

Search engine visibility: the effect of generic top-level domain choice.

Z. Hamdulay
e-Innovation Academy, Cape Peninsula University of Technology
Cape Town
South Africa
zhamdulay@mweb.co.za

M. Weideman
e-Innovation Academy, Cape Peninsula University of Technology
Cape Town
South Africa
melius@gmail.com

Abstract

The principle objective of this research project is to determine whether or not search engines rank websites in dot-com domains higher than alternative generic top-level domains (gTLDs). The research question is: Should websites in these other gTLDs have to increase their marketing exposure of their websites?

The methods to be employed in this research project are to build separate websites each having domain names in different gTLDs, e.g. mysite.com, mysite.net, mysite.biz etc. The content on all websites will be identical and will be submitted to major local and international search engines. Each website will be search engine crawler-friendly and will include title tags in hyperlinks, alt tags in graphics, definitive webpage titles and keyword phrases in the meta tags etc.

These websites will not only be crawler friendly but also user friendly. All sites will be written in HyperText Markup Language (HTML). The domain names chosen will be descriptive and no scripting languages will be used e.g. Javascript. Each website will be thoroughly tested in known browsers e.g. Netscape and Microsoft Internet Explorer. Data will also be gathered via interviews with senior officials at local e-commerce traders and/or search engines. This interview will mainly consist of open-ended questions regarding their marketing infrastructure towards successfully branding their websites to Internet users.

Expected results from the methods employed above could prove that websites in dot-com domains are ranked higher than alternative gTLDs by search engine crawlers. Further outcomes could also prove that websites in alternative gTLDs have to spend more money when branding their websites to Internet users.

This study is intended to make website owners aware of the importance of choosing the appropriate URL to increase website visibility on the Internet.

Keywords

generic top-level domain, e-business, website visibility, search engine

1. Introduction

The growth of domain name registrations continues to increase. The total number of registered domain names for the second quarter of 2004 was 66.3 million (Verisign, 2004). Dot-com remains to be the largest gTLD, which accounts for 47% of the total number of registered domain names. Alternate gTLDs such as ".biz", ".info", ".name", ".net" and ".org" collectively add up to 17%, with an average of 90.000 additional registrations per month (Verisign, 2004).

Thurow (2003) stated that over 300 million searches are performed daily on search engines and directories. This author further mentions that proper preparation of a website for search engine visibility increases the chances of Internet users finding the website.

Brighty & Markham (2002) mention that search engines place more importance on gTLDs and strongly recommends the addition of keywords in domain names. However, no empirical experiments have been found in order to determine the impact a gTLD has on website visibility.

The research problem is that owners of e-Commerce websites have no clarity as to the effect the use of alternate domain names have on the visibility of their mission critical websites.

2. Background to the research problem

Research has shown that almost all English words have been registered as part of domain names in the dot-com domain, which is evidence of "exhaustion of the namespace", and therefore supporting the need for new gTLDs (Gritter, 2002). With the introduction of these gTLDs, an opportunity is given to e-commerce website owners who could not register their "preferred domain name" in the .com domain due to trademark conflicts, to now purchase their names in the new gTLDs. E-commerce website owners can now purchase more meaningful names to increase their website visibility e.g. autoweb. Also, content can now be segregated into more definable groups enabling domain-specific search engines to index websites more logically and therefore increase customer traffic.

With all these advantages, Berners-Lee (2004) argues that these gTLDs could cause more harm than good. This could initiate confusion to customers where the same brand name is used on separate gTLDs for different e-business websites. The cost for major e-commerce websites to ensure that they own their domain branding across all gTLDs could be lower than for smaller e-businesses. This could also introduce a flurry of cybersquatters occupying the domain space in the new gTLDs (Melville, 2000) and introduce a number of legal disputes (Apke, 2003).

The addition of gTLDs may not affect trademarks e.g. Microsoft Corporation, purchasing domain names such as microsoft.net or microsoft.org, but may be more problematic for generic domain names as it has no trademark protection. Gritter (2002) mentions that adding new gTLDs will not increase the availability of more descriptive names.

Internet Corporation for Assigned Names and Numbers (ICANN) does provide "sunrise periods", which are exclusively open to trademark owners to purchase their domain name. These "sunrise periods" are therefore only available to trademark owners

(OECD, 2004). Thus, big corporations will simply reserve all the different variations of the name, i.e. business.tel, business.mail, business.mobi, etc.

ICANN is a non-profit organisation charged with the management of the domain name system (DNS). It is also responsible for the resolution of cybersquatting issues and the creation of new TLDs (Weideman & Moncrief, 2000).

How does this affect the visibility of the website? The author performed a simple search on Google and Yahoo! using the keyword "used cars", which returned 22 900 000 and 22 400 000 results respectively. The sites, usedcars.biz and usedcars.com are separately owned websites that sells used cars. The results returned can be seen in Table 2.1. Do search engines rank ".com" domains higher than any other domains? Why has the ".biz" domain been listed 550th on Yahoo and not listed at all on Google?

Table 1: Search engine hits for similar domains.

Site	Ranked on Google	Ranked on Yahoo!	Keyword
usedcars.biz	Not listed	550	used cars
usedcars.com	4	2	used cars

Two of the reasons for ICANN to allow additional gTLDs is to increase business on the Internet as well to give entrepreneurs an opportunity to purchase memorial domain names in gTLDs other than dot-com. There is strong evidence that the dot-com space is overcrowded.

3. Literature study

The growth of the Internet in terms of number of websites has convinced Internet companies that the responsibility of luring potential customers to one's site, and maintaining customer interest and loyalty, is a daunting task. Many traditional brick-and-mortar companies hastily expanded their businesses to the Internet with the ambition to create websites to serve as storefronts, offices, and headquarters and distribution centers (Nguyen, 2001).

For many of these companies, starting an online business having the right domain name was a first step towards branding. Companies would often spend a substantial amount of money for the right domain name in the hope that it would reduce future costs for online advertising and branding campaigns to increase name recognition. Other companies secured a domain name first, then developed the business and traded around that name (Nguyen, 2001).

Therefore, owning the right "memorial" domain name is having the description of the product or online service being sold in the generic domain name. As an example, sex.com, was sold for \$250 million and business.com for \$7.5 million. It appears as if the more generic the domain name the more costly the virtual real estate becomes.

Nguyen (2001) stated that in the nineties, millions of successful or famous generic domain names gave rise to "e" and "i" prefix domain names. The "e" prefix indicates "electronic", and the "i" prefix is for "Internet". This introduced domain names such as

eToys.com and iWireless.com. Domain names mirroring a corporate name may be more of a valuable asset, as it facilitates communications with reaching customers.

With the introduction of additional gTLDs, domain name branding can become even more confusing to consumers. Does an Internet website owner now have to purchase their domain name in all gTLDs? For example, does business.com have to acquire business.tel, business.mail or business.mobi, etc to protect its trademark and to avoid confusion or misappropriation (Berners-Lee, 2004)?

The purchase of domain names in all gTLDs might be affordable for large companies, but for smaller companies the cost becomes very significant (Schmitz & Sint 2003). For most businesses, the part between www and the domain name is its trademark. Will customers now have to remember which gTLD to distinguish between company brands?

3.1 Information on the Internet

Today, the Internet is the youngest and fastest growing medium, and its accelerating growth suggests that it has not yet reached its highest expansion period (Anon., 2000).

This author further concludes that:

- the static Web consists of 2.5 billion documents with a growth rate of 7.3 million pages per day,
- the amount of information on the static Web totals somewhere from 25 to 50 terabytes and
- the information on the dynamic Web is 400 to 500 times larger than that of the static Web.

Static webpages are available to spiders or users that visit a website. These are pages that have been created manually and the content can be updated manually (Green, 2000; Stojanovski, 2001). These pages are indexable to all spiders whereas the dynamic Web is not indexable to spiders. All content on the dynamic Web is supplied by a database; these include pages which require authentication.

Lawrence and Giles (1999) further add:

- eighty-three per cent of information on sites is commercial content, 6 per cent is scientific or educational content, while 1.5% of sites contain pornographic content and
- no single search engine indexes more than 16% of the estimated size of the publicly indexable Web.

The Web has become very rich with information in almost every field; however, it has become increasingly difficult to search for desired information as users are facing information overload. Chau et al. (2002) state that because a user cannot specify a search domain, a search on a general-purpose search engines such as Google results in thousands of hits.

3.2 Internet domains

A domain name ensures computers are located easily on the Internet and enables users to reach websites. The DNS matches the IP address, which identifies individual host computers on the Internet. Domain names are easier for people to remember than IP addresses. Postel and Paul Mockapetris developed the DNS system and the name space is divided into TLDs (Weinberg, 2002).

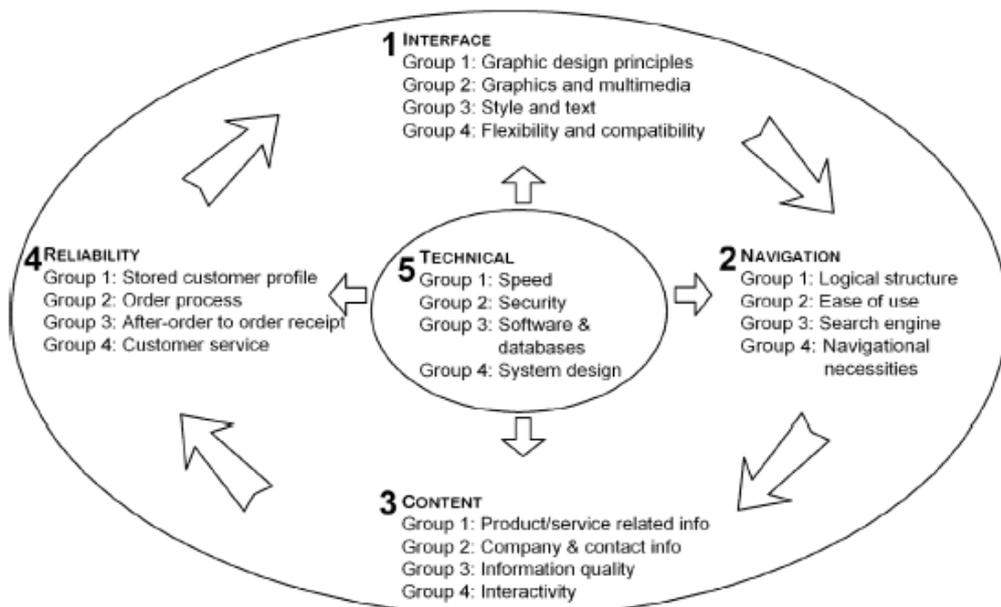
3.3 An evaluation framework

Bekker & Van der Merwe (2003) developed a framework for e-evaluating e-commerce websites. Figure 1 represents this framework. It contains 5 levels; each containing some criteria groups:

- Level 1 – Interface.
- Level 2 – Navigation.
- Level 3 – Content.
- Level 4 – Reliability.
- Level 5 – Technical.

The purpose of this framework and criteria groups is to find a comprehensive way to evaluate e-commerce websites. In this research paper, the focus will be on level 2, criteria group 3. This criteria group evaluates or measures webpages within a website in order for them to be easily located by search engines.

Figure 1 : e-Commerce website evaluation framework [Source: Bekker & Van der Merwe (2003)].



The authors used this model for the evaluation of six websites, see Figure 2. In the Navigation column (level 2), kalahari.net, pcbooks.co.za and vsonline.co.za received the lowest scores. The dot-com domains, amazon.com/books, bn.com and exclusivebooks.com all received high scores in at that level.

Table 2 : e-Commerce website evaluation framework (Matrix for the books industry) [Source: Bekker & Van der Merwe (2003)].

Site name	Site address	Interface	Navigation	Content	Reliability	Technical	Total
Amazon.com Books	www.amazon.com/books	17	35	30	31	28	141
Barnes & Noble	www.bn.com	15	29	25	27	19	115
Exclusive Books	www.exclusivebooks.com	14	35	18	27	23	117
Kalahari.net	www.kalahari.net	-15	11	13	24	12	45
PC Books, Inc.	www.pcbooks.co.za	-19	-4	1	16	22	16
VS Online	www.vsonline.co.za	6	14	-9	6	11	28

3.4 Trademarks

A business registering a domain name wants one that is unique and easy to remember for consumers, which is also not easily confused with other names. Domain names represent the "user friendly" addresses of websites on the Internet which is easily identifiable to users (Amir et al., 2003). The term trademark suggests that it is a mark used with reference to a particular trade (Panwar-Mridul, 2003).

Partridge (1997) classifies trademarks into four categories:

- generic,
- descriptive,
- suggestive and
- arbitrary and fanciful.

Generic domain names have no trademark protection, whereas descriptive terms may become protectable if the plaintiff can establish that the term has acquired "secondary meaning" in the relevant market. Suggestive, arbitrary and fanciful trademarks are afforded protection without an indication of secondary meaning.

3.5 Internet Corporation for Assigned Names and Numbers (ICANN)

They formed to develop and manage Internet policy and logistics related to:

- Internet protocols,
- IP addresses and
- domain names, (Weinberg, 2002 & Nilsen, 2002).

The ICANN board of directors added the ".museum" TLD to the Internet domain name system (DNS) for restricted use only in November 2000. Schuler & Kurtz (2002) mention some of the advantages of this TLD registration:

- it enables museum institutions to separate themselves from the broad ".org", ".com", ".net" and country code sites,
- the ".museum" designation can only be identified with museums,
- it can provide smaller institutions with a brand name that characterises their offerings to site visitors and
- Internet users can be assured a high degree of confidence that all content found at ".museum" sites is reliable since it only recognises legitimate registered museums.

3.6 Uniform Dispute Resolution Policy (UDRP)

The ICANN Board has adopted the UDRP in August 1999 for defining how disputes over domain name registrations are to be resolved. The UDRP is responsible for abusive and bad faith domain name registrations in all global top-level domains, i.e., cybersquatting. Any trademark owner can file a UDRP complaint internationally.

In order to have the domain name cancelled, the trademark holder must establish:

- that he/she has a legally recognized trademark in a name that is identical or confusingly similar to the domain name,
- that the current registrant of the domain name has no legitimate rights in the name and
- that there has been some evidence of bad faith or abuse.

3.7 Internet search engines

3.7.1 Search engine sizes

Google is the most famous search engine on the Web (Chau et al., 2002; George, 2004), with 250 million searches performed on it each day (Sullivan, 2005). A larger search index size does not imply better search results. However, using a search engine with a large index could help when searching for unusual or hard-to-find information.

3.7.2 Search engine directories and domain-specific search engines

A Web directory contains a list of webpages compiled by human editors. This list of pre-defined webpages is categorised according to subject and topic with a short description and related URL's (Hart & Rolletschek, 2003). Domain-specific search engines are search engines that only return webpages related to a specified domain. Google is not a domain-specific search engine (Oyama et al., 2004).

3.7.3 General-purpose search engines

Search engines are indispensable for Internet users and have been built to aid users in locating information on the Web. Murnane (2003) states that 3 out of 4 Internet users use search engines, and they have a less than 50% success rate. Weideman (2001) claims a typical success rate of 34% for Internet searching.

The components of a typical search engine are:

- Spider – a program that traverses the Web in search of data by using different algorithms.
- Indexer – a program that indexes webpages retrieved and store the results in a database.
- Retrieval and ranking – to read the indexed webpages and present ranked results to users.
- User interface – allows the users to query the database (Chau et al., 2002).

3.7.4 Spiders

Search engines usually use spiders (also known as "bots", "crawlers", "worms" or "wanderers" (Andrus, 2003; Chau et al., 2002)) to retrieve pages from the Web by following URL links in pages using standard HTTP protocols. Spiders use different algorithms to control their search.

The following are methods generally used by spiders to explore the Web searching for webpages relevant to a particular domain:

- as many domains have specialised content, spiders are restricted to stay in particular Web domains (Manber *et al.*, 1997). For example, most webpages within the toyota.com domain will be pertinent to cars (Chau *et al.*, 2002),
- some spiders are constrained to gather only webpages which are a fixed number of links away from the initial URL or domain (Manber *et al.*, 1997; Sumner et al., 1998) and
- more advanced spiders analyse webpages and hyperlinks to distinguish which documents are relevant and can be downloaded (Cho *et al.*, 1998; McCallum et al., 1999).

3.7.5 Ranking

The main purpose of a search engine is to attempt to return relevant results that will satisfy a users request. Search engines present their results as an ordered list of websites. This list is ranked with the most relevant websites first, which is determined by the search engines own unique ranking algorithm. Weideman (2001) states that ranking refers to a technique used by search engines to classify and display search results. The results returned are based on the most relevant answer to the least.

A number of factors affect the ranking decision taken by search engine algorithms – content is claimed to be an important one by Zhang *et al* (2005). Other factors include graphics, frames, metatags, code and several other back-end techniques (Weideman *et al* 2005).

4. Research design and methodology

The method employed in this research project will be the construction of separate websites, each having domain names in different gTLDs, e.g. mysite.com, mysite.net, mysite.biz etc. Each one will be populated with the same content, and will be submitted to the same major search engines. Data will be gathered on a weekly basis by statistically monitoring each website for search engine crawler and user visitations, and search engine ranking. Each website will be search engine crawler-friendly and include title tags in hyperlinks, alt tags in graphics, definitive webpage titles and keyword phrases in the meta tags.

More specifically, the following series of steps constitute the methodology:

4.1 Register URLs

Register the domain name "southafricanbooksale" in gTLDs: ".com", ".net", ".org", ".info" and ".biz". This domain name is currently available in these gTLDs.

4.2 Design Sites

All websites will be designed with the same content and will be developed in flat HTML.

4.3 Upload Sites

All websites will be hosted at the same location.

4.4 Register Sites with Search Engines

Each website will be individually registered with the same search engines at the same time.

4.5 Monitor Crawler and User Visits

Each website will be monitored for crawler and user visitation e.g. which search engine crawlers visited the site, from which search engines the user referred to etc.

4.6 Analyze and Summarize Results

Results will be monitored on a weekly basis and analyzed statistically.

Conclusions

Conclusions will be reached based on results received.

An example of how the statistical data, which will be collected on a weekly basis, can be represented:

Table 3: Sample of statistics recording format.

Week :		site: southafricanbooksale				
		.com	.net	.org	.info	.biz
	Search Engine					
Crawlers visited	Google					
	Yahoo					
Is Ranked	Google					
	Yahoo					
Highest Ranking	Google					
	Yahoo					

In parallel with this empirical experiment, data is also being gathered via interviews with senior officials at local e-commerce traders and search engines. These interviews will consist mainly of open-ended questions regarding their marketing infrastructure towards successfully branding their websites to Internet users.

These interviews will be quantitative in nature and will aim to contribute to the research objective by extracting the opinions from industry related persons to effectively determine whether an understanding is known of the importance a gTLD has on e-commerce website visibility.

5. Expected results and conclusion

The research will be aimed at both startup e-businesses who are thinking of purchasing a domain name and established e-businesses in all gTLDs. Startup e-businesses will find the results of the study useful when choosing a domain name in a gTLD other than “.com”.

Results of this study will also assist in determining how much marketing existing e-businesses, with gTLD other than “.com”, has to accomplish to have their domain name recognized to consumers.

Expected results from the methods employed above could prove that search engine crawlers rank websites in dot-com domains higher than alternative gTLDs. Further outcomes could also include that websites in alternative gTLDs have to spend more money when branding their websites to Internet users.

6. REFERENCES

- Amir, Y., Massey, D. & Tutu, C. 2003. A new look at the old domain system. [Online]. Available WWW: www.cnds.jhu.edu/pub/papers/cnds-2003-2.pdf (Accessed 26 August 2005).
- Andrus, W. 2003. Domain-specific search engine. [Online]. Available WWW: www.cs.sunyit.edu/~andrusw/projects.html (Accessed 3 April 2006).
- Anonymous. 2000. How much information? [Online]. Available WWW: www.sims.berkeley.edu/research/projects/how-much-info/internet.html (Accessed 18 May 2006).
- Apke, T. 2003. Legal strategies in resolving domain name disputes. *Industrial Management & Data Systems*, 103(5):332-338.
- Bekker, J. & Van der Merwe, R. 2003. A framework and methodology for evaluating e-commerce web sites. *Internet Research Electronic Networking Application and Policy*, 13(5):330-341.
- Berners-Lee, T. 2004. World Wide Web Consortium, New Top Level Domains Considered Harmful. [Online]. Available WWW: www.w3.org/DesignIssues/TLD (Accessed 18 April 2006).
- Brighty, M. & Markham, D. 2002. *Winning e-brand strategies – developing your on-line business profitability*. USA: Spiro Press.
- Chau, M., Chen, H., Qin, J., Zhou, Y., Qin, Y., Sung, W. & McDonald, D. 2002. Comparison of two approaches to building a vertical search tool: a case study in the Nanotechnology domain. In: *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'02)*, Portland, Oregon, 14-18 July 2002, New York: ACM Press, 135-144.
- Cho, J., Garcia-Molina, H. & Page, L. 1998. Efficient crawling through URL ordering. In: *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia. Available WWW: www.dbpubs.stanford.edu:8090/pub/1998-51 (Accessed 3 April 2006).
- George, P. 2004. Who runs the .gov and .edu domains? [Online]. Available WWW: searchenginewatch.com/searchday/article.php/3106221 (Accessed 25 April 2006).
- Green, D. 2000. The evolution of Web searching. *Online Information Review*, 24(2):124-137.

- Gritter, M. 2002. Stanford University, Segregating content by name: Legislation and Technology. [Online]. Available WWW: www.stanford.edu/~mgritter/domain-policy/legislating-content-segregation.html (Accessed 18 August 2005).
- Hart, T. & Rolletschek, G. 2003. The challenges of regulating the Web. *Info*, 5(5):6-24.
- Lawrence, S. & Giles, L. 1999. Accessibility and distribution of information on the Web. *Nature*, 400:107-109.
- Manber, U., Smith, M. & Gopal, B. 1997. WebGlimpse – Combining browsing and searching. In: *Proceedings of the Usenix 1997 Annual Technical Conference*, 6-10 January 1997, Anaheim, California.
- McCallum, A., Nigam, K., Rennie, J. & Seymore, K. 1999. A machine learning approach to building domain-specific search engines. In: *Proceedings of the International Joint Conference on Artificial Intelligence. IJCAI-99*. [Online]. Available WWW: www.ri.cmu.edu/pub_files/pub1/mccallum_andrew_1999_1/mccallum_andrew_1999_1.pdf (Accessed 3 April 2004).
- Melville, A.L. 2000. New cybersquatting law brings mixed reactions from trademark owners. *Journal of Science and Technology*, 6(13).
- Murnane, L.G. 2003. *The invisible Web: What search engines can't find and why*. [Online]. Available WWW: www.lib.umd.edu/groups/digref/laura.html (Accessed 23 April 2006).
- Nilsen, J.O. 2002. Mixing oil with water – Resolving the differences between domain names and trademark laws. *Journal of high technology law*, 1(1):47-62.
- Nguyen, X.T. 2001. Shifting the paradigm in e-commerce: move over inherently distinctive trademarks – the e-brand, i-brand and generic domain names ascending into power? *American University Law Review*, 50:937-970.
- OECD. 2004. Generic top level domain names: market development and allocation issues. *Organisation for Economic Co-operation and Development*. [Online]. Available WWW: www.oecd.org/LongAbstract/0,2546,en_2649_37441_32996952_119666_1_1_37441,00.html (Accessed 21 August 2005).
- Oyama, S., Kokubo, T. & Ishida, T. 2004. Domain-specific web search with keyword spices. *Knowledge and Data Engineering, IEEE Transactions*, 16(1):17-27.
- Panwar-Mridul, M. 2003. The impact of new generic top level domains on trademarks. *e-Law: Murdoch University Electronic Journal of Law*, 10(1):6-24.
- Partridge, M. 1997. A review of recent trademarks and unfair competition cases in the U.S. *IDEA: The Journal of Law and Technology*, 38(57).
- Schmitz, S.W. & Sint, P.P. 2003. B2C eCommerce strategy and market structure: The survey based approach. *Industrial Organization*. [Online]. Available WWW: www.ideas.repec.org/p/wpa/wuwpio/0301012.html (Accessed 1 March 2006).

Schuler, D.K & Kurtz, J.M. 2002. An Internet opportunity for museums: '.museum'. *International Journal of Nonprofit and Voluntary Sector Marketing*, 7(1):13-18.

Stojanovski, J. 2001. Search engines in-depth. In: *Proceedings of the CUC 2000: Quest for information Conference*, Zagreb. [Online]. Available WWW: http://www.carnet.hr/cuc/cuc2000/radovi/prezentacije/F/F2/f2_f.pdf (Accessed 13 April 2006).

Sullivan, D. 2005. Nielsen netratings search engine ratings. [Online]. Available WWW: www.searchenginewatch.com/reports/article.php/2156451 (Accessed 2 January 2006).

Sumner, R.G., Yang, K. & Dempsey, B. 1998. An interactive WWW search engine for user-defined collections. In: *Proceedings of the 3rd ACM Conference on Digital Libraries*, Pennsylvania, USA, June 1998: 307-308.

Thurow, S. 2003. *Search engine visibility*. Indianapolis, IN: New Riders.

Verisign. 2004. The domain name industry brief. *The Verisign domain report*, 1(4).

Weideman, M. and Chambers, R. 2005. The application of best practice towards improving website visibility to search engines: an empirical pilot study. *South African Journal of Information Management*, 7(4). [Online]. Available WWW: <http://www.sajim.co.za>. (Accessed 2 February 2006).

Weideman, M. 2001. Internet Searching as a study aid for Information Technology and Information Systems learners at a tertiary level. Unpublished Ph.D. Thesis, University of Cape Town.

Weideman, M. & Moncrief, M. 2000. *The status quo of technical management issues underlying the .co.za domains: a literature survey*. Unpublished Working Paper.

Weinberg, J. 2002. ICANN, 'Internet stability,' and new top level domains. *Communications Policy and Information Technology: Promises, Problems, Prospects*. [Online]. Available WWW: www.law.wayne.edu/weinberg/icannetc.pdf (Accessed 15 April 2006).

Zhang, J., and Dimitroff, A. 2005. The impact of webpage content characteristics on webpage visibility in search engine results (part I). *Information Processing and Management: an International Journal*, 41(3):665-690.