). Based on the interviews and the experiments carried out in this study, the research recommends a keyword density of 3% to 6% to be acceptable by human readers.

5.3 SIGNIFICANCE OF THE STUDY

The data collected from the interviews, literature analysed as well as the experimental study on keyword density, have shown a significant difference between the perceptions of SEO practitioners' and the SE algorithm reaction reality. The triangulation method used enabled the researcher to clearly identify areas of divergences and uncertainties displayed by SEO practitioners and website designers in understanding keyword density, keyword stuffing and the way in which a penalty is implemented by search engines.

It is, however, understood that if the scholars, SEO practitioners, website designers and SE algorithm programmers incorporate this study the following may be expected:

- higher user satisfaction due to adherence to end user specifications,
- minimum bounce rate as webpages will be more content centred than search engine centred,
- scholars adopting the new keyword evolution displayed by search engines,
- increased conversion rate due to strategies and methodology changes and

- the ISP might have identified that the site was applying unethical keyword stuffing and therefore disabled it,
- the site owner could have taken it down or
- the owner might have deleted the files from the FTP server to update them and later upload new files.

Therefore, the researcher concluded that keyword stuffing is not a big concern that requires significant attention, which is why designers should not spend valuable time fine-tuning keyword densities.

However, this finding may lose relevance as search engine algorithms constantly change. Nevertheless, keywords are the foundation of SEO, as all strategies emanate from using the right keywords for deriving correct conversions. Keyword density should be evaluated as a business strategy or considered as a marketing perspective.

Penalisation may exist, but there is no crossover point of keyword rich website text to spamdexing that could result in penalisation of a website. This researcher, however, did not find any empirical results to confirm the banning of websites. The researcher is of the opinion that since search engines has a reporting facility on their Webmaster tool, appropriate action is taken if a site is reported for abusing or using black hat techniques (Yahoo! 2010c).

The researcher further concluded that SEO practitioners need to spend time doing experiments to check the validity of some of the information they gather from different sources. Bloggers may assist in alerting and providing valuable information but, on the other hand, the information should be verified.

The research confirms that there is a speedy evolution of technology and that SEO practitioners need to quickly adjust and understand that this significant evolution is not about indexing and ranking anymore. However, it is also about the volume of conversion being made, the degree at which customer's particular needs are being addressed and reducing bounce rate. The design and optimisation of each webpage should meet usability and search standards so that attention is not only given to SE operability and reaction toward the site.

This research has found empirical evidence that scraping is actively implemented and some website designers continue exploiting this technique. The researcher also

established that a scraped site can be negatively affected in terms of its indexing time frame and in some situations the website may completely fail to be indexed.

5.5 RECOMMENDATIONS

SEO practitioners should consider how to alter their perspective of keyword stuffing.

In an attempt to earn high rankings and indexing, SEO practitioners must avoid frustrating users visiting a site by including irrelevant information and not answering their question.

It would be more effective to spend resources on making webpage content interesting, relevant and engaging, rather than compromising webpage relevance by keyword stuffing. The cost of retaining a client is lower than winning a new one and clients' intent should be well understood.

Webmaster Tools are fundamental strategic tools that ought to be used for analytical judgement and powerful decision making, as well as trying to push for relevant traffic to a webpage. These tools assist one to make factual recommendations and decisions. Therefore, the SEO practitioners should utilise them and supplement with other open source and commercial software to verify their facts for strategic decisions.

The companies' methodology or strategy must be reviewed constantly in accordance with the algorithm changes, as each algorithm change could possibly have a counter positive or negative impact on the company's strategy being applied. A relevant example is the Google Panda "Farmer" algorithm update that was released at the time of writing.

In order to avoid penalisation or a webpage not being indexed, attention should be paid to consumer perspective rather than searching engine requirements.

SEO tools should be utilised regularly as they provide a basic structural architecture of how the website is built and how it can be accessed on the Web. They further offer a range of recommendations that assists in amendments and other strategic decisions that have to be made.

Keyword density not be used as a deceptive technique for website ranking; instead, it must be used legitimately in ways that increase a site's visibility, usability, conversion and accessibility.

As Murphy, Christian and Kielgast (2008:90-97) stated, more than 85% of all Internet purchases started with a search. Hence, enough descriptive website content should be identified and strategically positioned on a webpage in order to capture the interest of both the user and the search engine. By doing so, the website would be fully interpreted by search engines, as its content will have better ranking odds. The use of generic content in websites enhances the build-up of trust and credibility of websites by search engines in relation to the purpose of a particular website. The aforesaid would result in website objectives being achieved.

5.6 FUTURE RESEARCH AREAS

This study paved the way for future studies to be executed in areas that include scraping, as well as the relevancy of information displayed on the SERP. In addition, research needs to be done on other spamdexing techniques in order to test the validity of existing claims. The results may possibly enhance the quality of information a search query can retrieve from SE indices.

5.7 RESEARCH SUMMARY

Although it is important to adhere to SEO guidelines, there is a need for experiments to be carried out in order to determine the level at which search engines penalise websites. Whilst a high keyword density has a negative impact on a site with respect to usability and relevance, this author found no crossover point from keyword rich website text to spamdexing. The author has proved that search engines' interpretation of spamdexing on varying keyword densities is not based on a scale; a site can be indexed regardless of how low or high the keyword density is. This author has also proven that Yahoo! and Bing favour sites with high keyword density when indexing, compared to the ones with a low keyword density. The research has further proved that, other than keyword stuffing and cloaking, there are other unethical techniques that are still being used on the Internet such as content scraping.

REFERENCES

- Abernethy, J., Chapelle, O. & Castillo, C. 2009. WITCH: Web spam identification through content and hyperlinks. *In Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb '08)*, Beijing, China, 22 April 2008: 1-23.
- Abernethy, J., Chapelle, O. & Castillo, C. 2010. The Graph Regularization Methods for Web Spam Detection. *Journal Machine Learning – ML*, 81(2):207-225.
- Alimohammadi, D. 2003. Meta-tag: A means to control the process of Web indexing. *Online Information Review*, 27(4): 238-242.
- Angrosino, M. 2007. *Doing Ethnographic and Observational Research: The SAGE Qualitative Research Kit*. London: SAGE Publications Ltd.
- Apexpacific, 2010. *SEO Suite: How long does it take to have my website indexed?* <http://www.apexpacific.com/knowledgebase/submission/fagdetail.asp?id=8> [25 November 2010].
- Appleton, R. 2010. *Are Spam Tactics Stopping Your Site From Reaching Its Rankings Potential?* <http://www.searchmarketingstandard.com/spam-tactics-stopping-site-reaching-rankings-potential> [19 April 2010].
- Axandra. 2008. *Proven Methods for successful Search Engine Marketing (SEO)*. <http://www.scribd.com/doc/25413811/search-engine-mktg> [15 April 2010].
- Bartow, A. 2010. *The future of Internet and how to stop it*, by Jonathan Zittrain (Book Review). *Michigan Law Review*, 108(6):1079-1093.
- Beel, J. & Gipp, B. 2010a. Academic Search Engine Spam and Google Scholar's Resilience Against it. *The Journal of electronic publishing*, 13(3). <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0013.305> [21 January 2011].
- Beel, J. & Gipp, B. 2010b. On the Robustness of Google Scholar Against Spam. *In Proceedings of the 21st ACM Conference on Hypertext and Hypermedia (HT'10)*, Toronto, Canada, 13 - 16 June 2010:297-298.
- Benczur, A.A., Erdelyi, M., Masanes, J. & Siklosi, D. 2009. Web Spam Challenge Proposal for Filtering in Archives. *In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb2009)*, Madrid, Spain, 21 - 21 April 2009:61-62.
- Bing. 2010a. *Bing central home page*. <http://www.bing.com/> [22 October 2010].
- Bing. 2010b. *Bing Site Submission Console*. <http://www.bing.com/Webmaster/SubmitSitePage.aspx> [10 October 2010].
- Bing. 2011. *Guidelines for successful indexing*. http://help.live.com/help.aspx?mkt=en-us&project=wl_Webmasters [15 February 2011].
- Blanche, T.M., Durrheim, K. & Painter, D. 2006. *Research in Practice: Applied Methods for Social Sciences*. Cape Town: University of Cape Town Press.
- Bless, C., Smith, H.C. & Kagee, A. 2006. *Fundamentals of Social Research Methods: An African Perspective*. Cape Town: Juta & Co. Ltd.

- Borchardt, J. & Weideman, M. 2008. A theoretical discussion of the role the use of images and videos play in the achievement if top search ranking. *Working Paper*, Munich University of Applied Science, Munich.
http://dk.cput.ac.za/cgi/viewcontent.cgi?article=1052&context=inf_papers. [15 November 2010].
- Borglum, K. 2009. Getting your website to show up in search ranking. (Practice Management Q&A), *Medical Economics*, GALE, 86(14):30.
- Bowman, J. L. 2004. *Optimization with keywords*. <http://spiderfood.net/keywords.html> [09 June 2010].
- Bruemmer, P. 2001a. *SEO and the Web Design Process*.
<http://www.clickz.com/clickz/column/1706004/seo-web-site-design-process>. [05 August 2010].
- Bruemmer, P. 2001b. *Understanding the Power of Search engine marketing*.
<http://www.clickz.com/search/opt/article.php/909181> [05 August 2010].
- Canel, F. 2010. *Bingbot, the Sequel*. <http://www.webownertools.com/2010/09/29/bingbot-the-sequel/> [12 October 2010].
- Castillo, C., Chellapilla, K. & Davison B.D. 2008. Adversarial Information Retrieval on the Web (AIRWeb'07). *ACM SIGIR Forum*, 42(1):68-72.
- Castillo, C. Gyongyi, Z. Jatowt, A. & Tanaka, K. 2011. Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2011). *In Proceedings of the Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2011) In conjunction with 20th international World Wide Web Conference*, Hyderabad, India, 28 March - 1 April 2011:313-314.
- Charlesworth, A. 2009. *Internet Marketing: A practical approach*. Oxford: Butterworth-Heinemann Publishing.
- Chen, L. 2010. Using a two-stage technique to design a keyword suggestion system. *Information Research (IR)*, 15(1). <http://informationr.net/ir/15-1/paper425.html> [10 April 2011].
- Cho, A. 2008. What is Web 3.0?: The Next Generation Web: Search Context for Online Information. <http://www.suite101.com/content/what-is-web-30-a61407>. [21 October 2010].
- Clay, B. 2009. *How search engines define spam*.
<http://seotoolset.com/information/antispam.html> [15 April 2010].
- Clay, B. & Esparza, S. 2009. *Search Engine Optimization All-In-One for Dummies*. Hoboken, New Jersey: Wiley Publishing Inc.
- Corazzon, R. 2009. *Theory and history of ontology: A resource guide for philosophers*.
<http://www.formalontology.it/> [24 June 2010].
- Crowell, G. 2004. *Search Engine Watch: Retaining traffic after website redesign*.
<http://Searchenginewatch.com/article/2065265/Retaining-Traffic-after-a-Web-Site-Redesign>. [10 September 2010].
- Curran, K. 2004. Tips for achieving high positioning in the results pages of the major search engines. *Information Technology Journal*, 3(2):202-205.
- CyberWyre. 2006. *Highest paying AdSense keywords*. <http://www.cwire.org/2006/03/23/updatedhighest-paying-adsense-keywords> [21 September 2010].

- Dai, N., Davison, B.D. & Qi, X. 2009. Looking into the past to better classify web spam. *Proceedings of the 5th international workshop on Adversarial information retrieval on the web (AIRWeb 2009)*, Madrid, Spain, 21 April 2009:1-8. <http://airweb.cs.clehigh.edu/2009/papers/pl-dai.pdf> [20 April 2011].
- Knowzit, Y. 2009. *Ontology Vs Epistemology* . <http://yemovonknowzit.blogspot.com/> [22 August 2011].
- DeJarnette, R. 2009. *Search Engine Optimization for Bing*. <http://www.bing.com/community/blogs/Webmaster/archive/2009/09/03/search-engine-optimization-for-bing.aspx> [11 October 2010].
- Dejarnette, R. 2010. *Chasing the long tail with keyword research (SEM 101)*. <http://www.bing.com/toolbox/blogs/Webmaster/archive/2010/02/25/chasing-the-long-tail-with-keyword-research-sem-101.aspx> [11 October 2010].
- East, P. 2007. Search Engine Marketing: A 40-minute primer on using search engines to grow your business, *MedicBD Health*. <http://www.slideshare.net/hanapinmarketing/search-engine-marketing-a-40minute-primer-on-using-search-engines-to-grow-your-business-presentation> [15 January 2011].
- Egele, M., Kolbitsch, C. & Platzer, C. 2009. Removing web spam links from search engine results. *Journal in Computer Virology*, 7(1):51-62.
- Eisenberg, B., Quarto-vonTivadar, J., Davis, L.T. & Crosby, B. 2008. *Always be testing: The complete guide to Google website optimizer*. Sybex: Indianapolis.
- Erdelyi, E., Benczur, A.A. , Masanes, J. & Siklosi, D. 2009. Web spam filtering in internet archives. *Proceedings of the 5th international workshop on Adversarial information retrieval on the web (AIRWeb 2009)*, Madrid, Spain, 21 April 2009:17-20.
- Erdelyi, M., Garzo, A. & Benczur, A.A. 2011. Web spam classification: a few features worth more. *In Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2011) In conjunction with the 20th International World Wide Web Conference*, Hyderabad, India, 28 March 2011: 27-35.
- Featherstone, M., Adam, S. & Borstorff, P. 2009. A methodology for extracting quasi-random samples from World Wide Web domains. *Journal of International Business Research*, 8(1):63-67. <http://www.gale.cengage.com> [10 September 2010].
- Fiol-Roig, G., Miró-julià, M. & Herraiz, E. 2011. Data Mining Techniques for Web Page Classification. *Advanced in Intelligent and Soft Computing*, 89:61-69.
- Fox, W. & Bayat, S.M. 2007. *A guide to Managing Research*. Cape Town: Juta & Co Ltd.
- George, D. 2005. *The ABC of SEO: Search Engine Optimization Strategies*. Morrisville: Lulu Press.
- Goldsborough, R. 2005. Get your site noticed without spamming (Personal Computing). *Black Issues in Higher Education*, 21(26):40-42.
- Google. 2008a. *Technology overview*. <http://www.google.com/corporate/tech.html> [09 September 2010].

- Google. 2008b. *What's an SEO? Does Google recommend working with companies that offer to make my site Google-friendly?* <http://www.google.com/support/Webmasters/bin> [09 September 2010].
- Google. 2010a. *Google central home page.* <http://www.google.co.za/> [22 October 2010].
- Google. 2010c. *Google Site submission console.* <http://www.google.com/addurl/?continue=/addurl> [10 October 2010].
- Google. 2010d. *Webmaster Guidelines and Keyword stuffing* <http://www.google.com/support/Webmasters/bin/answer.py?answer=66358> [15 February 2011].
- Google. 2010e. *Google Webmaster Central.* <http://google.com/support/Webmasters/bin/answer.py?answer=66358> [15 April 2010].
- Gori, M. & Witten, I. 2005. The Bubble of the Web Visibility: Promoting visibility as seen through the unique lens of search engines, Technical Opinion. *Communication of the ACM*, 48(3):115-117.
- Jacoby, G. A. & Luqi. 2007. Intranet model and metrics. *Communications of the ACM*, 50(2):43-50. <http://cacm.acm.org/magazines/2007/2/5736-intranet-model-and-metrics/abstract> [16 August 2010].
- Guang-Gang Geng, Z., Jin, X. & Zhang, D. 2011. Evaluating web content quality via multi-scale features. In *Proceedings of the Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2011) In conjunction with 20th international World Wide Web Conference*, Hyderabad, India, 28 March 2011:18-24.
- Gyongyi, Z. & Garcia-Molina, H. 2005. Web spam taxonomy. In *Proceedings of the 1st International Workshop on Adversarial Information Retrieval on the Web (AIRWeb2005) Held in conjunction with the 14th International World Wide Web Conference*, Chiba, Japan, 10 May 2005:39-47. <http://ilpubs.stanford.edu:8090/771/2005-9.pdf>. [25 April 2011].
- Hayati, P. & Potdar, V. 2009. Evaluation of Spam Detection and Prevention Frameworks for Email and Image Spam - A State of Art. *Digital Ecosystems and Business Intelligence (DEBI) Institute Curtin University of Technology.* <http://debii.curtin.edu.au/~pedram/research/1-webspam/76-evaluation-of-spam-detection-and-prevention-frameworks-for-email-and-image-spam-a-state-of-art.html> [19 January 2011].
- Henzinger, M. R., Motwani, R. & Silverstein, C. 2002. Challenges in Web Search Engines. *ACM SIGIR Forum*, 36(2):1-12. <http://ijcai.org/Past%20Proceedings/IJCAI-2003/PDF/278.pdf> [18 March 2011].
- Hotho, A., Benz, D., Jaschke, R. & Krause, B. (editors). 2008. *Proceedings of the Workshop at 18th Europ. Conf. on Machine Learning ECML'08 / 11th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases PKDD'08, 2008*, Antwerp, Belgium, 15-19 September 2008.
- Iler, H. 2006. *Maximizing Visibility for Multilingual Websites*. Ontario: Wintranslation.com. http://www.sempo.org/resource/resmgr/Docs/wintranslation_for_SEMPO.pdf. [15 April 2010].
- Internet World Stats organization. 2010a. *Internet Usage and World Population Statistics.* <http://www.Internetworldstats.com/stats.htm> [07 September 2010].

- Jerkovic, I. J. 2009. *SEO Warrior*. Sebastopol, Canada: O`REILLY Media.Inc.
- Johnson, J. 2007. *GUI Bloopers 2.0: Common User Interface Design Dont's and Dos*. Morgan Kaufmann: San Francisco.
- Kassotis, C. 2009. Keyword Stuffing and Anchor Text - Mortal Enemies! *Article Submission Directory, Information Resource, RSS Syndication*.
<http://www.easyarticlesubmit.com/Article/Keyword-Stuffing-and-Anchor-Text--Mortal-Enemies/193162> [01 April 2011].
- Kennedy, D. 2009. Broadening search: new tools, approaches raise the question: is Google enough? *ABA Journal*, November 95(11):28. <http://www.gale.cengage.com> [10 September 2010].
- Kimura, M., Saito, K. & Sato, S. 2005. Detecting Search Engine Spam from a Trackback Network in Blogspace. *Proceedings of the 9th KES International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES2005)*, Australia, Melbourne, 14-16 September 2005:723-729.
- Krause, B., Schmitz, C., Hotho, A. & Stumme, G. 2008. The Anti-Social Tagger – Detecting Spam in Social. Bookmarking Systems. *Proceeding AIRWeb '08 Proceedings of the 4th international workshop on Adversarial information retrieval on the web (AIRWeb 2008)*, Beijing, China, 22 April 2008:61-68.
- Kritzinger, W.T. 2005. The Effect Webpage Body Keyword Location has on Ranking in Search Engine Results study: An Empirical. Unpublished Masters Dissertation, Cape Peninsula University of Technology, Cape Town.
- Kritzinger, W.T. & Weideman, M. 2007. Key word placing in Webpage body text to increase visibility to search engines. *South African Journal of Information Management*, 9(1).
<http://sajim.co.za/index.php/SAJIM/article/view/16> [16 August 2010].
- Kritzinger, W.T. & Weideman, M. 2008. Finding the synergy: search engine optimization vs. website usability. *Proceedings of the 10th Annual Conference on World Wide Web Applications*, Cape Town, 3-5 September 2008. <http://www.zaw3.co.za> [04 October 2010].
- Kumar, R. 2005. *Research Methodology-A Step-by-Step Guide for Beginners: (2nd.ed.)*. Singapore: Pearson Education.
- Langville, A.N. & Meyer, C. D. 2006. *Google's PageRank and Beyond – The science of Search Engine Ranking*. Princeton: Princeton University Press.
- Leedy, D.P., Ormrod, J. E. 2005. *Practical Research Planning and Design*. 8th Edition. New Jersey: Pearson Education Inc.
- Leedy, D.P. & Ormrod, J. E. 2010. *Practical Research Planning and Design*. 9th Edition. New Jersey: Pearson Education Inc.
- Lim, Z. 2010. *Affiliate Marketing Q & A – Will You Be Penalized For Keyword Stuffing In Landing Page*. <http://www.slideshare.net/ZackLim/affiliate-marketing-q-a-will-you-be-penalized-for-keyword-stuffing-in-landing-page> [8 April 2011].
- Maceviciute, E. 2010. Review of: Theimer, Kate. *Web 2.0 tools and strategies for archives and local history collections*. London: Facet Publishing, 2010. *Information Research*, 15(1).
<http://informationr.net/ir/reviews/revs367.html> [18 February 2011].

- Malaga, R.A. 2007. The Value of Search Engine Optimization: An Action research product at a New E-Commerce Site. *Journal of Electronic Commerce in Organizations*, 5(3):68-82.
- Malaga, R.A. 2009. Web 2.0 Techniques for search engine optimization: Two case studies. *Review of Business Research*, 9(1):132-139.
- MarketingSherpa. 2005. *MarketingSherpa study results: 3,271 marketers reveal search marketing costs, clicks, & conversions*. <http://www.marketingsherpa.com/sample.cfm?contentID=3078> [21 September 2010].
- Mathews, J. 2011. *Get on Google front page: 2011 SEO tips and tools to rise in search engine rankings*. http://books.google.co.za/books?id=6n70oyVmgAQC&pg=PT7&dq=how+long+does+it+take+to+index+a+website&hl=en&ei=f1R3TYymIpC38QPho2qDA&sa=X&oi=book_result&ct=result&resnum=8&ved=0CE8Q6AEwBw#v=onepage&q&f=false [09 March 2011].
- Mbikiwa, F. & Weideman, M. 2006. Implications of search engine spam on the visibility of South African e-commerce Websites. *South African Journal of Information Management*, 8(4). http://www.sabinet.co.za/abstracts/info/info_v8_n4_a5.html [16 June 2010].
- MegaBusiness. 2011. Keyword stuffing website. <http://www.megabsolutions.com/terminal> [15 January 2011].
- Mishne, G., Carmel, D. & Lempel, R. 2005. Blocking Blog Spam with Language Model Disagreement. *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web AIRWeb 2005 Held in conjunction with the 14th International World Wide Web Conference*, Chiba, Japan, 10 May 2005: 11-16.
- Mouton, J. 2001. *How to succeed in your Master's & Doctoral Studies: A South African guide and resource book*. Pretoria: Van Schaik.
- Murphy, H. C., Christian D. Kielgast, (2008) "Do small and medium-sized hotels exploit search engine marketing?", *International Journal of Contemporary Hospitality Management*, Vol. 20 Iss: 1, pp.90 – 97. <http://www.emeraldinsight.com/journals.htm?articleid=1662898&show=html> [20 March 2011].
- Nel, J.P., Com, D. 2007. *Epistemology Vs Ontology*. <http://mentornet.co.za/files/documents/EPISTEMOLOGYVSONTOLOGY.pdf> [16 June 2010].
- Nielsen, J. 2004. *Statistics for Traffic Referred by Search Engines and Navigation Directories to Use IT*. <http://www.useit.com/about/searchreferrals.html> [04 September 2010].
- Nielsen, J. & Loranger, H. 2006. *Prioritizing web usability*. New Riders Press: Berkeley.
- Nikulin, V. 2010. Web-mining with Wilcoxon-based feature selection, ensembling and multiple binary classifiers. *In Proceedings of the ECML/PKDD 2010 Discovery Challenge*, Barcelona, Spain, 28 September 2010:87-100.
- Olsen, W. 2004. *Triangulation in Social Research: Qualitative and Quantitative Methods Can Really Be Mixed*. <http://www.ccsr.ac.uk> [20 October 2010].
- Paulussen, S. 2004. Online news production in Flanders: How Flemish online journalists perceive and explore the Internet's potentials. *Journal of Computer Mediated Communication*, 9(4):449-466.

Perez, S.K. & Anthony, K.S. eds. 1995. *Technical Report for IRDS Conceptual Schema, Part 1: Conceptual Schema for IRDS, Part 2: Modelling Language Analysis*. New York: American National Standards Institute.

Peshkin, A. 1993. The goodness of qualitative research. *Educational researcher*, 22(2): 23-29.

Raisinghani, M.S. 2005. Future trends in search engines. *Journal of Electronic Commerce in Organizations*, 3(3).

<http://www.allbusiness.com/reports-reviews-sections/case-studies/13066978-1.html> [14 August 2010].

Raisinghani, M. S. 2005a. Search engine technology: A Closer Look at Its Future. *Information Resource Management Journal*, 18(2):i-vii.

Ralf, F., Pettijohn, W. & James, B. 2008. Search engine Optimisation: a premier on outsourcing key tasks. *Journal of Direct, Data and Digital Marketing Practice*, 10(2). <http://www.gale.cengage.com> [13 August 2010].

Ramos, A. & Cota, S. 2004. *Insider Guide to SEO: How to Get Your Website to the Top of the Search Engines*. Fremont, California: Jain Publishing Company.

Ricca, F., Tonella, P., Girardi, C. & Pianta, E. 2004. An Empirical Study on Keyword-based Web Site Clustering. In *Proceeding of the 12th IEEE International Workshop on Program Comprehension (IWPC'04)*, Bari, Italy, 24-26 June 2004:204-213.

Romow Web Directory. 2007. Search Engine Optimization NO-NOS, blog entry by Peter, 26 April. <http://www.romow.com/blog/search-engine-optimization-no-nos> [09 September 2010].

Ron, B. & Zsolt, K. 2011. The Role of Search Engine Optimization in Search Marketing. *Social Science Research Network*. <http://ssrn.com/abstract=1745644> [15 April 2011].

Sharma, D.K. & Sharma, A.K. 2010. Deep web information retrieval process: A technical survey (Report). *International Journal of Information Technology and Web Engineering*, 5(1):1-22.

Shin, Y., Gupta, M. & Myers, S. 2011. Prevalence and Mitigation of Forum Spamming 2011. School of Informatics and Design, Indiana University. Bloomington, IN, USA. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5935048 [28 April 2011].

Singh, A.M. 2002. The Internet – strategies for optimal utilization in South Africa. *South African Journal of Information Management*, 4(1). www.sajim.co.za/index.php/SAJIM/article/download/152/149 [18 October 2005].

Sitemaps. 2009. *Sitemaps Protocol*. <http://www.sitemaps.org> [13 September 2009].

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

Snack, K. 2011. *Search Engine Market Share (April 2011)*. <http://www.karmasnack.com/about/search-engine-market-share/> [07 April 2011].

SPSS Manual. 2007. SPSS Advanced Statistics 17.0: *Kaplan-Meier Survival Analysis* <http://www.hks.harvard.edu/fs/pnorris/Courses/A%20SPSS%20Manuals/SPSS%20Advanced%20Statistics%2017.0.pdf> [18 March 2011].

- Stamoulis, N. 2009. Search Engine Optimization (SEO) Training. *Search Engine Optimization Journal*.
<http://www.slideshare.net/brickmarketing/seo-training-by-brick-marketing-on-may-13-2009/download?lead=2009da08079e7ce929204ebdda212a4e17f887a90> [07 April 2011].
- Stenbacka, C. 2001. Qualitative research requires quality concepts of its own. *Management Decision*, 39(7):551-555.
- Sterling, G. 2010. Nielsen: Bing Passes Yahoo! to become #2 Search Engine.
<http://searchengineland.com/nielsen-bing-passes-yahoo-to-become-2-search-engine-50483>
 [15 September 2010].
- Setup, M. 2005. *Epistemology*. In Zalta, E. N (ed) *The Stanford Encyclopaedia of Philosophy* (Spring 2010 Edition). <http://plato.standord.edu/archives/spr2010/entries/epistemology/> [16 June 2010].
- Stolz, C. & Barth, M. 2007. Website performance analysis success assessment of information driven website on user traces. *International Journal of Information Technology and Web Engineering*, 2(3)37-52. <http://www.gale.cengage.com> [12 August 2010].
- Struwig, F.W. & Stead, G.B. 2004. *Planning, designing and reporting research*. Cape Town: Pearson Education South Africa.
- Telnic.com. 2009. *Search engine optimisation and the .tel domain*.
<http://dev.telnic.org/docs/SEO.pdf> [20 August 2010].
- Thelwall, M. 2001. Commercial website links. *Internet Research: Electronic Networking Applications and Policy*, 11(2):114-124.
- Leonard-Wilkinson, T. A. 2002. Search Engine Optimisation: Designing a Search-Friendly Site. *STC Intercom* 49(10):30-31 (ABCs of Web Design column).
- Thomas, A. B. 2004. *Research skills of Management Studies*. London: Routledge.
- Thurow, S. 2003. *What can I do to optimize my website for search engines?*
<http://www.devwebpro.com/articles/0303b.html> [09 September 2010].
- Todaro, M. 2007. *Internet Marketing Methods Revealed: The complete guide to becoming an Internet marketing expert*. Florida: Atlantic Publishing Group Inc.
- Visser, E.B. & Weideman, M. 2011. An empirical study on website usability elements and how they affect search engine optimisation. *South African Journal of Information Management* –doi: 10.4102/sajim. V13i1.428 13(1).
<http://www.sajim.co.za/index.php/SAJIM/article/view/428/459> [07 April 2011].
- Webconfs.com. 2010. *SEO Tutorial: "List of best and worst practices for designing a high traffic website"*. <http://www.webconfs.com/15-minute-seo.php> [10 September 2010].
- Weideman, M. 2004a. Empirical evaluation of one of the relationships between the user, search engines, metadata and websites in three-letter .com websites. *South African Journal of Information Management*, 6(3). www.sajim.co.za/index.php/SAJIM/article/view/312. [16 June 2010].
- Weideman, M. 2004b. Ethical issues on content distribution to digital consumers via paid placements as opposed to website visibility in search engine results. *Proceedings of ETHICOMP*, Syros, Greece, 14-16 April 2004:904-915.

Weideman, M. 2005. *Internet Searching Articles*. <http://www.mwe.co.za/html/internet-searching-articles.htm>. [09 September 2010].

Weideman, M. 2008. Internet Searching and other Research Challenges: Publish or Perish. Cape Town: Cape Peninsula University of Technology. Inaugural Speech. http://dk.cput.ac.za/cgi/viewcontent.cgi?article=1030&context=inf_papers&sei-redir=1#search=%22Internet%20Searching%20other%20Research%20Challenges%3A%20Publish%20or%20Perish.%20Cape%20Town%3A%22 [12 August 2009].

Weideman, M. 2009. *Website Visibility: The theory and practice of improving rankings*. Oxford: Woodhead Publishing Limited.

Weideman, M. & Kritzinger, W. 2003. Search Engine Information Retrieval: Empirical Research On The Usage Of Metatags To Enhance website Visibility And Ranking Of E-Commerce websites. *Proceedings of The 7th World Conference on Systemics, Cybernetics and Informatics*, Orlando, Florida, USA, 27-30 July 2003:231-236.

Weiss, A. & Weideman, M. 2008. An empirical study on the importance of keyword use in webpage naming conventions, in achieving top webpage rankings. *Working Paper*, Munich University of Applied Science, Munich. http://dk.cput.ac.za/cgi/viewcontent.cgi?article=1051&context=inf_papers [20 November 2010].

Wikipedia. 2010. *Keyword Density*. http://en.wikipedia.org/wiki/Keyword_density [27 October 2010].

Wilson, R. F. & Pettijohn, J. B. 2008. Search engine optimisation: A primer on outsourcing key tasks. *Journal of Direct, Data and Digital Marketing Practice*, 10(2):133-149.

WordStream. 2009. *Search Engine Optimization (SEO) Best Practices Eight ways to increase reach and exposure through higher natural search rankings (White Paper)*. www.wordstream.com/download/docs/wordstream-seo.pdf [15 January 2011].

Yahoo!. 2010a. Yahoo! central home page. <http://yahoomail.com/> [22 October 2010].

Yahoo!. 2010b. *Yahoo! site Submission console*. <http://siteexplorer.search.yahoo.com/> [10 October 2010].

Yahoo!. 2010c. *Report Search Abuse, search Spam and Deletions*. http://help.yahoo.com/l/us/yahoo/search/deletion/?_ylt=Auj0f4TvoNIH3wbN8271SysuqCN4 [15 April 2010].

Yahoo!. 2011. Yahoo! Search Content Quality Guidelines. <http://help.yahoo.com/l/us/yahoo/search/basics/basics-18.html> [15 February 2011]

Yung, M.E. 2011. *Search Engine Optimisation: Have you been crawled over?* <http://www.scribd.com/doc/48757721/Search-Engine-Optimisation-Have-You-Been-Crawled-Over> [01 April 2011].

Zahorsky, R. M. 2010. Search and deceive: a Web trick catches a venerable law directory. *ABA Journal*. 96(2):32-33.

Zaslavsky, D. 2010. *Pro Webmasters: How long did it take for your new website to be indexed by search engines*. <http://Webmasters.stackexchange.com/questions/1574/how-long-did-it-take-for-your-new-website-to-be-indexed-by-search-engines> [25 November 2010].

- Zhang, J. & Dimitroff, A. 2004. Internet search engines' response to metadata Dublin Core implementation. *Journal of Information Science*. 30(4):310-320.
- 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.
- Zhao, L. 2004. Jump higher: Analyzing web-site rank in Google. *Information Technology and Libraries*, 23(3):108-119.
- Zuze, H. & Weideman, M. 2010. Keyword density in website body text – search engine spamdexing? *Informatics & Design Papers and Reports*. http://dk.cput.ac.za/inf_papers/58 [15 January 2011].

APPENDICES

APPENDIX A

LETTER TO INTERVIEWEES



RE: Letter of Authorization to Conduct Interview for Research

Dear Sir/Madam

As a follow-up to a telephone conversation I had with you, earlier this week, I'd just like to repeat myself and outline a few things I introduced you to.

I am a Masters candidate at the Cape Peninsula University of Technology under the supervision of **Prof Melius Weideman**. I am hereby requesting your assistance in allowing me to conduct an interview on the area of my study. My research topic is **THE CROSSOVER POINT BETWEEN KEYWORD RICH WEBSITE TEXT AND SPAMDEXING**.

The purpose of my study is to determine how the SEO practitioners and Website developer view spamdexing, specifically keyword stuffing.

I am requesting that you provide me with information regarding the current situation of SEO, SE, your understanding of spamdexing and your experience in dealing with it at your work places. Your answers will provide the data to develop an aggregate planning model and may also identify potential areas for further model development and study.

All information gathered through the interviews will be completely confidential. For your privacy your name and organization name will not appear in the thesis or reports based on this study. Any specific quotes made will be assigned a numerical ID (e.g. Interviewee 1), and when necessary, your role (e.g. Manager of XYZ).

In appreciation of your time and effort I will provide you with an anonymous summary of all interviews I conducted. If you feel you may help in this regards, please kindly respond affirmatively to this email. I will be staging the interviews between the 13th and 18th of September 2010. Also kindly indicate the date and time in this period that will suite you.

Thank you in advance for your assistance in this study.

Sincerely,

Herbert Zuze (Student Number 210262508)

APPENDIX B

SEO QUESTIONNAIRE

SEO PRACTITIONERS SPAMDEXING KNOWLEDGE ASSESSMENT QUESTIONNAIRE

Project Name:

**THE CROSSOVER POINT BETWEEN KEYWORD RICH WEBSITE
TEXT AND SPAMDEXING**

Prepared by:

HERBERT ZUZE

Date

QUESTIONS

1. How many years have you been in the SEO industry and why did you consider being in this field?
2. What SEO tools do you regularly use and how do they work?
3. What areas do you think are currently the most important in organically ranking a site?
4. If a submitted site is not showing up on SERP, does this mean it has been banned and how long does it take for a site to be indexed and show up on the SERP?
5. Should one optimise for the singular or long tail form of keywords and why?
6. What do you understand by the phrase keyword stuffing?
7. How often do you carry out experiments on search engines to check whether their search and indexing algorithms have changed?
8. Do you follow what forums and bloggers say regarding the SEO industry?
9. Have you ever seen a blacklisted or banned website and what does it look like?
10. How do you measure the richness or poorness of a keyword in the body of a webpage?
11. How do Google, Yahoo! and Bing interpret keyword stuffing and do their algorithms stick with their respective document guidelines and procedures?
12. Do you think SEO practitioners and website developers understand spamdexing?
13. Do you think there is a crossover point from keyword rich text to spamdexing and how do you interpret it?
14. Are sitemaps helpful for webpage indexing?
15. Where do you think the SEO industry is headed?

APPENDIX C

E-MAILS TO GOOGLE, YAHOO! AND BING PROFESSIONAL EMPLOYEES

First email to Google professional employee

Google and spamdexing [Mtech BIS](#) | [X](#)

☆ **Melius Weideman** to answers-support [show details](#) 11/26/10 [Reply](#)

Hi Douwe,

I am addressing this email to you as being a Search Expert at Google. I am a research Professor at CPUT (Cape Peninsula University of Technology), specializing in research on SEO, PPC and information retrieval. If you do a search for the terms: Melius Weideman on any of the three big search engines, you will see listings of some of my research work.

One of the projects I am currently working on includes search engine interpretation of keyword stuffing as a form of spamdexing. I would appreciate your comments, from Google's side, on this issue. To make it easier, I have proposed some questions in this regard below, but feel free to address the issue in any way you like.

1. How do I know if a website has been banned by Google?
2. What is the optimum keyword density and the crossover point from a good keyword density to keyword stuffing?
3. Approximately how long does it take for a website to be visited for indexing by Google?
4. Is there any form of notification that you provide to website owners if their sites are blacklisted?

I really appreciate your time in assisting me in this matter!

Regards,
Prof M. Weideman.

Second email to Google professional employee

[Reply](#) [Reply to all](#) [Forward](#)

☆ **Melius Weideman** to answers-support [show details](#) 12/10/10 [Reply](#)

Hi Douwe,

Please refer to the message I sent to you at the end of November below.
Any comments?

Thanks,
Prof Weideman.

#####

- Show quoted text -

First e-mail to Yahoo! professional employee

Yahoo & spamdexing [Mtech BIS](#) | [X](#)

☆ **Melius Weideman** to Jeremy

[show details](#) 11/26/10

[Reply](#) ▼

Hi Jeremy,

I am addressing this email to you as being a Search Expert at Yahoo!. I am a research Professor at CPUT (Cape Peninsula University of Technology), specializing in research on SEO, PPC and information retrieval. If you do a search for the terms: Melius Weideman on any of the three big search engines, you will see listings of some of my research work.

One of the projects I am currently working on includes search engine interpretation of keyword stuffing as a form of spamdexing. I would appreciate your comments, from Yahoo's side, on this issue. To make it easier, I have proposed some questions in this regard below, but feel free to address the issue in any way you like.

1. How do I know if a website has been banned by Yahoo?
2. What is the optimum keyword density and the crossover point from a good keyword density to keyword stuffing?
3. Approximately how long does it take for a website to be visited for indexing by Yahoo?
4. Is there any form of notification that you provide to website owners if their sites are blacklisted?

I really appreciate your time in assisting me in this matter!

Regards,

Prof M. Weideman.

Second e-mail to Yahoo! professional employee

[Reply](#) [Reply to all](#) [Forward](#)

☆ **Herbert Zuze** Thanks for the emails.

11/28/10

☆ **Melius Weideman** to Jeremy

[show details](#) 12/10/10

[Reply](#) ▼

Hi Jeremy,

I sent the email below to you during the end of November.

Any comments?

Thanks,

Prof M Weideman

#####

- Show quoted text -

- Show quoted text -

[Reply](#) [Reply to all](#) [Forward](#)

First e-mail to Bing professional employee

Bing and spamdexing [Mtech BIS](#) | [X](#)

☆ **Melius Weideman** to bingbot [show details](#) 11/26/10 [Reply](#)

Hi Dean/Fabrice,

I am addressing this email to you as being a Search Expert at Bing. I am a research Professor at CPUT (Cape Peninsula University of Technology), specializing in research on SEO, PPC and information retrieval. If you do a search for the terms: Melius Weideman on any of the three big search engines, you will see listings of some of my research work.

One of the projects I am currently working on includes search engine interpretation of keyword stuffing as a form of spamdexing. I would appreciate your comments, from Bing's side, on this issue. To make it easier, I have proposed some questions in this regard below, but feel free to address the issue in any way you like.

1. How do I know if a website has been banned by Bing?
2. What is the optimum keyword density and the crossover point from a good keyword density to keyword stuffing?
3. Approximately how long does it take for a website to be visited for indexing by Bing?
4. Is there any form of notification that you provide to website owners if their sites are blacklisted?

I really appreciate your time in assisting me in this matter!

Regards,
Prof M. Weideman.

Melius Weideman, PhD

Second e-mail to Bing professional employee

☆ **Melius Weideman** to bingbot [show details](#) 12/10/10 [Reply](#)

Hi Dean/Fabrice,

Please refer to the message I sent to you at the end of November below.

Any comments?

Thanks,
Prof Weideman.

#####

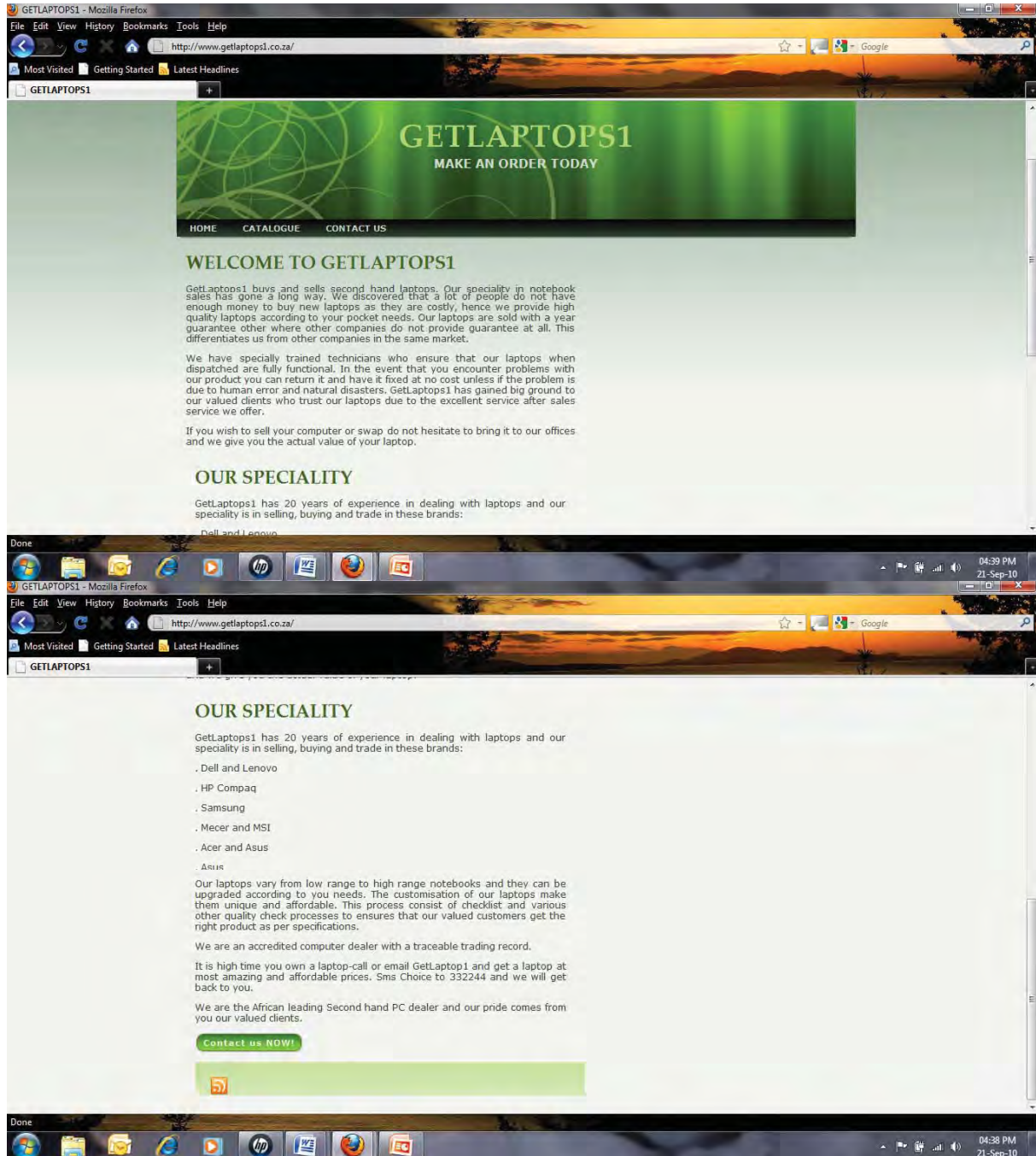
- Show quoted text -
- Show quoted text -

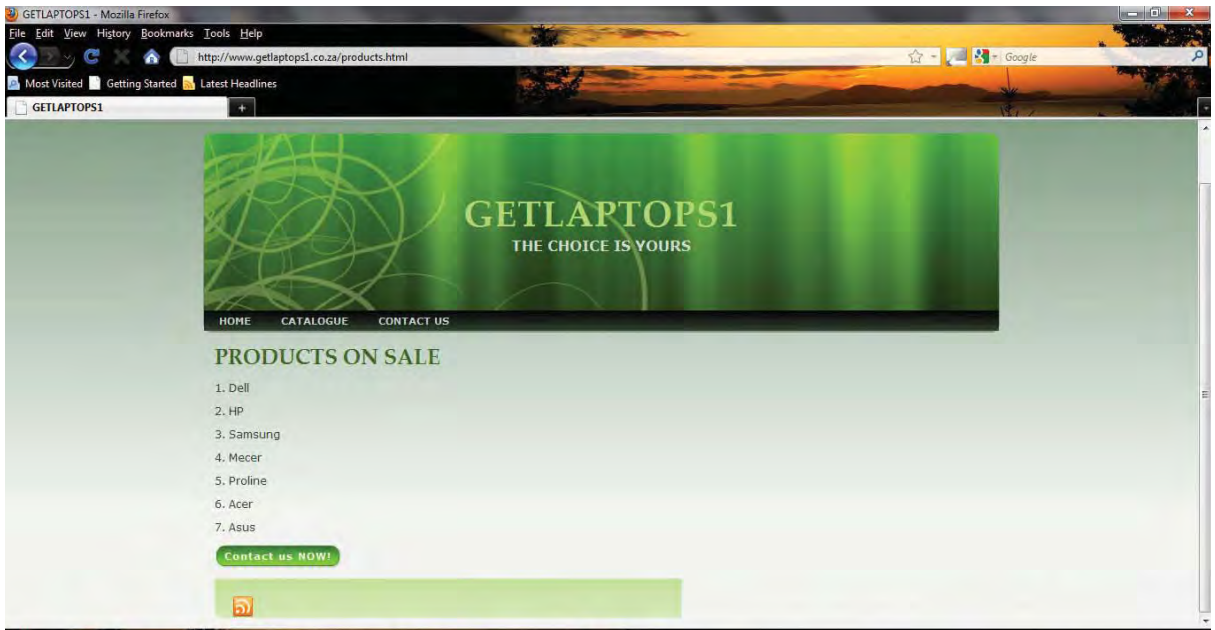
----- Forwarded message -----
From: "Melius Weideman" <WeidemanM@cput.ac.za>
To: <bingbot@microsoft.com>
Date: Fri, 26 Nov 2010 13:57:07 +0200
Subject: Bing and spamdexing
Hi Dean/Fabrice,

APPENDIX D

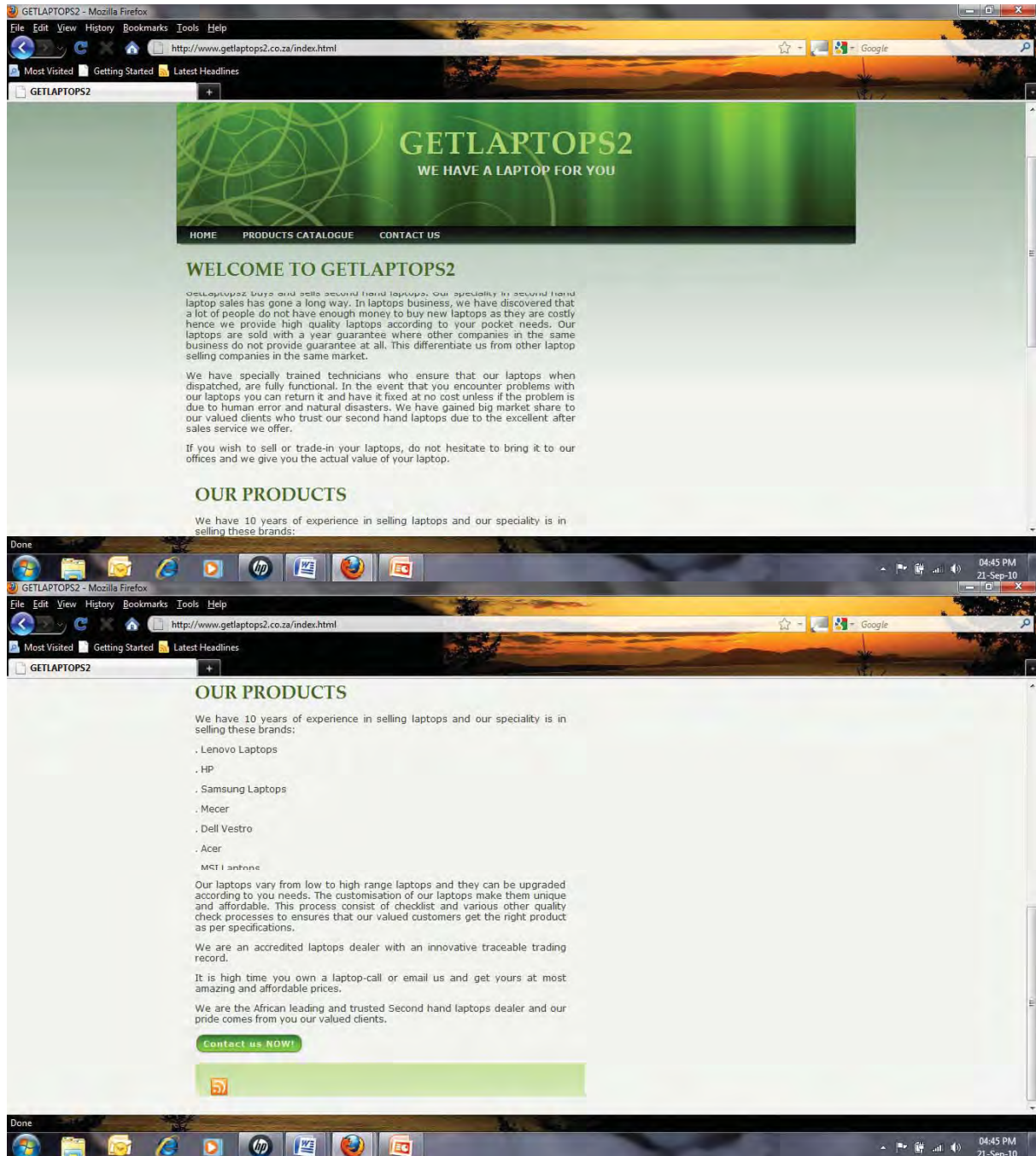
PHASE1 WEBSITES SNAPSHOTS

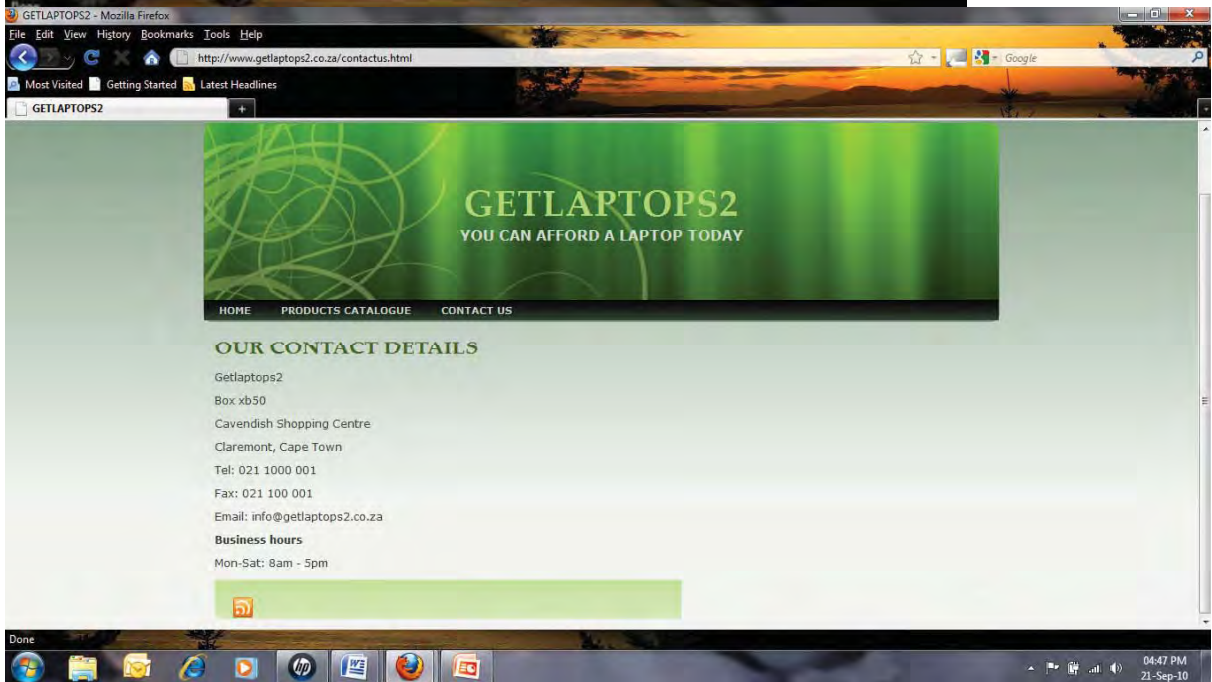
Getlaptops1



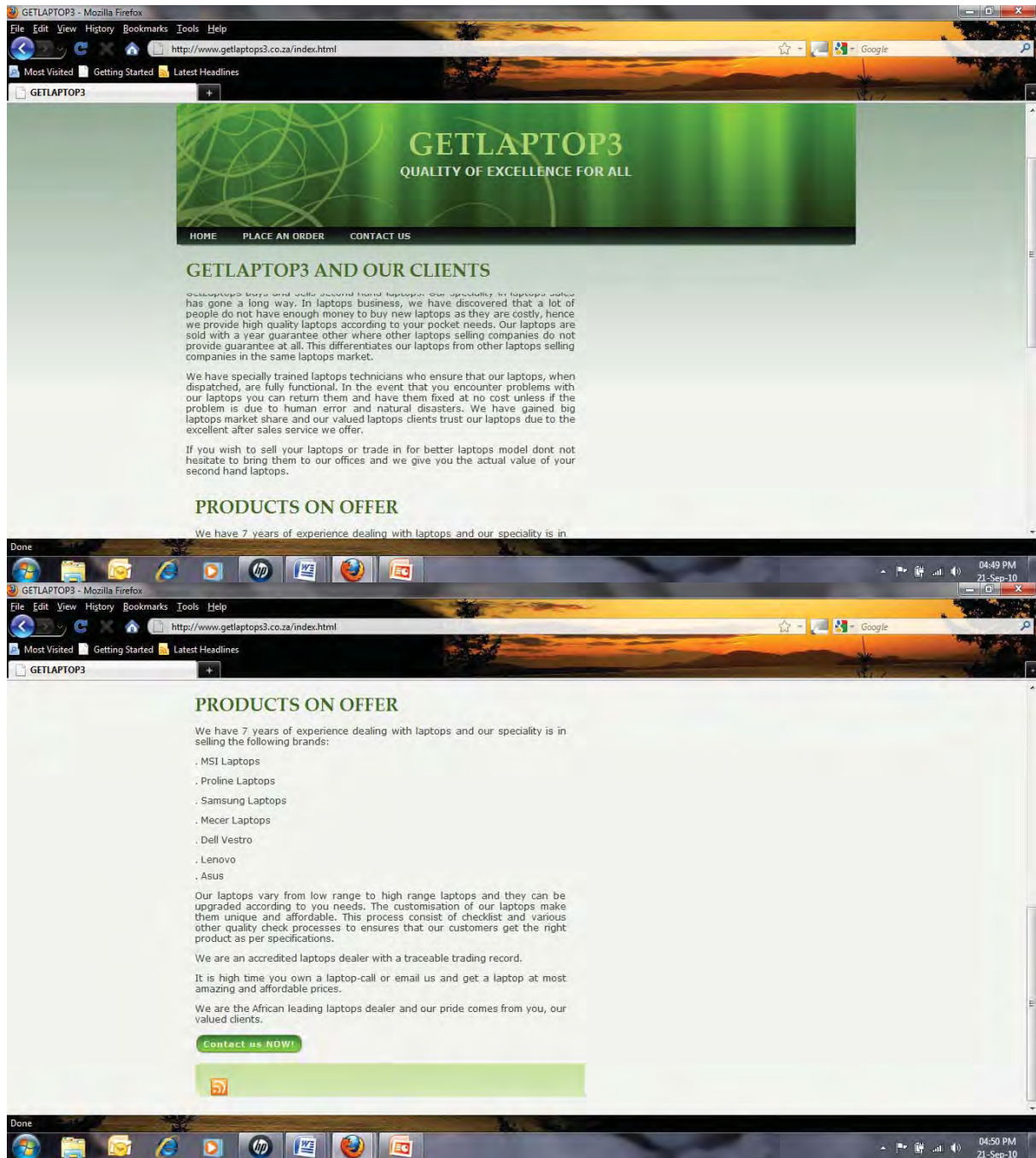


Getlaptops2





Getlaptops3



GETLAPTOPS3 - Mozilla Firefox
 http://www.getlaptops3.co.za/products.html

GETLAPTOPS3

BEST DEALS

1. Lenovo for R1500 each
2. Asus for R2500 each
3. Samsung for R500 each
4. Dell Vestro for R2500 each
5. HP Compaq for R1000 each
6. Acer for R500 each
7. Asus for R3500 each

[Contact us NOW!](#)

Done
 GETLAPTOPS3 - Mozilla Firefox
 http://www.getlaptops3.co.za/contactus.html

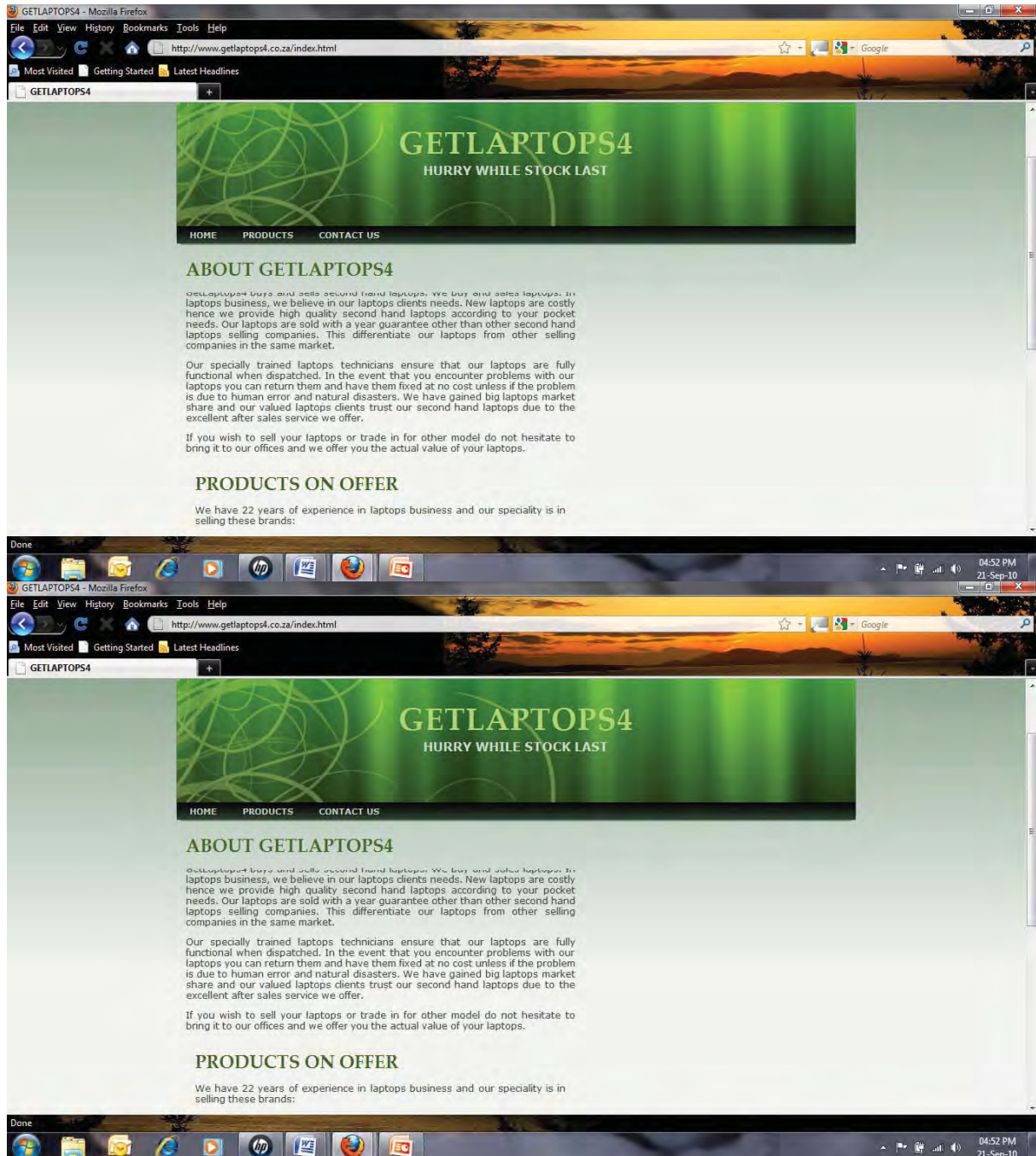
GETLAPTOPS3

OUR CONTACT DETAILS

Getlaptops3
 Box 71
 Crossroads Shopping Centre
 Table View, Durban
 Tel: 031 1000 002
 Fax: 031 100 002
 Email: info@getlaptops3.co.za

Opening hours
 Mon-Sat: 9am - 5pm

Getlaptops4



GETLAPTOPS4 - Mozilla Firefox
http://www.getlaptops4.co.za/products.html

GETLAPTOPS4

YOU CAN AFFORD A LAPTOP

HOME PRODUCT CONTACT US

BEST DEALS

1. Lenovo stock available
2. Asus stock available
3. Samsung stock available
4. Dell Vostro limited stock
5. HP Compaq stock available
6. Acer stock available
7. Asus out of stock

[Contact us NOW!](#)

Done
04:52 PM
21-Sep-10

GETLAPTOPS4 - Mozilla Firefox
http://www.getlaptops4.co.za/contactus.html

GETLAPTOPS4

VISIT US NOW

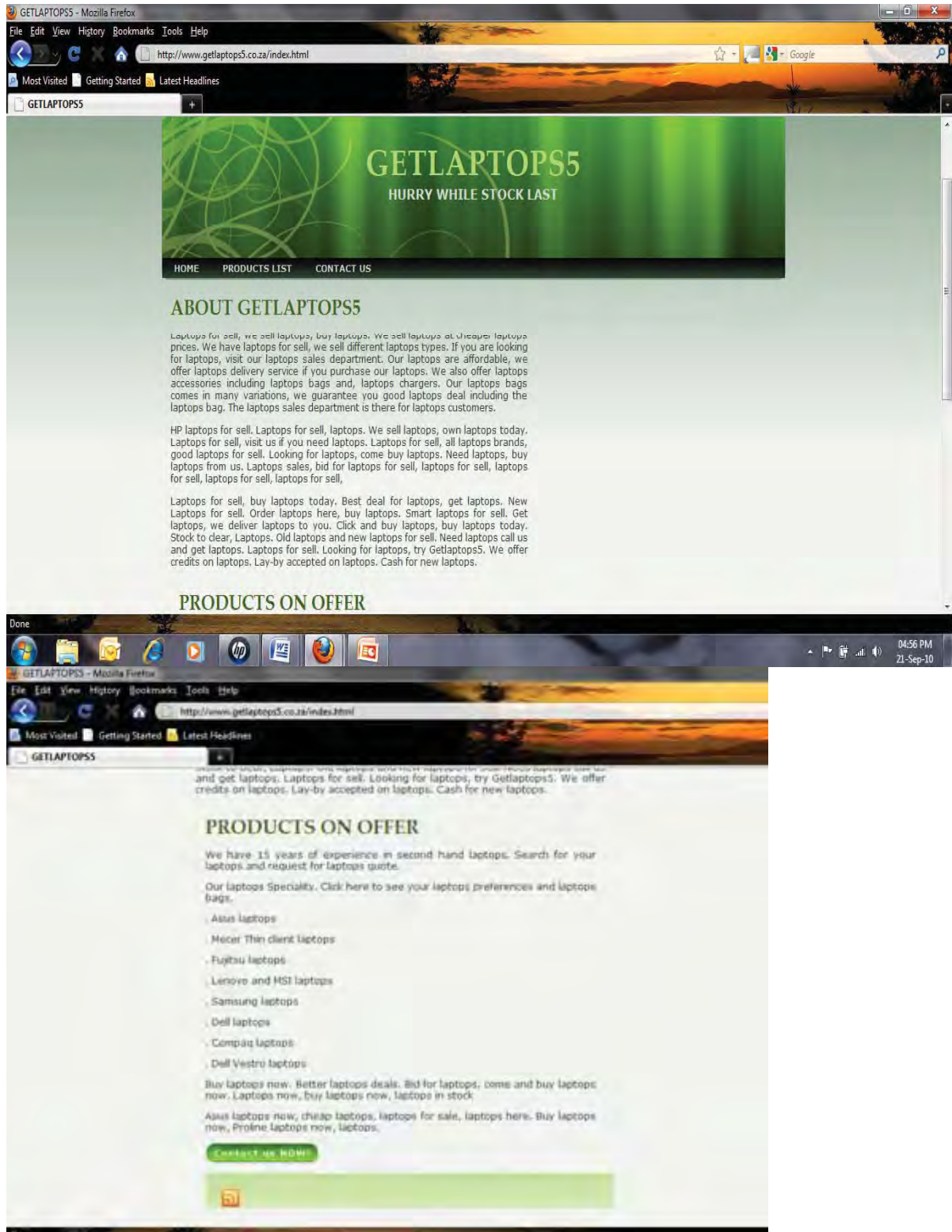
HOME PRODUCTS CONTACT US

CONTACT DETAILS

Getlaptops4
Box 2009
Warrior street
Port Elizabeth
Tel: 051 1000 0042
Fax: 051 100 042
Email: info@getlaptops4.co.za

Business Hours
Mon-Sat: 9am - 4pm

Getlaptops5



GETLAPTOPS5 - Mozilla Firefox
http://www.getlaptops5.co.za/products.html



HOME PRODUCT LIST CONTACT US


BEST DEALS

1. Lenovo out of stock
2. Asus
3. Samsung stock available
4. Dell Vostro limited stock
5. HP Compaq stock available
6. Acer stock available
7. Asus out of stock

[Contact us NOW!](#)

Done
04:55 PM
21-Sep-10

GETLAPTOPS5 - Mozilla Firefox
http://www.getlaptops5.co.za/contactus.html



HOME PRODUCTS LIST CONTACT US

CONTACT DETAILS

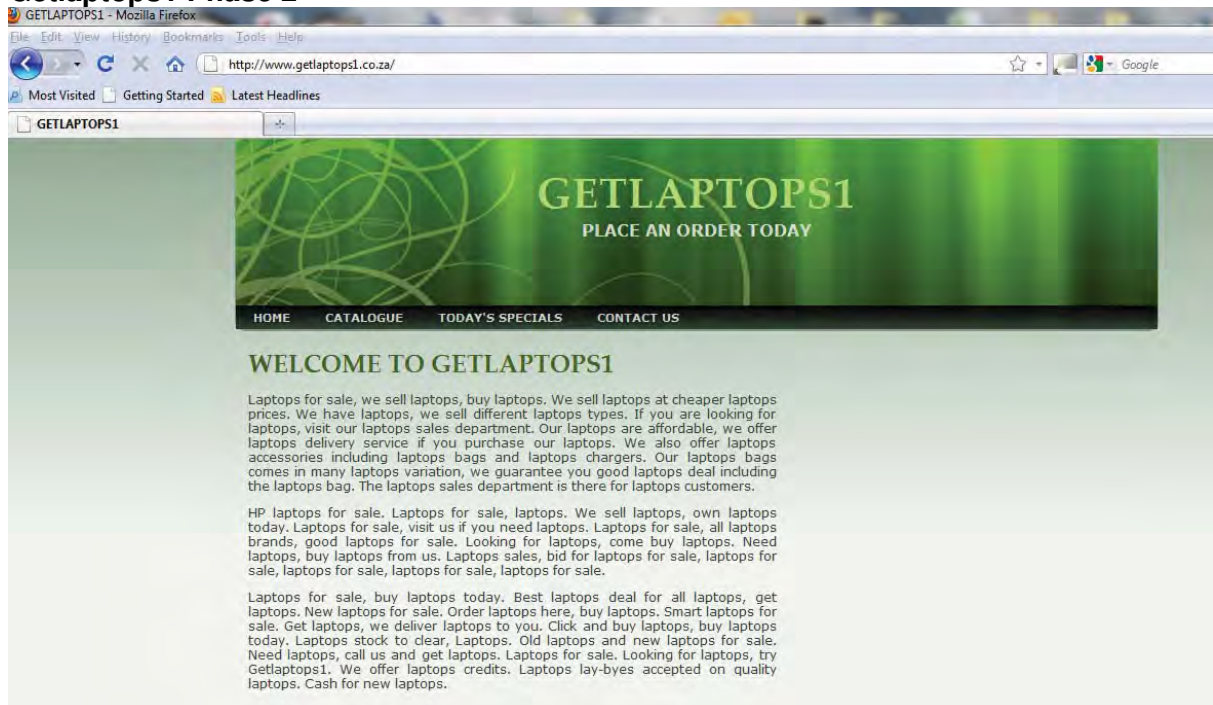
Getlaptops5
Box B509
Andrews street
Pretoria
Tel: 081 1007 082
Fax: 081 1007 082
Email: info@getlaptops5.co.za

Business Hours
Mon-Sat: 8am - 4pm

APPENDIX E

PHASE 2 WEBSITES SNAPSHOTS

Getlaptops1 Phase 2



OUR SPECIALITY

Search for your laptops and request a laptop quote. Get laptops today, buy laptops. Cheap laptops here. Great laptops deals. Our laptops Speciality. Below are laptops preferences and laptops bags.

- Asus laptops
- Mecer laptops
- Fujitsu laptops
- Lenovo laptops
- Samsung laptops
- Dell laptops
- Compaq laptops

Buy laptops now. Better laptops deals. Bid for laptops, come and buy laptops now. Laptops now, buy laptops now, laptops in stock. Laptops, yes laptops. Buy laptops.

Asus laptops now, cheap laptops, laptops for sale, laptops here. Buy laptops now, laptops now, laptops.

[Contact us NOW!](#)



GETLAPTOPS1 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.getlaptops1.co.za/getlaptops1_todayspecials.html

Most Visited Getting Started Latest Headlines

GETLAPTOPS1

GETLAPTOPS1

PLACE AN ORDER TODAY

HOME CATALOGUE TODAY'S SPECIALS CONTACT US

TODAY'S SPECIALS

BRAND NAME	MODEL	COST	QUANTITY IN STOCK
ACER	ASPIRE 53H-1729 NETBOOK	R5900	5
	ASPIRE AS5742-7120	R9000	6
DELL	INSPIRON 15 NOTEBOOK	R10000	2
	LATITUDE E5510	R11900	1
FUJITSU	LIFEBOOK AH530	R15000	3
	LIFEBOOK AH530	R12300	4
SONY	AVIO VPCM121AX/P NETBOOK	R17000	4
	AVIO VPCEE23FX/BI	R9600	7
HP	MINI 110-1144NR NETBOOK	R6720	2
	PROBOOK 4520s	R12500	6

You can also look at the following products in this week specials

- Docking stations
- Network adapters
- Graphics cards and
- LCD Monitors

Getlaptops2 Phase 2

GETLAPTOPS2 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.getlaptops2.co.za/index.html

Most Visited Getting Started Latest Headlines

GETLAPTOPS2

HOME PRODUCTS CATALOGUE OUR CHRISTMAS SPECIALS ABOUT US

WELCOME TO GETLAPTOPS2

High-quality laptops, super laptops, buy laptops. Good laptops. Buy laptops, more laptops. Cheaper laptops, different laptops types. First-class laptops, visit our laptops sales department. Affordable laptops, get laptops, laptops delivery also available. Superior laptops. Laptops accessories, laptops bags and laptops chargers. Laptops bags come in many laptops variation. Laptops guarantee is included. Cheap laptops bag. Fine laptops for excellent laptops clients.

HP laptops. Laptops for sale, laptops. Excellent laptops, own laptops today. Outstanding laptops, visit us for laptops. Brilliant laptops, all laptops brands, good laptops. First-choice laptops, come buy laptops. Need laptops, buy laptops from us. Laptops sales, bid for laptops, exceptional laptops, discounted laptops, admirable laptops, laptops for sale.

Inexpensive laptops, buy laptops. Best laptops deal for all laptops, get laptops. New Laptops and old laptops. Order laptops, buy laptops. Smart laptops for sale. Get laptops, we deliver laptops. Low-cost laptops, inexpensive laptops. Laptops stock to clear, Laptops. Old laptops and new laptops for sale. Need laptops, call us and get laptops. Superior laptops, fancy laptops. Looking for laptops, try Getlaptops5. We offer laptops credits. Laptops Lay-byes accepted on quality laptops. Cash for new laptops.

OUR LAPTOPS

Get laptops quote. Get laptops today, buy laptops. Cheap laptops here. Great laptops deals. Our laptops speciality. Laptops preferences and laptops bags.

- Asus laptops
- Mecer laptops
- Fujitsu laptops
- Lenovo laptops
- Dell laptops
- Compaq laptops

Amazing laptops, unique laptops, impressive laptops. Worthy laptops, creditable laptops, praiseworthy laptops, estimable laptops.

Buy laptops now. Better laptops deals. Bid for laptops, buy laptops, buy laptops, marvellous laptops. Laptops now, buy laptops now, laptops in stock. Laptops, yes laptops. Buy laptops. Good laptops, cheap laptops, smart laptops. Dell laptops, HP laptops, Mecer laptops, Lenovo laptops, Compaq laptops, Samsung laptops, Fujitsu laptops, Asus laptops, Acer laptops, MSI laptops.

Asus laptops, cheap laptops, laptops for sale, laptops here. Buy laptops now, laptops now, laptops. Special laptops, incomparable laptops and estimable laptops. Extraordinary laptops, commendable laptops, splendid laptops and marvellous laptops.

[Contact us NOW!](#)

GETLAPTOPS2 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.getlaptops2.co.za/Getlaptops2_OurChristmasSpecials.html

Most Visited Getting Started Latest Headlines

GETLAPTOPS2

HOME CATALOGUE OUR CHRISTMAS SPECIALS CONTACT US

OUR CHRISTMAS SPECIALS

BRAND NAME	MODEL AND NUMBER	COST PRICE	NUMBER IN STOCK
DELL	VESTRO 530 NETBOOK	R12 300	5
	INSPIRON MINI 10 NETBOOK	R10 900	8
HP	MINI 110-3030NR NETBOOK	R13 960	1
	ENVY 10-1110NR	R9 420	5
FUJITSU	LIFEBOOK LH530	R8 600	9
	LIFEBOOK NH570	R9 254	4
ACER	ASPIRE AT&T NETBOOK	R10 000	12
	ASPIRE 5734 AS5734ZM	R9600	2
SONY	AVIO VPCY216FX/V NETBOOK	R7 892	6
	AVIO VPCEA22FX/P	R5 680	1

We are also selling the following products

- CDs and DVDs
- Sound Cards
- Switches and

Getlaptops3 Phase 2


GETLAPTOPS3 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.getlaptops3.co.za/index.html

Most Visited Getting Started Latest Headlines

GETLAPTOPS3



HOME PLACE AN ORDER SUMMER SPECIALS OUR DETAILS

GETLAPTOPS3 CLIENTS

Aspire laptops, Mesh-black laptops, Timelinux laptops, As5742 laptops, MacPro laptops, Mac-Air laptops, Macbook laptops, AspirePro laptops, GatewayPro laptops, AS578645 laptops, Mb9962LL/A laptops, Core-Processor laptops, I3-Processor laptops, Dual-Core laptops, Pentium-Processor laptops, i5-Processor laptops, i7-Processor laptops.

Dual-Processor laptops, laptops now. Better laptops, laptops deals. AsusPro laptops, buy laptops, Thin-Light laptops, Bamboo laptops, buy laptops, laptops stock. Laptops, yes laptops. AS2F-XA1 laptops, Good laptops, cheap laptops, smart laptops. Fujitsu laptops, Dell laptops, HP laptops, Mecer laptops, Lenovo laptops, Compaq laptops, Samsung laptops, Asus laptops, Acer laptops, MSI laptops. Laptops, black laptops. White laptops, Laptops. U35F laptops and new laptops. Acer laptops, Macbook laptops. Laptops, superior laptops, fancy laptops. Lenovo laptops, laptops, try Getlaptops3. Mesh laptops, laptops. Green laptops, covered laptop. Colourful laptops, laptops, thin-light laptops.

Inspiron laptops, Ideapad laptops, Studio laptops. Grey laptops, super laptops, buy laptops. Y560 laptops. G5550-Series laptops, more laptops. Edge laptops, laptops, different laptops. ThinkPad laptops, our laptops sales. Mars-Black laptops, Lenovo laptops, laptops delivery, Super laptops. HP Laptops, laptops bags, laptops mouse. Laptops keyboards. Kindle laptops, laptops, laptops hard-drives.

HP laptops, Free laptops, laptops. Gs50-Series laptops, laptops. Outstanding laptops, laptops. Samsung laptops, all laptops, IdeaPad-550 laptops. First-choice laptops, buy laptops. Need laptops, buy laptops. Laptops sales, bid laptops, exceptional laptops, discounted laptops, admirable laptops, laptops chargers.

Studio laptops, laptops motherboard, laptops memory. Satellite laptops, laptops, fusion laptops. TruBrite Laptops and old laptops. Order laptops, buy laptops. Potege laptops Meta-Grey laptops, omega laptops. Qosmio laptops, Omega-Black laptops.

LAPTOPS ON OFFER

True laptops, laptops quote. Fusion-DS123 laptops, laptops, buy laptops. Gemini-Red laptops, laptops here. General laptops, laptops covers. Laptops list, laptops bags.

- Laptops
- Laptops
- Laptops
- Laptops
- Laptops
- Laptops
- Laptops

Sony-VPC laptops, unique laptops, pink laptops. VAIO laptops, white laptops, praiseworthy laptops, brown laptops.

Sony-AVIO laptops, AS5570 laptops, laptops, laptops here. HP-Probook laptops, laptops, laptops now, red laptops. MSI laptops, gaming laptops, laptops. Dark-Brown laptops, commendable laptops, GXd7800 laptops and marvellous laptops. Versatile Entertainment laptops, touch-screen laptops, republic-gamers laptops, laptops. Toshiba laptops, laptops chargers.

[Contact us NOW!](#)

GETLAPTOPS3 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.getlaptops3.co.za/Getlaptops3_SummerSpecials.html

Most Visited Getting Started Latest Headlines

GETLAPTOPS3

GETLAPTOPS3
CREDIBILITY, TRUST AND EXCELLENCE

HOME CATALOGUE SUMMER SPECIALS CONTACT US

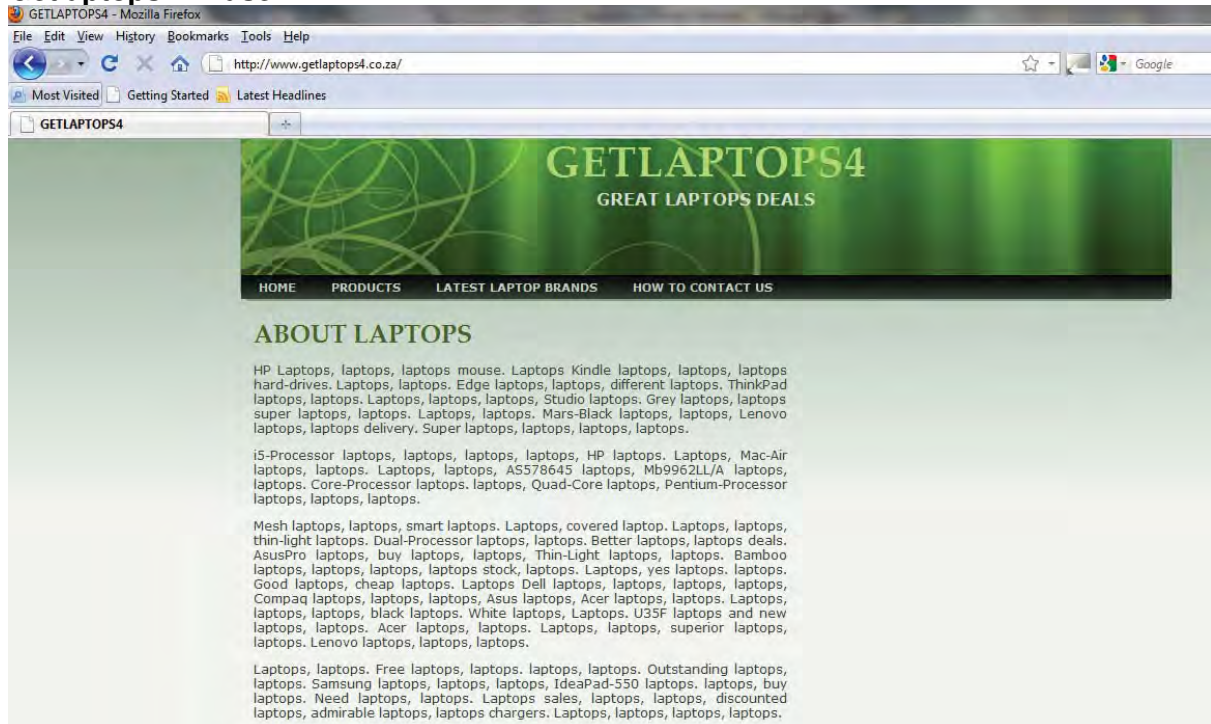
SUMMER SPECIALS

BRAND NAME	MODEL	COST	STOCK
SONY	AVIO VPCX369TX/V NETBOOK	R9 892	5
	AVIO VYHCEA22FX/P	R15 680	11
HP	COMPAQ PRESARIO CQ62-210US NETBOOK	R3 960	6
	ENVY 1450-2510NR	R29 420	15
FUJITSU	LIFEBOOK AAS9830	R18 600	29
	LIFEBOOK CDF3670	R19 254	44
DELL	VETRO 3500 NOTEBOOK	R2 300	12
	STUDIO 17 NOTEBOOK	R9 900	2
ACER	ASPIRE AS4820T NETBOOK	R15 00	1
	ASPIRE ES1551-4650	R7 600	12

CHECK OUT OUR NEW PRODUCTS

- Hubs
- Notebook memory
- Pc memory and
- Hard drives

Getlaptops4 Phase 2



The screenshot shows a Mozilla Firefox browser window with the address bar displaying <http://www.getlaptops4.co.za/>. The browser's menu bar includes File, Edit, View, History, Bookmarks, Tools, and Help. Below the address bar, there are tabs for 'Most Visited', 'Getting Started', and 'Latest Headlines'. The main content area features a green header with the text 'GETLAPTOPS4 GREAT LAPTOPS DEALS' and a navigation menu with links for 'HOME', 'PRODUCTS', 'LATEST LAPTOP BRANDS', and 'HOW TO CONTACT US'. The main body of the page is titled 'ABOUT LAPTOPS' and contains three paragraphs of text, each starting with a different laptop brand or feature: 'HP Laptops', 'i5-Processor laptops', and 'Mesh laptops'. The text in these paragraphs is repetitive and appears to be placeholder or filler text.

GETLAPTOPS4
GREAT LAPTOPS DEALS

HOME PRODUCTS LATEST LAPTOP BRANDS HOW TO CONTACT US

ABOUT LAPTOPS

HP Laptops, laptops, laptops mouse. Laptops Kindle laptops, laptops, laptops hard-drives. Laptops, laptops. Edge laptops, laptops, different laptops. ThinkPad laptops, laptops. Laptops, laptops, laptops, Studio laptops. Grey laptops, laptops super laptops, laptops. Laptops, laptops. Mars-Black laptops, laptops, Lenovo laptops, laptops delivery. Super laptops, laptops, laptops, laptops.

i5-Processor laptops, laptops, laptops, laptops, HP laptops. Laptops, Mac-Air laptops, laptops. Laptops, laptops, AS578645 laptops, Mb9962LL/A laptops, laptops. Core-Processor laptops, laptops, Quad-Core laptops, Pentium-Processor laptops, laptops, laptops.

Mesh laptops, laptops, smart laptops. Laptops, covered laptop. Laptops, laptops, thin-light laptops. Dual-Processor laptops, laptops. Better laptops, laptops deals. AsusPro laptops, buy laptops, laptops, Thin-Light laptops, laptops. Bamboo laptops, laptops, laptops, laptops stock, laptops. Laptops, yes laptops. laptops. Good laptops, cheap laptops. Laptops Dell laptops, laptops, laptops, laptops, Compaq laptops, laptops, laptops. Asus laptops, Acer laptops, laptops. Laptops, laptops, laptops, black laptops. White laptops, Laptops. U35F laptops and new laptops, laptops. Acer laptops, laptops. Laptops, laptops, superior laptops, laptops. Lenovo laptops, laptops, laptops.

Laptops, laptops. Free laptops, laptops. laptops, laptops. Outstanding laptops, laptops. Samsung laptops, laptops, laptops, IdeaPad-550 laptops. laptops, buy laptops. Need laptops, laptops. Laptops sales, laptops, laptops, discounted laptops, admirable laptops, laptops chargers. Laptops, laptops, laptops, laptops.

LAPTOPS ON OFFER

Studio laptops, laptops, laptops motherboard, laptops memory. Satellite laptops, laptops, fusion laptops. TruBrite Laptops, laptops. laptops, buy laptops. Potege laptops laptops, omega laptops. laptops, laptops, Omega-Black laptops, laptops. True laptops, laptops quote. laptops, laptops, buy laptops. Gemini-Red laptops, laptops here. General laptops, laptops, laptops, laptops, covers. Laptops laptops, laptops bags.

- Laptops, laptops, laptops, laptops, laptops,
- Laptops, laptops, laptops, laptops, laptops,
- Laptops, laptops, laptops, laptops, laptops,
- Laptops, laptops, laptops, laptops, laptops,
- Laptops, laptops, laptops, laptops, laptops,
- Laptops, laptops, laptops, laptops, laptops,
- Laptops, laptops, laptops, laptops, laptops,

Sony-VPC laptops, laptops, laptops. VAIO laptops, laptops, laptops, brown laptops. Laptops, AS5570 laptops, laptops, laptops here. laptops, laptops, laptops now, laptops. MSI laptops, laptops, laptops. laptops, commendable laptops, laptops and marvellous laptops. Versatile-Entertainment laptops, touch-screen laptops, laptops, laptops. Toshiba laptops, laptops chargers. laptops, laptops, laptops, laptops, laptops,

[Contact us NOW!](#)



GETLAPTOPS4 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.getlaptops4.co.za/Getlaptops4_Latest_LaptopBrands.html
Most Visited Getting Started Latest Headlines
GETLAPTOPS4

GETLAPTOPS4

GREAT LAPTOPS DEALS

HOME CATALOGUE LATEST LAPTOP BRANDS CONTACT US

LATEST LAPTOP BRANDS

BRAND NAME	MODEL	COST	STOCK
FUJITSU	NOTEBOOK MJY9630	R5 986	2
	NETBOOK OPF3670	6 994	4
HP	COMPAQ PRESARIO CQ62-392HGT NETBOOK	R7 890	1
	ENVY 5924-689LPR	R2 920	1
SONY	AVIO VPCFG2560TX/V NETBOOK	R5 592	5
	AVIO VYEDSEA2562FX/P	R3 680	44
ACER	ASPIRE BG48820T	R8 000	1
	ASPIRE BG356T NOTEBOOK	R15 900	12

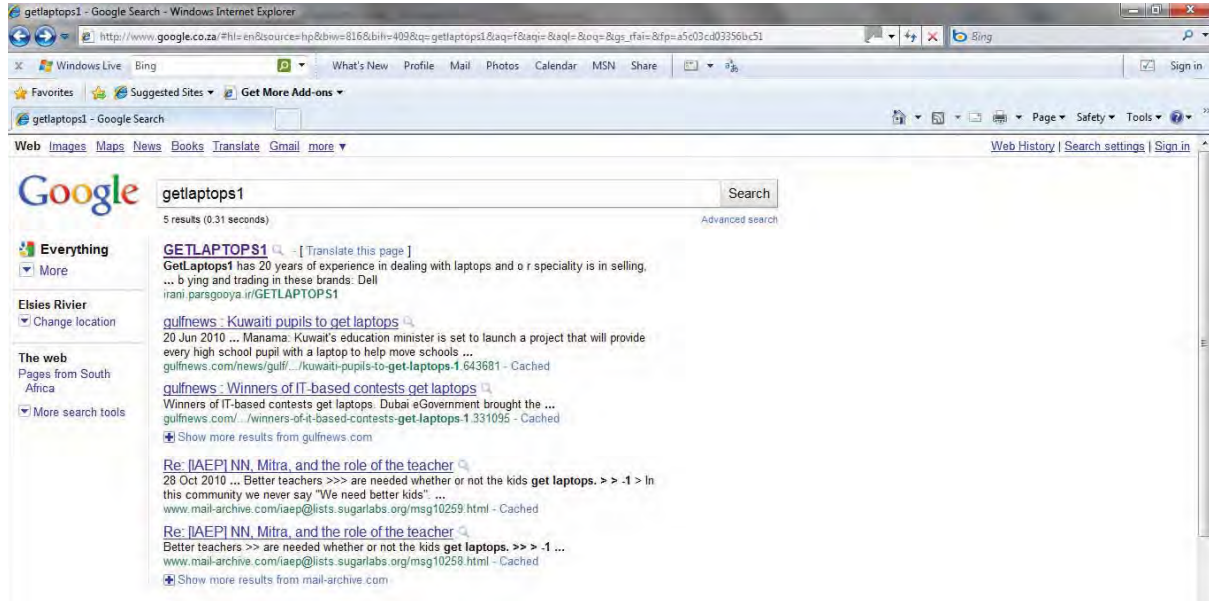
STOCK TO CLEAR

- CPU cooling fans and heat sink
- Flash disks
- Flash disks
- External hard drives

APPENDIX F

IRANIAN SCRAPING SITE SNAPSHOTS

Iranian Website





برای پیش بردن برونس، ایجاد سوسن سانسیتی، پرست سر برین روبریا، استناد به نفس، آرامش، تعادل، شادی و سلامتی، شجاعت، چشم زخم، محافظت در برابر حوادث، تقویت اعصاب، دستگاه گوارش، ریه، استخوان، کمر، ششونایی، چشم ها نگهداری: نگهداری از عقیق بسیار ساده است: شما می توانید آن را در هفته یک بار در آب جاری بشوید و یا معرض در اشعه خورشید قرار دهید. عقیق برای بیماری ها و ناتوانی های چشمی بسیار مؤثر است و همچنین دارویی تقویتی برای دستگاه تولید مثل می باشد. می توانیم از این سنگ برای درمان سر دردها، گرفتگی های عضلانی و دردهای کمر و ستون فقرات استفاده کنیم. برای معالجه بیماری های صرع، کلیه، روده، پرستات، احتقاق و ورمها نیز می توانیم از آن کمک بگیریم. سنگ عقیق ریه ها، معده، کبد و مغز را تقویت می کند و محرکی برای دستگاه گوارش است و بیان سخنوری را تسهیل می کند. ارتعاشات این سنگ استداد و موانع درونی را از بین می برد و انرژی منفی را از ما دور می کند. توصیه می شود آن را در آبرتمان ها بکار برد تا از نیروهای منفی به دور باشید. این سنگ از تجمع کوآرتزهایی که با طرافت خاصی منطقه ای شکل گرفته است تشکیل می شود. و از گروه کلسدوئین (سیلیسهای نیم شفاف متیلور شده ای که در عهد باستان در امر جواهر سازی از آن بسیار استفاده می کردند) است

تعداد:

23200 تومان

ارسال با پست

جایزه های ویژه

جراج نونال کور



فروش بی واسطه نونال کور 2 وارداتی اصل نونال از کمپانی معتبر UZAY SPOR ترکیه نونال کر وسیله ای شگفت انگیز است که ورزش دراز و نشست را برای شما شیوه سازی می کند با این تفاوت که با این دستگاه و درجالت نشستنه هیچ فشاری به ستون فقرات شما وارد نمی شود و در عرض چند هفته می توانید نتیجه آن را بر روی اندام خود مشاهده نمایید. نونال کر جدید ترین و جالب ترین وسیله ورزشی مناسب اندام در دنیا می باشد. قیمت 108000 تومان قیمت جدید فقط 79000 تومان این دستگاه بهترین وسیله برای عضلانی و کوچک شدن شکم و پهلوها می باشد که بدون فشار آوردن به ستون فقرات و عضلات گردن و همچنین بدون ایجاد حساسیتی مفرط برای فرد استفاده کننده در بازه زاویه ای 90 درجه انجام می شود. همگامی این وسیله ورزشی بر اساس حرکت رفت و برگشت دو جرخ موجود در قسمت پشت صندلی این دستگاه است. برخی از امکانات و قابلیتها: وزن قابل تحمل این وسیله 120 کیلوگرم، کوچک کردن شکم (به راحتی) ساخت کشور تایوان، تحت لیسانس موسسه مارشال آلمان، بدون ایجاد حساسیتی مفرط برای فرد، نمایانگر دیجیتالی تعداد دراز و نشست، میزان کالری مصرفی شده و مدت زمان ورزش، صندلی ضد عرق و ضد حساسیت از جنس EVA، توب ورزشی جهت تعریبات ورزشی، DVD آموزشی و تعریبات دستگاه، راهمک فارسی به همراه برنامه رژیم غذایی بهترین

تعداد:

79000 تومان

ارسال با پست

بک مستند کامل bbc life



این مجموعه به تلاش موجودات زنده برای ادامه بقا می پردازد . تلاش های شکار کردن و شکار نشدن ، تصاویری بسیار مهیج و زیبا از صحنه های شکار جانداران ، جنگیدن برای زنده ماندن در شرایطی طاق فرسا ، تلاشی که جاندار دارین از آن به عنوان مبارزه برای وجود نام می برد. این مجموعه نیز از تکنولوژی بسیار فیلمبرداری استفاده شده است که برای اولین بار در Planet Earth به کارگیری شده بود علاوه بر این در این مجموعه از تکنیک های منحصر به فردی استفاده شد تا تصاویری متفاوت و بسیار واقعی را به مخاطب خود نشان دهند . تکنیک هایی مثل فیلمبرداری ثابت با هلیکوپتر ، حس پرواز در کنار پروانه ها و کارهایی که اجازه می داد تا در کنار جانورانی از جمله گوزن شمالی به فیلمبرداری از آنها بپردازند و به نوعی جزئی از محیط جانوران شوند. در این مجموعه از دوربین های پر قدرتی که امکان تصویر برداری تا 8000 فریم در ثانیه را دارند ، استفاده شده است که این قابلیت نیز در شکار صحنه های جذاب شکار که تا کنون دیده نشده اند خیلی مفید و مؤثر بود. این مجموعه باورنکردنی برای اولین با بصورت کامل در 10 قسمت با کیفیت در حد HD و با زیرنویس کامل فارسی عرضه میشود و در تمامی دستگاه های خانگی و کامپیوتر قابل اجرا میباشد چالش های زندگی (Challenges of Life): این قسمت با

تعداد:

15400 تومان

ارسال با پست

1000 مکان قبل از مرگ باید دید



مجموعه ای فوق العاده دیدنی از زیباترین و دیدنی ترین شهرها و مکان های دنیا . زوج جوان و میلیونی که در تصمیمی شگفت انگیز خانه و محل زندگی خود را ترک کرده و به همراه تجهیزات کامل و گروه فیلمبرداری تصمیم دارند تا پایان عمرشان به 1000 مکان دیدنی در سراسر دنیا سفر کنند. در این مجموعه زیباترین شهرها ، مکان ها ، جنگلها ،سواحل حتی مردمان آن مناطق و آداب و رسوم آنان به صورت کاملا زنده و مستند به تصویر کشیده شده. در این مجموعه سفر به 13 کشور و منطقه دیدنی جهان از جمله استرالیا ،برزیل ،کامبوجا ،کانادا ،فرانسه ،آلاسکا ،برو ،آفریقای جنوبی ،ایتالیا ،تایلند ،هاوایی ،هندوستان و مکزیک به همراه تمام صحنه ها و خاطرات حین سفر و تمامی مکانها و جاهای دیدنی این مناطق با کیفیت فوق العاده HD می باشد. برخی از جاهای دیدنی..... Cherry Blossom Viewing, Japan The Pyramids of Giza, Egypt St. Andrew's Cathedral, Singapore Palacio de Cristal, Retiro Park, Madrid Siena, Italy St Basil's Cathedral, Moscow, Russia این مستند ارزشمند در 6 دی وی با بالاترین کیفیت عرضه می گردد

تعداد: 10500 تومان

ارسال با پست

برترین مستند حیات وحش bbc planet



Planet Earth در سال 2006 توسط بخش مستند کانال BBC و به کارگردانی Alastair Fothergill (کارگردان شهیر فیلم های مستند) با بودجه 25 میلیون دلار در طی مدت 5 سال ساخته شد. George Fenton برای این مجموعه موسیقی باشکوهی را تصنیف کرده است که لذت تماشای این چنین اثری را چندین برابر کرده است. در زمان تماشای این مجموعه بارها و بارها سکانس های باورنکردنی را خواهید دید که گوش سپردن به این موسیقی متن (با کیفیت صدای دالسی دیجیتال 6 کاناله) بر هیجان شما خواهد افزود این مجموعه در 62 کشور و 204 محل متفاوت فیلمبرداری شده است. گوینده این مستند David Attenborough یکی از شناخته شده ترین گویندگان و طبیعت شناسان جهان می باشد و لقب Sir را نیز دریافت کرده است. این مستند دارای 11 اپیزود (قسمت) می باشد. در اولین اپیزود یک دید عمومی درباره مناطقی که در 10 اپیزود بعدی قرار است به آنها پرداخته شود، داده می شود. در هر کدام از 10 اپیزود بعدی به بررسی یکی از بوم های طبیعی روی زمین و جانورانی که در آن ناحیه زندگی می کنند، می پردازد. در این مستند رفتارهایی از حیوانات که تا پیش از این دیده نشده بود به تصویر کشیده شده است. هم مانند تعقیب گوزن توسط گرگ ها از بالا، بچه خرس های گریزلی که برای اولین بار از غار بیرون می آیند، حمله ناگهانی صدها شمشیرماهی و کارشناسان

تعداد: 10200 تومان

ارسال با پست

عینک ۳ بعدی آنالگراف (فیلم ها را سه بعدی ببینید)



عینک 3 بعدی + درایور 3 بعدی ساز کامپیوتر آیا تا کنون فیلمهای 3 بعدی را تجربه کرده اید؟ لذت دیدن فیلم ها و عکسهای 3 بعدی را با عینکهای آنالگراف تجربه کنید آیا دوست دارید فیلم های 3 بعدی مثل آواتار را تماشا کنید؟ یکج استثنایی آنالگراف شامل یک عینک 3 بعدی و 1 عدد DVD آیا می دانید 50 درصد فیلمهای پرفروش هالیوود در سال 2009 و 2010 سه بعدی بوده اند 70 درصد بازیهای جذاب موجود در بازار 3بعدی هستند!!! شامل یک عینک 3 بعدی و 1 عدد DVD شامل درایور 3 بعدی ساز کامپیوتر از این به بعد قادر به تماشای فیلمهای 3 بعدی خواهید بود

تعداد: 10000 تومان

ارسال با پست

شعبه باران قرن بیستم



۲- مجموعه‌ی تمام شعبده بازها که اجرا شده توسط کریس آنجل در 4 dvd (بیش از ۳۰ ساعت با کیفیت عالی) از جمله کارهای اجرا شده: - جدا کردن بدن انسان در پارک - پرواز کریس آنجل در آسمان - راه رفتن روی آب - زن و مردی که در یک نقطه 15 متر به عقب برده می شوند بدون اینکه خودشان متوجه بشوند - کریس آنجل و شیشه - خوردن سوراخ ها و سپس بیرون آوردن آنها از شکم - کریس آنجل و پرواز 2 - فردی که کاری را در ذهنش انتخاب می‌کند و کریس آن کارت را بدون دخالت دست از بین تمام کارت‌ها بیرون می‌کشد. - یک خانوم روی یک کاغذ پروانه‌ای میکشند و بعد کریس آن کاغذ را تا میکند بعد از باز کردن کاغذ یک پروانه واقعی پرواز میکند و جانب اینکه که نشان می‌دهد از کاغذ پاک میشود - پرواز از بالا به پایین و از پایین به بالا بر روی یک باساز بلند - کریس آنجل و آتش - کریس آنجل و سازه - کریس آنجل در عرض 1 ثانیه وارد ماشین حمل پول میشود (مانندینی که هیچ راه نفوذی ندارد) - بازیگر هالیوود از بین چندین هزار سعی دگ یکی را انتخاب و کریس آنجل به او می‌گوید که چه سیدی را انتخاب کرده است. - عبور شیشه از وسط سیر - مجموعه‌ای شامل بردستی های حرفه‌ای دیوید کاپر فیلم David Copperfield در ۱ DVD « عبور از دیوار چین » غیب کردن قطار - هواپیما و مجسمه آزادی آمریکا در حضور مردم

تعداد 9900 تومان

ارسال با پست

رازهای شعبده بازی و جادوگری



مستندی بسیار جالب در رابطه با آشکار کردن رازهای شعبده بازی، لایم به ذکر است که تمامی این ترفندها توسط یکی از شعبده بازیان حرفه می شود و بدلیل ترس از به خطر افتادن حرفه و جان خودش در نمایش این مجموعه که شامل 3 قسمت هستای آموزش داده ماسکوی بر صورت دارد! این مجموعه بدین صورت هست که شما در ابتدا شعبده بازی و نمایش می کنید و سپس بار دیگر به طور مفصل چگونه انجام اون به شما نمایش داده میشود، پیشنهاد میکنم حتما نمایش کنید مجموعه شکستن راز شعبده باز ، بزرگ ترین رازهای شعبده بازی را افشا می کند! تا کنون هیچ شعبده بازی جرات نمایش آن چه که در پشت صحنه شعبده بازی اتفاق می افتد را نداشته است. اما اکنون، «شعبده باز نقاب دار» رازهایی را که بینندگان را مسحور خود می ساختند به تصویر می کشد. همیشه این سوال بوده که چگونه ممکن است یک انسان نصف شود، یک قبل غیب شود، توری از بدن انسان رد شود، از یک تانک آب فعل شده فرار کرد و الان برای اولین بار همه ها و پشت پرده این تردستی ها را خواهید دید این مجموعه بی نظیر در 5 دگ وی دگ با قیمت 9300 تومان با بهترین کیفیت عرضه می گردد .

تعداد 9300 تومان

ارسال با پست

مستند زیباترین سواحل جهان



زیباترین سواحل مجموعه فوق العاده زیبایی و زیبا از زیباترین سواحل جهان یک مستند فوق العاده برای کسانی که دنبال آرمش هستند صحنه ها و مناظر این جزایر به قدری زیباست که گویی تکه ای از بهشت را نظاره می کنید و هیچ راهی سخن نمی گوید چون تصاویر گویای همه چیز است فورنگ زیبا و آرامش بخشی که در فیلم بخش میشود آرمش مطاعفی را به بیننده القا می کند این مجموعه با کیفیت HD بخش می شود. این مجموعه در 3 دگ وی دگ با کیفیت بسیار بالا و با قیمت 7500 تومان عرضه می گردد

تعداد 7500 تومان

ارسال با پست

ورود به دنیای شگفت انگیز

ورود به دنیای شگفت انگیز



مجموعه مستند استثنایی دیگر از BBC با کیفیت بسیار بالا به علاقه‌مندان عرضه شد ، نوع فیلمبرداری غیر قابل باور است شما با این نوع فیلمبرداری به شناخت عمیق و بسیار نزدیک موضوعات بحث شده می رسید همگان به بزرگی خنایند فعال ایمان داریم اما این بزرگی را وقتی بیشتر احساس می کنیم که با دنیای اطراف و ناشناخته او آشنا شویم ، در این مستند شما به زوایای زندگی ، بشر ، حیوانات ، تاریخ و ... برخورد خواهید کرد که در تصویر نگارش و حتی در تصویر افکار شما تاثیر خواهد گذاشت بزرگی دنیای پیرامون را بیشتر حس خواهید کرد و به قدرت و تفکر خدای جوش بیشتر از گذشته دست پیدا می کنید . این مستند در 3جله دی وی دگ با کیفیت بالا و شامل موارد زیر است: 1- بهترین مکانهای جهان (Greatest Places) کیفیت: 720p2 سفر به عتراهای جذاب (Journey Into Amazing Caves) کیفیت : 720p3 طبیعت صخره های مرجایی در دریا (Greatest Places) کیفیت : 720p4 زندگی حشرات (Bugs) کیفیت : 1080p5 عجايب رود نیل (Mystery Of The Nile) کیفیت : 720p6 اکتشاف سرعت (Super Speedway) کیفیت : 720p7 پرش از ارتفاع و چترپارگی (Adrenaline Rush) کیفیت : 720p8 جنگل زیبای زیر دریا (Forest Ocean) کیفیت : 720p9 سیرزمین های عجیب در اقیانوسها (Wonderland) کیفیت : 1080p10 سرعت (Speed) کیفیت : 1080p11 زندگی دریایی (The Living Sea) کیفیت : 1080p این مجموعه مستند در 3 dvd با فرمت d1x و همچنین کیفیت واقعی

تعداد 6100 تومان

ارسال با پست

اسرار فضا : مستند ریباک ناسا



این مستند در آمریکا بنام ناسا چه چیزی را مخفی می کند معروف است . این مجموعه به بررسی فعالیت های ناسا و اقدامات مخفی ناسا در طی سالها متمادی پرداخته است و به بررسی فضای بیکران پرداخته است . این محصول به یکی از زیباترین مستند هایی است که در رابطه با هوا و فضا ساخته است و دیدن آن را به دوستداران این گونه مستندات توصیه می کنیم. این مجموعه در 3 دی وی با کیفیت بسیار بالا و قابل بحثی در پلیرهای خانگی و کامپیوتر و نسخه اصلی می باشد .

تعداد: 6100 تومان

ارسال با پست

Copyright © 2005 - 2009 parsگویا.com. Designed by parsگویا.com

Translated pages



The screenshot shows a Google Translate interface. At the top, there is a Persian banner with the text: "تهرانی ها کمتر از 24 ساعت محصول سفارشی خود را دریافت می نمایند تمامی محصولات با قابهای بسیار شیک و محکم می باشد". A red banner below it contains several colorful advertisements for products like "کامپیوتر", "مشقک دیبای", "مهره مار", "تورنچین ریباک", and "با خواص شفا بخش". A green arrow points to a close button on the left side of the red banner.

If you click close (shown by the arrow) on the red banner, you get to

تهرانی ها کمتر از 24 ساعت محصول سفارشی خود را دریافت می نمایند تمامی محصولات با قابهای بسیار شیک و محکم می باشد

Advertise

- World's most beautiful beaches Documentary
- 1000 places see before you die
- Total Blind Auction
- Top wildlife documentary bbc planet
- Naglyf 3D glasses (see three-dimensional videos)
- Full documentation pack bbc life
- Training new scarf and scarves Close

Talking Shop Pars

Snake charm bead generator (Article original)



Not sure about the snake ever heard? Extra attention without the authority of people you attract the opposite sex and emotions this is not a superstition exact explanation for this miracle to read .. price: USD 16,000. Vertebra snake causing appeal (Article original) snake vertebrae treated first meeting may reveal issues superstitious, but the metaphysics of modern science, including modern science is human Vjdyd Vvaqyat hide many of the issues and scientific discovery Vasbat Mymand Stone Therapy Bsyarjzy one of the branches that are now able Many human acute problems to solve about metaphysics like snake vertebrae Ntr is stated: Hrshy' the universe due to its structural essence is the magnetic force behind Vbraynd the energy released from the bodies as effective Vashya' Vqabl observed disorder can develop in the cosmos expression in the Koran that they also mentioned that the verses had to look to the stars Vzmyndrasmanha with the permission of Allah Qadmtal invisible columns are mentioned. constrictor Vmadrh male because their anatomic Azshrayt specific situation to have a different motivation for blending the Weber.its Pydakndd overcome the situation.

Home

Necklace

Advertise

- Log Wonderful World
- Twentieth century magicians
- New aerobics and weight training

- Land terrible Anaconda
- Secrets of magic and sorcery
- Who is the most powerful woman in Egypt?
- Life Pstandam
- Penguin parade
- Desert Documentary Valves
- Documentary civilization - CIVILIZATION
- Planet Earth's strong
- Secrets of Space: NASA's beautiful documentary
- Twentieth century magicians
- Log Wonderful World
- Log Wonderful World

USD 16,000
ارسال با پست

- Planet Earth's strong
- Documentary civilization - CIVILIZATION
- Desert Documentary Valves
- Penguin parade
- Life Pstandam
- Who is the most powerful woman in Egypt?
- Bracelet Beautiful Turquoise 2 Stone of Healing
- Very beautiful necklace of turquoise stones healing cutting eaten
- Beautiful stone bracelet healing or fluorine (Fivrayt)
- Beautiful bracelets tiger eye stone cut
- Beautiful stone bracelet healing if cut quartz (2)
- Beautiful stone bracelet cut Mytyst (2)
- Beautiful stone bracelet healing Zbrjd
- Beautiful stone bracelet healing or fluorine Fivrayt
- Beautiful stone bracelet healing Opal (Opal) Stone wishes
- Beautiful bracelets mobile hematite healing stones
- Agate stone bracelet healing
- Beautiful stone bracelet healingo
- Beautiful stone bracelet healing Bsydari
- Beautiful stone bracelet healing Mytyst

Bracelet Beautiful Pink Quartz Healing Stone



Pink Quartz Properties: This stone is a great combination of energy healing Brhy effects mentioned are: gain energy and provide positive energy lost by improving facial and body stain also for treatment of stomach diseases (except ulcer) strengthening positive wave energy; gain positive energy and provide energy lost by Usage: If a piece of quartz into a glass of cold water thrown and after ten minutes outside Nodehi stone and water to drink and rock for twenty minutes on his third eye place. improved stain the face or body treatment for stomach diseases (except ulcer) Usage: A piece of quartz to ten minutes in the hot noon sun and then put it into some warm milk (not boiled mode) put ten minutes, (if the milk of sheep or goats is better). Then out of the stone has three tablespoons a day before each meal and eat for an hour then eat something else. This method can repeat up to five days. Reinforce positive energy waves Usage: If a piece of quartz into a piece of blue cloth or white cotton and laid eyes on his third eye for twenty minutes. Better lie down somewhere quiet. Recommendation:: If a piece of quartz into a piece of blue cloth or white cotton and laid eyes on his third place for twenty minutes. Better lie down somewhere quiet. Recommendation:: Magnetic Force

RS 9900
ارسال با پست

- Beautiful stone bracelet healing or fluorine Fivrayt
- Beautiful stone bracelet healing Opal (Opal) Stone wishes
- Beautiful bracelets mobile hematite healing stones
- Agate stone bracelet healing
- Beautiful stone bracelet healingo
- Beautiful stone bracelet healing Bsydari
- Beautiful stone bracelet healing Mytyst

Very beautiful necklace of turquoise stones healing cutting eaten



1. Turquoise Stone Energy Properties available in this stone is useful for the following: stimulating Vrah Chakrahay forces and people start to relax, awaken the inner forces people treat various gastric diseases such as: ulcers, swelling, pain and movement Cha Cry ... stabilize the individual types of energy healing and treatment of headache, migraine, muscle aches and bone to treat constipation Awakening Third Eye Vrah stimulate up human forces and Chakrahay Usage: A piece of turquoise stones to ten minutes thrown into a glass of water and then drink the water and rocks on the third eye chakra fifth or twenty minutes for your escape. relax , waking up those inner forces of energy available in this stone also relax, causing awaken the inner forces of forces of energy available in this stone also relax, causing awaken the inner forces of the people are. This energy having energy raiser, some of the internal energy and the person Ratqvyt electromagnetic circuits around the body sends. Habas energy exchange this kind of negative energy going to kill and replace the body waking Baanrzhayhay raiser and forces be built (if the course

USD 24,000
ارسال با پست

- Bracelet Beautiful Pink Quartz Healing Stone
- Snake charm bead generator (Article original)
- Secrets of magic and sorcery
- Agate necklace eaten Lathe
- Land terrible Anaconda
- Human destiny after death
- Training new scarf and scarves Close

Agate necklace eaten Lathe



Among the winners of fear, cause lucky, blessed day in the provision, confidence, relaxation, balance, happiness and health, courage, eye ulcers, protects against Hvdastqvyt: neurology, gastrointestinal tract, lung, bone, back, hearing, eye maintenance: maintenance of the agate is very simple: you can make it once a week, running water or wash exposed place in the sun rays. Aghigh for eye diseases and disabilities is very effective and also for strengthening dna production unit is like. We can treat the stone head nains.

muscle cramps and even pain and spine to ease to ease... kidney, bladder, colon, prostate, and Vermha Ahtgaq can also take the help. Stone Aghigh lungs, stomach, liver and brain will strengthen and drive to the gastrointestinal tract and facilitates eloquent expression. Vibration of this stone and obstruction of internal barriers and eliminates the negative energy away from us will. It is recommended apartments used to cut the negative forces around you. Kvarztz'hay accumulation of stone with a special regional delicacy is formed will be formed, and the group Klsdyvyn (Syls'hay a half-transparent crystallized Testament Ancient jewelry in the much used it) is

USD 23,200

Total

ارسال با پست

Naglyf 3D glasses (see three-dimensional videos)

Top wildlife documentary bbc planet

Total Blind Auction

1000 places see before you die

World's most beautiful beaches Documentary

Suggested products

Total Blind Auction



Direct selling Core 2 total imports from Taiwan companies originally valid UZAY SPOR Trkyhtvtal chorus is amazing tool that exercise sit for you to simulate with the difference mode with this device and sat no pressure to your spine is not within a few weeks can result it can find on their limbs. TOTAL deaf through new and most interesting sport in the fitness world. 108 000 Price USD 79,000 USD New Price Only this device and the best tool for muscle shrinkage of abdominal and flank is no pressure to the spine and neck muscle fatigue and excessive for a person without using a 90 degree angle in the interval is performed. mechanism based on the sport by moving the two-wheel trip In the back seat is the device. Some features and capabilities: the intolerable weight of 120 kg by stomach shrink (easily) made in Taiwan under license from the German Marshall Institute for the individual without causing viewer fatigue sit-digital number, calories consumed and exercise duration and antiperspirants chairs made of EVA allergy ball exercise DVD for sports education and training system with Persian Guide main diet plan

USD 79,000

Total

ارسال با پست

Full documentation pack bbc life



This collection attempts to continue living organisms survive deals, trying to hunt down and not hunting, is very exciting and beautiful images of animal chase scenes, fighting to survive in the absence of overwhelming, that Charles Darwin's attempt to fight for it Despite the name .. this collection take too much video technology has been used for the first time on Planet Earth had been employing in addition Mjvh unique techniques were used to extremely different and real images to contacts his show. Proven techniques such as filming with a helicopter, flying along feeling butterflies and things allowed near animals such as moose and filming them pay some kind of animals are part of the environment. In this collection of powerful cameras which allows imaging up to 8000 frames per second have been used in the hunting scenes also feature attractive hunting so far have not been very useful and effective. The incredible first set with 10 episodes in full quality the full range of HD and Farsi with subtitles are offered in all devices and home computers is applicable life challenges (Challenges of Life); This Section

USD 15,400

Total

ارسال با پست

1000 places see before you die



Wonderful and interesting collection of the most beautiful and most interesting cities and places of the world. Millionaire couple and the surprising decision at home and leave their place of residence and with complete equipment and camera crews have decided to end their lives in a picturesque location in 1000 to travel around the world. In this beautiful collection of cities, places, forests, coastal areas, even its people and customs as they live and fully documented depicted. In this collection of 13 country and regional travel to see the world, including Australia, Brazil, Cambodia, Canada, France, Alaska, Peru, South Africa, Italy, Nepal, Hawaii, India and Mexico, along with all the scenes and memories during the trip and all sightseeing places and areas with HD quality is fantastic. Some sightseeing Cherry Blossom Viewing, Japan The Pyramids of Giza, Egypt St. Andrew's Cathedral, Singapore Palacio de Cristal, Retiro Park, Madrid Siena, Italy St Basil's Cathedral, Moscow, Russia This documentary valuable sixth highest quality DVD will be released

USD 10,500

Total

ارسال با پست

Top wildlife documentary bbc planet



Planet Earth in 2006 by sector and directed documentary channel BBC Alastair Fothergill (Director famous documentary film) with a budget of 25 million dollars! During the 5 years was built. George Fenton for the noble collection composed music has such a pleasure watching the effect has been multiplied. When watching this series again and again you will see incredible sequences that listening to the soundtrack (with six-channel Dolby Digital sound quality) will be added on your excitement! The set in 62 countries and 204 different locations has been filming. Narrator David Attenborough documentary, one of the most well-known speakers and experts of nature and the world is also the title Sir is received. It has a 11 episode documentary (part) is. In the first episode of a general vision about where the next episode 10 is supposed to be dealt with, will be given. Which of the 10 episodes each subsequent review of the natural ecology on land and animals that live in the area, pays. The behavior documented in animals that had not been seen before this has been portrayed. Like: chasing deer by wolves from the top, baby grizzly bears for the first time come out of the cave, and sudden attack hundreds Shmshymahy Experts

USD 10,200

Total

ارسال با پست

Naglyf 3D glasses (see three-dimensional videos)



3D Glasses + 3D driver instrument PC ever 3D films you've experienced? Enjoy seeing videos and pictures with 3D glasses to experience Naglyf you like movies like 3D avatar watch? Naglyf Exceptional Package including glasses 3D 1 x DVD you know 50 percent of films selling Hollywood in 2009 and 2010 three-dimensional have been 70 percent Bazyhay attractive on the market 3D are! including glasses 3-D and 1 x DVD includes Drivers 3 Next maker of computer capable of the following films will be 3D

USD 10,000

Total

ارسال با پست

Twentieth century magicians



3 - sets all magic performed by Chris Angel games in 4 dvd (more than 20 hours with excellent quality), including work performed: - separate the human body in the park - flying in the sky Chris Angel. Walk on water, and women. The man



who in a moment, 15 meters will be taken back without realizing they are subject - Chris Angell and glass - and then eat needles out of their afferent abdominal - Chris Angell and Flight 2 - A person who chooses the card in his mind, and Chris the card without the intervention of between draws out all the cards. - A woman stabbed on a paper butterfly and then Chris is the paper to a paper after the opening butterfly flies true and interestingly butterfly painting paper is clear - R. flight top down and bottom-up on a long passage - Chris Angell and Fire - Chris Angell and shadows - Chris Angell within seconds into a money machine is carrying (the machine has no influence) - Hollywood actor among Several thousand CDs and Chris Angell select one says what he has chosen CD. - glass passes from the middle of the table - including a set of professional sleight David Field David Copperfield Capri in 1 DVD »wall passing» Unseen by train - Statue of Liberty, America's aircraft and the people

RS 9900

Tota

ارسال با پست

Secrets of magic and sorcery



Very interesting documentary about revealing secrets of magic, It is worth mentioning that all these tricks by a professional magicians be compromised due to fear of professional and life itself in all this collection includes three episodes taught Hstay puts on a mask is the face! this collection is as follows the first trick you can watch and then again in greater depth how do you display it offline. I propose to break the mystery series sure to watch the magician, the greatest secrets will expose magicians! Ever dare display any magic behind the scenes what magic is not occurring. But now, "masked magician" secrets that your visitors are enchanted depicts built. The question was always how a person may be half an invisible elephant, the body of Thierry is rejected, one water tank escaped and locked Now for the first time behind the trick and will juggle the see this unique collection of 5 DVDs in 9300 USD price with best quality are presented.

RS 9300

Tota

ارسال با پست

World's most beautiful beaches Documentary



Beautiful beaches and beautiful collection of fantastic dream world of beautiful beaches a great documentary for those who are seeking relaxation scenes and landscapes so beautiful that it Chzayr as if watching a piece of paradise and you do not speak no narrator says as images tell Everything is beautiful and soothing music in the film is broadcast to viewers Mzafy induced relaxation does this collection of HD video is played. This set of three DVDs in High Definition and will be released price rs 7500

RS 7500

Tota

ارسال با پست

Log Wonderful World



Another BBC documentary series with exceptionally high quality were presented to those interested. Type your camcorder is incredible video with this kind of deep understanding and very close to the issues discussed could be as big as everyone believe in God, but great when we feel more and unknown around the world get to know him. In the context of this documentary you, man, animals, history and Approach will that change the attitude and even in changing your thoughts will affect the world's great about the most sense will and power and thought god his most of the past reach there. This documentary in three rings with DVD quality include the following: 1 - the world's best venues (Greatest Places) Quality: 720p2-trip to fascinating caves (Journey Into Amazing Caves) Quality: 720p3-natural coral reefs in the sea (Greatest Places) Quality: 720p4-living insects (Bugs) Quality: 1080p5-Wonders of the Nile (Mystery Of The Nile) Quality: 720p6-end speed (Super Speedway) Quality: 720p7-jump height and skydiving (Adrenaline Rush) Quality: 720p8-beautiful forest under the sea (Sea Forest) Quality : 720p9-lands on a strange ocean (Ocean Wonderland) Quality: 1080p10-speed (Speed) Quality: 1080p11-Marine Life (The Living Sea) Quality: 1080p This documentary series on 3 dvd divx format as well as the actual quality

RS 6100

Tota

ارسال با پست

Secrets of Space: NASA's beautiful documentary



NASA called this documentary in America to hide what is known. This series examines the activities of NASA and NASA's actions in hiding after many years has paid and paid review space is infinite. This product is one of the most beautiful documentaries that are with air and space and made it to the fans to see these documents we recommend. This 3 DVD collection with high quality and playable on computer and home Plyrhay is original.

RS 6100

Total

ارسال با پست



APPENDIX G

WEBSITES PHASE 1 SAMPLE CODE

Websites Phase 1 Sample Code

Index Page for Getlaptops5

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="en-US" xml:lang="en">
<head>
<!--
Designed by Herbert Zuze
http://validator.w3.org : "This page is valid XHTML 1.0 Transitional"
-->
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<meta http-equiv="X-UA-Compatible" content="IE=EmulateIE7" />
<meta name="google-site-verification" content="_4uD0h_2QVxJovhZuxCEKy3RvQszlFOOs-uVEj2Hxx8" />
<title>GETLAPTOPS5 </title>

<link rel="stylesheet" href="style.css" type="text/css" media="screen" />
<META NAME="KEYWORDS" CONTENT="Laptops,laptop,getlaptops5,computers,notebook">
</head>
<body>
<div id="art-page-background-simple-gradient">
</div>

<div id="art-main">
<div class="art-Sheet">
<div class="art-Sheet-tl"></div>
<div class="art-Sheet-tr"></div>
<div class="art-Sheet-bl"></div>
<div class="art-Sheet-br"></div>
<div class="art-Sheet-tc"></div>
<div class="art-Sheet-bc"></div>

<div class="art-Sheet-cl"></div>

<div class="art-Sheet-cr"></div>
<div class="art-Sheet-cc"></div>
<div class="art-Sheet-body">
<div class="art-Header">
<div class="art-Header-png"></div>
<div class="art-Header-jpeg"></div>
<div class="art-Logo">

<h1 id="name-text" class="art-Logo-name"><a href="#">GETLAPTOPS5</a></h1>

<div id="slogan-text" class="art-Logo-text"> Cheap Laptops for Sale</div>
</div>
</div>
<div class="art-nav">
<div class="l"></div>
<div class="r"></div>

<ul class="art-menu">
<li><a href="index.html" class="[ActiveItem]"><span class="l"></span><span class="r"></span><span class="t">Home</span></a></li>

<li><a href="products.html"><span class="l"></span><span class="r"></span><span class="t">Laptops Model List</span></a></li>
<li><a href="Getlaptops5_SeasonSpecials.html"><span class="l"></span><span class="r"></span><span class="t">Season Specials</span></a></li>

<li><a href="contactus.html"><span class="l"></span><span class="r"></span><span class="t">Details</span></a></li>

</ul>
</div>
<div class="art-contentLayout">
<div class="art-content">
<div class="art-Post">
<div class="art-Post-body">
<div class="art-Post-inner">

<h2 class="art-PostHeaderIcon-wrapper">

<span class="art-PostHeader">About Getlaptops5 and Laptops</span>

</h2>

<div class="art-PostContent">

<p>Economical laptops for sale, we sell laptops, buy laptops. We sell laptops at cheaper laptops prices. We have laptops for sale, we sell different laptops types. If you are looking for laptops, visit our laptops sales department. Our laptops are affordable, we offer laptops delivery service if you purchase our laptops. We also offer laptops accessories including laptops bags and laptops chargers. Our laptops bags come in many variations, we guarantee you good laptop deals including the laptops bag. The laptops sales department is there for laptops customers.</p>
<p>HP laptops for sale. Low-priced laptops for sale, laptops. We sell low-cost laptops, laptops today. Inexpensive laptops for sale, visit us if you need laptops. Discounted laptops for sale, all laptops brands, good laptops for sale. Looking for laptops, come buy laptops. Need laptops, buy laptops from us. Laptops sales, bid for laptops for sale, laptops for sale, laptops for sale, laptops for sale, laptops for sale.</p>

<p>We have 15 years of experience in second hand laptops. Laptops. Search for your laptops and request a laptop quote.</p>

```

```

<div class="art-Post">
  <div class="art-Post-body">
    <div class="art-Post-inner">
      <h2 class="art-PostHeaderIcon-wrapper">
        <span class="art-PostHeader">Laptops for sale</span>
      </h2>
      <div class="art-PostContent">
        <p>Our laptops Speciality. Below are different laptops preferences and laptops bags:</p>
        <p>
          <ul>
            <li>Asus laptops and laptops bags</li>
            <li>Mecer laptops and laptops bags</li>
            <li>Fujitsu second hand laptops and laptops bags</li>
            <li>Lenovo laptops and laptops bags</li>
            <li>Samsung second hand laptops and laptops bags</li>
            <li>Dell laptops and laptops bags</li>
            <li>Proline laptops and laptops bags</li>
            <li>Compaq laptops and laptops bags</li>
            <li>Toshiba laptops and laptops bags</li>
            <li>Acer laptops and laptops bags</li>
          </ul>
        </p>
        <p>We provide the best laptops deals in Cape Town. Getlaptops sales high performance laptops. Buy laptops now. Better laptops deals. Laptops on sale. Bid for laptops, come and buy laptops now. Laptops now, buy laptops now, laptops in stock.</p>
        <p>Asus laptops now, mini laptops on sale. Refurbished laptops for sale. Cheap laptops, laptops for sale, laptops here. Buy laptops now, Proline laptops now, laptops. Shop cheap laptops from Getlaptops5. We now offer online laptops purchasing.</p>
        <p>
          <span class="art-button-wrapper">
            <a href="contactus.html"><span class="l"></span><span class="r"></span><span class="art-button">Contact us NOW!</span></a>
          </span>
        </p>
        <p>
        </p>
      </div>
      <div class="cleared"></div><div class="art-Footer">
        <div class="art-Footer-inner">
          <a href="#" class="art-rss-tag-icon" title="RSS"></a>
          <div class="art-Footer-text">
            </div>
          </div>
          <div class="art-Footer-background"></div>
        </div>
      </div>
    </div>
  </div>
</body>
</html>

```

Products Page for Getlaptops5

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="en-US" xml:lang="en">
<head>
  <!--
  Designed by Herbert Zuze
  http://validator.w3.org : "This page is valid XHTML 1.0 Transitional"
  -->
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
  <meta http-equiv="X-UA-Compatible" content="IE=EmulateIE7" />
  <title>GETLAPTOPS5</title>

  <link rel="stylesheet" href="style.css" type="text/css" media="screen" />
  <META NAME="KEYWORDS" CONTENT="Laptops,laptop,getlaptops1,computers,notebook">
</head>
<body>
<div id="art-page-background-simple-gradient">
  </div>

  <div id="art-main">
    <div class="art-Sheet">
      <div class="art-Sheet-tl"></div>
      <div class="art-Sheet-tr"></div>
      <div class="art-Sheet-bl"></div>
      <div class="art-Sheet-br"></div>
      <div class="art-Sheet-tc"></div>
      <div class="art-Sheet-bc"></div>

      <div class="art-Sheet-cl"></div>

      <div class="art-Sheet-cr"></div>
      <div class="art-Sheet-cc"></div>
      <div class="art-Sheet-body">
        <div class="art-Header">
          <div class="art-Header-png"></div>
          <div class="art-Header-jpeg"></div>
          <div class="art-Logo">

            <h1 id="name-text" class="art-Logo-name"><a href="#">GETLAPTOPS5</a></h1>

            <div id="slogan-text" class="art-Logo-text">Cheap Laptops for sale</div>
          </div>
        </div>
      </div>
    </div>
  </div>

```

```

</div>
<div class="art-nav">
  <div class="l"></div>
  <div class="r"></div>

  <ul class="art-menu">
    <li><a href="index.html" class="{ActiveItem}"><span class="l"></span><span class="r"></span><span class="t">Home</span></a></li>
    </li>
    <li><a href="products.html"><span class="l"></span><span class="r"></span><span class="t"> Laptops Model List</span></a></li>
    <li><a href="Getlaptops5_SeasonSpecials.html"><span class="l"></span><span class="r"></span><span class="t">Season Specials</span></a></li>

    <li><a href="contactus.html"><span class="l"></span><span class="r"></span><span class="t">Details</span></a></li>

  </ul>
</div>
<div class="art-contentLayout">
  <div class="art-content">
    <div class="art-Post">
      <div class="art-Post-body">
        <div class="art-Post-inner">

          <h2 class="art-PostHeaderIcon-wrapper">

            <span class="art-PostHeader">Best Deals</span>

          </h2>

          <p>

            <ol>

              <li>Lenovo out of stock </li>
              <li>Asus </li>
              <li>Samsung stock available</li>
              <li>Dell Vestro limited stock</li>
              <li>HP Compaq stock available</li>
              <li>Acer stock available</li>
              <li>Asus out of stock</li>

            </ol>

          <p>

            <span class="art-button-wrapper">

              <a href="contactus.html"><span class="l"></span><span class="r"></span><span class="art-button">Contact us NOW!</span></a>

            </span>

          </p>

        </div>
      </div>
    </div>
  </div>
  <div class="cleared"></div>
  <div class="art-Footer">
    <div class="art-Footer-inner">
      <a href="contactus" class="art-rss-tag-icon" title="RSS"></a>

      <div class="art-Footer-text">
        </div>
      </div>
    </div>
  </div>
</div>
</body>
</html>

```

Contact Us Page for Getlaptops5

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="en-US" xml:lang="en">
<head>
  <!--
  Designed by Herbert Zuze
  http://validator.w3.org : "This page is valid XHTML 1.0 Transitional"
  -->
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
  <meta http-equiv="X-UA-Compatible" content="IE=EmulateIE7" />
  <title>GETLAPTOPS5 </title>
  <link rel="stylesheet" href="style.css" type="text/css" media="screen" />
  <META NAME="KEYWORDS" CONTENT="Laptops,laptop,getlaptops5,computers,notebook">
</head>
<body>
<div id="art-page-background-simple-gradient">
</div>
  <div id="art-main">
    <div class="art-Sheet">
      <div class="art-Sheet-tl"></div>
      <div class="art-Sheet-tr"></div>
      <div class="art-Sheet-bl"></div>
      <div class="art-Sheet-br"></div>
      <div class="art-Sheet-tc"></div>
      <div class="art-Sheet-bc"></div>
      <div class="art-Sheet-cl"></div>
      <div class="art-Sheet-cr"></div>
      <div class="art-Sheet-cc"></div>
      <div class="art-Sheet-body">
        <div class="art-Header">
          <div class="art-Header-png"></div>
          <div class="art-Header-jpeg"></div>
          <div class="art-Logo">
            <h1 id="name-text" class="art-Logo-name"><a href="#">GETLAPTOPS5</a></h1>

            <div id="slogan-text" class="art-Logo-text">Cheap Laptops for sale</div>
          </div>
        </div>
      </div>
    </div>
  </div>
  <div class="art-nav">
    <div class="l"></div>

```

```

<div class="r"></div>
<ul class="art-menu">
  <li><a href="index.html" class="{ActiveItem}"><span class="l"></span><span class="r"></span><span class="t">Home</span></a></li>
  </li>
  <li><a href="products.html"><span class="l"></span><span class="r"></span><span class="t">Laptops Model List </span></a></li>
  <li><a href="Getlaptops_SeasonSpecials.html"><span class="l"></span><span class="r"></span><span class="t">Season Specials</span></a></li>
  <li><a href="contactus.html"><span class="l"></span><span class="r"></span><span class="t">Details</span></a></li>
</ul>
</div>
<div class="art-contentLayout">
  <div class="art-content">
    <div class="art-Post">
      <div class="art-Post-body">
        <div class="art-Post-inner">
          <h2 class="art-PostHeaderIcon-wrapper">
            <span class="art-PostHeader">Contact details</span>
          </h2>
          <p>
            <p> Getlaptops5</p>
            <p>P.O Box 4509</p>
            <p>Andrews street</p>
            <p>Pretoria 0182</p>
            <p>Tel: 0861 190 082</p>
            <p>Fax: 0861 190 082</p>
            <p>Email: info@getlaptops5.co.za</p>
            <p><b>Business Hours</b></p>
            <p>Mon-Sat: 8am - 4pm</p>
          </p>
        </div>
      </div>
    </div>
    <div class="art-Footer">
      <div class="art-Footer-inner">
        <a href="#" class="art-rss-tag-icon" title="RSS"></a>
      </div>
    </div>
    <div class="art-Footer-background"></div>
  </div>
</div>
</div>
</body>
</html>

```



```

<h1 id="name-text" class="art-Logo-name"><a href="#">GETLAPTOPS5</a></h1>
<div id="slogan-text" class="art-Logo-text">Cheap Laptops for sale</div>
</div>
<div class="art-nav">
<div class="l"></div>
<div class="r"></div>
<ul class="art-menu">
<li><a href="index.html" class="{ActiveItem}"><span class="l"></span><span class="r"></span><span class="t">Home</span></a></li>
</li>
<li><a href="products.html"><span class="l"></span><span class="r"></span><span class="t"> Laptops Model List</span></a></li>
<li><a href="Getlaptops5_SeasonSpecials.html"><span class="l"></span><span class="r"></span><span class="t"> Season Specials</span></a></li>
<li><a href="contactus.html"><span class="l"></span><span class="r"></span><span class="t">Details</span></a></li>
</ul>
</div>
<div class="art-contentLayout">
<div class="art-content">
<div class="art-Post">
<div class="art-Post-body">
<div class="art-Post-inner">
<h2 class="art-PostHeaderIcon-wrapper">
<span class="art-PostHeader">Best Deals</span>
</h2>
<p>
<ol>
<li>Lenovo out of stock </li>
<li>Asus </li>
<li>Samsung stock available</li>
<li>Dell Vestro limited stock</li>
<li>HP Compaq stock available</li>
<li>Acer stock available</li>
<li>Asus out of stock</li>
</ol>
<p>
<span class="art-button-wrapper">
<a href="contactus.html"><span class="l"></span><span class="r"></span><span class="art-button">Contact us NOW!</span></a>
</span>
</p>
<div class="cleared"></div><div class="art-Footer">
<div class="art-Footer-inner">
<a href="contactus" class="art-rss-tag-icon" title="RSS"></a>
<div class="art-Footer-text">
</div>
</div>
<div class="art-Footer-background"></div>
</div>
</div>
</div>
</body>
</html>

```


<td></td>
<td>COMPAQ PRESARIO G20 NETBOOK</td>
<td>R10 680</td>
<td>16</td>

</tr>

<tr>

<td>DELL</td>
<td>VETRO 5200s NOTEBOOK</td>
<td>R9 600</td>
<td>66</td>

</tr>

<tr>

<td></td>
<td>VETRO 6600s NOTEBOOK</td>
<td>R9 500</td>
<td>15</td>

</tr>

<tr>

<td>ACER</td>
<td>ASPIRE BOP990T NETBOOK</td>
<td>R9 000</td>
<td>115</td>

</tr>

<tr>

<td></td>
<td>ENVY 8824-289LP</td>
<td>R7 000</td>
<td>21</td>

</tr>

<tr>

<td>SONY</td>
<td>AVIO VPCFG2560TX/V NETBOOK</td>
<td>R8 000</td>
<td>5</td>

</tr>

<tr>

<td></td>
<td>AVIO VYEDSEA2562HY/P</td>
<td>3 696</td>
<td>12</td>

</tr>

<tr>

<td>FUJITSU</td>
<td>NOTEBOOK MJY9630</td>
<td>R5 950</td>
<td>20</td>

</tr>

<tr>
<td></td>
<td>NETBOOK NHF3670</td>
<td>R6 990</td>
<td>4</td>
</tr>

</tbody></table>

</p>

<p></p>

<p>OTHER PRODUCTS</p>

<p>

- Notebook chargers
- External hard drives
- Pc memory and
- Wireless keyboards and mouse

</p>

```

<p></p>

<div class="cleared"></div><div class="art-Footer">
  <div class="art-Footer-inner">
    <a href="#" class="art-rss-tag-icon" title="RSS"></a>
    <div class="art-Footer-text">
      </div>
    </div>
    <div class="art-Footer-background"></div>
  </div>
</div>
</body>
</html>

```

Contact Us Page for Getlaptops5

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="en-US" xml:lang="en">
<head>
  <!--
  Designed by Herbert Zuze
  http://validator.w3.org : "This page is valid XHTML 1.0 Transitional"
  -->
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
  <meta http-equiv="X-UA-Compatible" content="IE=EmulateIE7" />
  <title>GETLAPTOPS5 </title>
  <link rel="stylesheet" href="style.css" type="text/css" media="screen" />
  <META NAME="KEYWORDS" CONTENT="Laptops,laptop,getlaptops5,computers,notebook">
</head>
<body>
<div id="art-page-background-simple-gradient">
  </div>
  <div id="art-main">
    <div class="art-Sheet">
      <div class="art-Sheet-tl"></div>
      <div class="art-Sheet-tr"></div>
      <div class="art-Sheet-bl"></div>
      <div class="art-Sheet-br"></div>
      <div class="art-Sheet-tc"></div>
      <div class="art-Sheet-bc"></div>
      <div class="art-Sheet-cl"></div>
      <div class="art-Sheet-cr"></div>
      <div class="art-Sheet-cc"></div>
      <div class="art-Sheet-body">
        <div class="art-Header">
          <div class="art-Header-png"></div>
          <div class="art-Header-jpeg"></div>
          <div class="art-Logo">
            <h1 id="name-text" class="art-Logo-name"><a href="#">GETLAPTOPS5</a></h1>
            <div id="slogan-text" class="art-Logo-text">Cheap Laptops for sale</div>
          </div>
          <div class="art-nav">
            <div class="l"></div>
            <div class="r"></div>
            <ul class="art-menu">
              <li><a href="index.html" class="{ActiveItem}"><span class="l"></span><span class="r"></span><span class="t">Home</span></a></li>
              <li><a href="products.html"><span class="l"></span><span class="r"></span><span class="t">Laptops Model List </span></a></li>
              <li><a href="Getlaptops5_SeasonSpecials.html"><span class="l"></span><span class="r"></span><span class="t">Season Specials</span></a></li>
              <li><a href="contactus.html"><span class="l"></span><span class="r"></span><span class="t">Details</span></a></li>
            </ul>
          </div>
          <div class="art-contentLayout">
            <div class="art-content">
              <div class="art-Post">
                <div class="art-Post-body">
                  <div class="art-Post-inner">
                    <h2 class="art-PostHeaderIcon-wrapper">
                      <span class="art-PostHeader">Contact details</span>
                    </h2>
                    <p>
                      <p> Getlaptops5</p>
                      <p>P.O Box 4509</p>
                      <p>Andrews street</p>
                      <p>Pretoria 0182</p>
                      <p>Tel: 0861 190 082</p>
                      <p>Fax: 0861 190 082</p>
                      <p>Email: info@getlaptops5.co.za</p>
                      <p><b>Business Hours</b></p>
                      <p>Mon-Sat: 8am - 4pm</p>
                    </p>
                    <div class="cleared"></div><div class="art-Footer">
                      <div class="art-Footer-inner">
                        <a href="#" class="art-rss-tag-icon" title="RSS"></a>
                        <div class="art-Footer-text">
                          </div>
                        </div>
                        <div class="art-Footer-background"></div>
                      </div>
                    </div>
                  </div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</body>
</html>

```

APPENDIX I

PHASE 1 INDEXING RESULTS RECORDINGS

Webpage Indexing records(String search) :

G oogle

B Ing

Y aho

gipe1 Getlaptops1

"We have specially trained technicians who ensure that our laptops, when dispatched, are fully functional."

gipe2 Getlaptops2

"If you wish to sell or trade in your laptops, do not wait - bring it to our offices and we give you the actual value of your computer."

gipe3 Getlaptops3

"We are an accredited African leading laptops dealer with traceable trading records and our pride comes from you, our cherished clients"

gipe4 Getlaptops4

"We also sell laptops accessories including laptops bags and laptops chargers."

gipe5 Getlaptops5

"Economical laptops for sale, we sell laptops, buy laptops"

Daily check results

BLUE - Clobbered

RED - Not indexed Yet

GREEN - Indexed by this date

Date	G	Y	B
07-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
08-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
09-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
10-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
11-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
12-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
13-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
14-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
15-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
16-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
17-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
18-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
19-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
20-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
21-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
22-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
23-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
24-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
25-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
26-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
27-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
28-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
29-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
30-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
31-10-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
01-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
02-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
03-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
04-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
05-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
06-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
07-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
08-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5
09-11-10	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5	gipe1 gipe2 gipe3 gipe4 gipe5

APPENDIX J PHASE 2 INDEXING RESULTS RECORDINGS

Submitted on 13-12-10

Webpage Indexing records(String search):

G oogle
B ing
Y aahoo

- glps1** Getlaptops1
"Our laptops are affordable, we offer laptops delivery service if you purchase our laptops."
- glps2** Getlaptops3
"High-quality laptops, super laptops, buy laptops."
- glps3** Getlaptops3
"Aspire laptops, Mesh-black laptops, Timeliner laptops, Ad5742 laptops."
- glps4** Getlaptops5
"Grey laptops, laptops super laptops, laptops."
- glps5** Getlaptops5
"Laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops laptops"

Daily check results

BLUE= Cooked
RED = Not Indexed Yet
GREEN = Indexed by this date

Date	glps1	glps2	glps3	glps4	glps5
13-12-10	RED	RED	RED	RED	RED
14-12-10	RED	RED	RED	RED	RED
15-12-10	RED	RED	RED	RED	RED
16-12-10	RED	RED	RED	RED	RED
17-12-10	RED	RED	RED	RED	RED
18-12-10	RED	RED	RED	RED	RED
19-12-10	RED	RED	RED	RED	RED
20-12-10	RED	RED	RED	RED	RED
21-12-10	RED	RED	RED	RED	RED
22-12-10	RED	RED	RED	RED	RED
23-12-10	RED	RED	RED	RED	RED
24-12-10	RED	RED	RED	RED	RED
25-12-10	RED	RED	RED	RED	RED
26-12-10	RED	RED	RED	RED	RED
27-12-10	RED	RED	RED	RED	RED
28-12-10	RED	RED	RED	RED	RED
29-12-10	RED	RED	RED	RED	RED
30-12-10	RED	RED	RED	RED	RED
31-12-10	RED	RED	RED	RED	RED
01-01-11	RED	RED	RED	RED	RED
02-01-11	RED	RED	RED	RED	RED
03-01-11	RED	RED	RED	RED	RED
04-01-11	RED	RED	RED	RED	RED
05-01-11	RED	RED	RED	RED	RED
06-01-11	RED	RED	RED	RED	RED
07-01-11	RED	RED	RED	RED	RED
08-01-11	RED	RED	RED	RED	RED
09-01-11	RED	RED	RED	RED	RED
10-01-11	RED	RED	RED	RED	RED
11-01-11	RED	RED	RED	RED	RED
12-01-11	RED	RED	RED	RED	RED

APPENDIX K WEBMASTER TOOLS SNAPSHOTS

Google webmaster tools

Home
All Messages
You have no unread messages.

Forward to herbertsize@gmail.com | Language: English

Help with:
[Adding a site](#)
[Reporting spam](#)
[Getting help](#)
[Site reconsideration](#)
[Google TWT](#)
[Message formatting](#)

Sites

Sites	Verification
www.getlaptops1.co.za	Add or remove owners
www.getlaptops2.co.za	Add or remove owners
www.getlaptops3.co.za	Add or remove owners
www.getlaptops4.co.za	Add or remove owners
www.getlaptops5.co.za	Add or remove owners

© 2011 Google Inc. - [Webmaster Central](#) - [Terms of Service](#) - [Privacy Policy](#) - [Webmaster Tools Help](#)

YAHOO! SITE EXPLORER

Explore URL

Site Explorer

My Sites

http:// Add My Site

New Pending Authenticated Failed

Site	Feeds	Actions	Status
http://www.getlaptops1.co.za	0	0 Active	✓
http://www.getlaptops2.co.za	0	0 Active	✓
http://www.getlaptops3.co.za	0	0 Active	✓
http://www.getlaptops4.co.za	0	0 Active	✓
http://www.getlaptops5.co.za	0	0 Active	✓

Drive more traffic to your site. Get a \$25 credit on Yahoo! Search Marketing

Get more out of Site Explorer. Duplicate site content? Use a Delete URL Action.

Copyright © 2010 Yahoo! Inc. All rights reserved. Privacy Policy - Terms of Service - Copyright/IP Policy - Submit Your Site - Help

bing Webmaster

Home

RESOURCES: [Blogs](#), [Forums](#), [Support](#)

MESSAGES

View messages and updates about issues affecting your sites

Message	Site	Date
Welcome to the new Bing Webmaster Tools		10/10/2010 11:58 PM

SITES

Select a site to see how Bing crawls and indexes your pages. Newly added sites will begin showing data within 3 days.

Add Site Remove Site Export

Site	Clicks (%)	Impressions (%)	Pages Indexed (%)	Pages Crawled (%)
getlaptops1.co.za	0 (0%)	0 (0%)	1 (0%)	0 (0%)
getlaptops2.co.za	0 (0%)	0 (0%)	Not available	Not available
getlaptops3.co.za	0 (0%)	0 (0%)	1 (0%)	0 (0%)

SITES

Select a site to see how Bing crawls and indexes your pages. Newly added sites will begin showing data within 3 days.

Add Site Remove Site Export

Site	Clicks (%)	Impressions (%)	Pages Indexed (%)	Pages Crawled (%)
getlaptops3.co.za	0 (0%)	0 (0%)	1 (0%)	0 (0%)
getlaptops4.co.za	0 (0%)	0 (0%)	1 (0%)	0 (0%)
getlaptops5.co.za	0 (0%)	0 (0%)	1 (0%)	0 (0%)

APPENDIX L

TRANSFORMED SURVIVAL DATA

	<i>Phase</i>	<i>SE</i>	<i>KWC</i>	<i>Group</i>	<i>Time to being indexed</i>	<i>MaxTime</i>	<i>Status</i>
1	Phase 1	Google	13 keywords	1	67	67	Censored
2	Phase 1	Google	20 keywords	1	27	67	Event
3	Phase 1	Google	28 keywords	1	28	67	Event
4	Phase 1	Google	40 keywords	1	32	67	Event
5	Phase 1	Google	90 keywords	2	32	67	Event
6	Phase 1	Bing	13 keywords	1	9	67	Event
7	Phase 1	Bing	20 keywords	1	8	67	Event
8	Phase 1	Bing	28 keywords	1	9	67	Event
9	Phase 1	Bing	40 keywords	1	9	67	Event
10	Phase 1	Bing	90 keywords	2	4	67	Event
11	Phase 1	Yahoo	13 keywords	1	9	67	Event
12	Phase 1	Yahoo	20 keywords	1	8	67	Event
13	Phase 1	Yahoo	28 keywords	1	9	67	Event
14	Phase 1	Yahoo	40 keywords	1	9	67	Event
15	Phase 1	Yahoo	90 keywords	2	4	67	Event
16	Phase 2	Google	100 keywords	2	67	67	Censored
17	Phase 2	Google	132 keywords	2	10	67	Event
18	Phase 2	Google	170 keywords	2	67	67	Censored
19	Phase 2	Google	232 keywords	2	67	67	Censored
20	Phase 2	Google	321 keywords	2	67	67	Censored
21	Phase 2	Bing	100 keywords	2	23	67	Event
22	Phase 2	Bing	132 keywords	2	22	67	Event
23	Phase 2	Bing	170 keywords	2	22	67	Event
24	Phase 2	Bing	232 keywords	2	19	67	Event
25	Phase 2	Bing	321 keywords	2	15	67	Event
26	Phase 2	Yahoo	100 keywords	2	23	67	Event

27	Phase 2	Yahoo	132 keywords	2	22	67	Event
28	Phase 2	Yahoo	170 keywords	2	22	67	Event
29	Phase 2	Yahoo	232 keywords	2	30	67	Event
30	Phase 2	Yahoo	321 keywords	2	15	67	Event

APPENDIX M

SURVIVAL TABLE

SE	ID	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases	
				Estimate	Std. Error			
Google	1	132 keywords	10.000	Event	.900	.095	1	9
	2	20 keywords	27.000	Event	.800	.126	2	8
	3	28 keywords	28.000	Event	.700	.145	3	7
	4	40 keywords	32.000	Event	.	.	4	6
	5	90 keywords	32.000	Event	.500	.158	5	5
	6	100 keywords	67.000	Censored	.	.	5	4
	7	170 keywords	67.000	Censored	.	.	5	3
	8	232 keywords	67.000	Censored	.	.	5	2
	9	321 keywords	67.000	Censored	.	.	5	1
	10	13 keywords	67.000	Censored	.	.	5	0
Bing	1	90 keywords	4.000	Event	.900	.095	1	9
	2	20 keywords	8.000	Event	.800	.126	2	8
	3	13 keywords	9.000	Event	.	.	3	7
	4	28 keywords	9.000	Event	.	.	4	6
	5	40 keywords	9.000	Event	.500	.158	5	5

	6	321 keywords	15.000	Event	.400	.155	6	4
	7	232 keywords	19.000	Event	.300	.145	7	3
	8	132 keywords	22.000	Event	.	.	8	2
	9	170 keywords	22.000	Event	.100	.095	9	1
	10	100 keywords	23.000	Event	.000	.000	10	0
Yahoo	1	90 keywords	4.000	Event	.900	.095	1	9
	2	20 keywords	8.000	Event	.800	.126	2	8
	3	13 keywords	9.000	Event	.	.	3	7
	4	28 keywords	9.000	Event	.	.	4	6
	5	40 keywords	9.000	Event	.500	.158	5	5
	6	321 keywords	15.000	Event	.400	.155	6	4
	7	132 keywords	22.000	Event	.	.	7	3
	8	170 keywords	22.000	Event	.200	.126	8	2
	9	100 keywords	23.000	Event	.100	.095	9	1
	10	232 keywords	30.000	Event	.000	.000	10	0

APPENDIX N

KEYWORD SURVIVAL TABLE

KeywordGroup1	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases	
			Estimate	Std. Error			
40 or fewer keywords	1	8.000	Event	.	.	1	11
	2	8.000	Event	.833	.108	2	10
	3	9.000	Event	.	.	3	9
	4	9.000	Event	.	.	4	8
	5	9.000	Event	.	.	5	7
	6	9.000	Event	.	.	6	6
	7	9.000	Event	.	.	7	5
	8	9.000	Event	.333	.136	8	4
	9	27.000	Event	.250	.125	9	3
	10	28.000	Event	.167	.108	10	2
	11	32.000	Event	.083	.080	11	1
	12	67.000	Censored	.	.	11	0
More than 40 keywords	1	4.000	Event	.	.	1	17
	2	4.000	Event	.889	.074	2	16
	3	10.000	Event	.833	.088	3	15
	4	15.000	Event	.	.	4	14
	5	15.000	Event	.722	.106	5	13
	6	19.000	Event	.667	.111	6	12
	7	22.000	Event	.	.	7	11
	8	22.000	Event	.	.	8	10
	9	22.000	Event	.	.	9	9
	10	22.000	Event	.444	.117	10	8
	11	23.000	Event	.	.	11	7

12	23.000	Event	.333	.111	12	6
13	30.000	Event	.278	.106	13	5
14	32.000	Event	.222	.098	14	4
15	67.000	Censored	.	.	14	3
16	67.000	Censored	.	.	14	2
17	67.000	Censored	.	.	14	1
18	67.000	Censored	.	.	14	0

GLOSSARY

Terms/Acronyms/Abbreviations

Algorithm – In relation to this study, it is a well-defined step-by-step problem solving procedure used by search engines to determine the ranking of results in relation to the end-user search query.

Blacklisted – This refers to a website identified as generating spamdexing and is subsequently removed from the search engine index.

Body Text – Is the textual content of the website or webpage that is displayed to a user through the user's browser.

Cloaking – This is a spamdexing technique where content or URLs presented to search robots are different to the one presented to human visitors.

Crawlers – A complex computer program used by a search engine to systematically scan and analyse the text content of webpages in order to determine their importance, and to store the information in an index that can be used for ranking in the results of Internet searches for certain keywords.

Domain – An Internet site, unique descriptor or address.

Epistemology – The theory of knowledge concerning sources and scope of knowledge and how notions relate to each other and further deals with ambiguity in knowledge claims.

Index – A database or directory of links and webpages used by a search engine to locate information and match user search queries.

Internet – The interconnection of computers on the globe that can provide instant and infinite simultaneous connection.

Keywords – Focal, descriptive, natural language words that are utilised by Internet users when searching for information on search results.

Keyword Density – The frequency of a keyword or phrase as used in a webpage.

Keyword stuffing – Is an unethical search engine optimisation technique of overusing a keyword or phrase in a webpage to enhance ranking on search results.

Link farm – An array of webpages with a large number of hyperlinks exchanged to each other or other pages with the main motive of promoting one another, boosting link popularity and ultimately enhancing search engine rankings.

Metatag – Is a coding instruction in HTML (HyperText Markup Language) that describes some feature of the contents of a webpage to search engine and instructs it what to do with the content.

Ontology – Is a formal, explicit specification of knowledge as a set of concepts within a category, and the relationships between those concepts, the objects, and other entities that are assumed to be in some area of interest and the relationships that can exist.

PPC – Pay-Per-Click is an Internet advertising technique used on search engines by bidding for keywords, in order to obtain a higher ranking on the search results of a search engine.

Ranking – The position in terms of popularity given to a webpage by a search engine on its search result page. On the results page the webpage displayed first on the listing is considered better than the second one.

Search engine – A service that allows an Internet user to enter a keyword or phrase to search for information. The search engine displays results on its SERP.

SEM – Short for “Search Engine Marketing” or in short “Search Marketing”. It is the overall process of increasing website exposure and traffic through search engine optimisation (SEO) efforts or paid search advertising (PPC).

SEO – Search engine optimisation is the use of various techniques when designing webpages, in order to attain high website rankings and visibility in the search results of a search engine.

SERP – Short for “Search Engine Results Page”, the webpage displayed by a search engine in response to a user search query for a specific keyword or key phrase. The results page contains listings of items such as webpages, video, maps, books, images, etc. The page will have three components, namely the sponsored links, search result area (natural listing) and the assistive information area. The result returned from the search query comprises of the title, description information of the subject title and the Uniform Resource Locator (URL) link.

Sitemap – A site map is a comprehensive visual or textual structure of a website's content that informs search engines about pages on their sites that are available for crawling. It can be a plain text file but it is usually designed in XML and can include each URL's changes, as well as the last time it was updated.

Spam – Also referred to as unsolicited e-mail; this is the dissemination of unwanted e-mail for the sake of commercial reasons.

Spamdexing – Also known as search spam or search engine spam, it is a technique of fooling a search engine's indexing algorithm with the intention of increasing the webpage ranking in the search results of a search engine.

Traffic – The amount of distinctive human visitors conducting a search on a webpage; the higher the number of visitors the higher the conversion probability.

Triangulation – The application of various research methods in order to have exceptionally high factual analysis, conclusion and recommendation and further achieve a high level of reliable and valid outcome.

Usability – In the context of this study, usability pertains to the ability of an end-user to perform various applicable exercises on a webpage without experiencing forms of sophistications and frustrations but enhancement of fun, logic and enjoyable navigations.

Website visibility – The practice of using different techniques to optimise the content of a website in order to earn a high search engine ranking on its results page.

User – Generally refers to a human (regardless of the level of Internet knowledge) who uses the Internet for different reasons (e.g. research, buying and selling, games, chats and other commercial and non-commercial services).

Webpage – A HTML document that forms part of an entire website, mostly with content that describes the page. A webpage may contain a mixture of text, links, graphics, videos and chat rooms.

Website – Is a collection of related webpages or files and documents hosted on a Web server and accessed on the World Wide Web of the Internet.