IMPROVING WEBSITE VISIBILITY AND INFORMATION RETRIEVAL OF E-COMMERCE VENTURES: A SPECIFICATION TO PLEASE THE CRAWLERS

M. Weideman
e-Innovation Academy, Cape Peninsula University of Technology

R. Chambers
e-Innovation Academy, Cape Peninsula University of Technology

ABSTRACT

Research has shown that even though the web provides numerous opportunities, many e-Commerce ventures are often ill equipped to exploit the web's commercial potential. To develop a visibility specification, the authors conducted research to identify potential elements which could provide a possible increase in website visibility. The evaluative results produced possible elements / designing techniques. These were used to formulate a specification to develop a visible website that is not only supported by arrant research, but also through real current applications. The research concluded that the application of these specifications to the design of e-Commerce websites will very likely lead to improved visibility.

1. INTRODUCTION

The aim of this research was to produce specifications based on empirical evidence, to guide e-Commerce designers in the construction of visible websites, which could contribute towards increased website traffic and subsequent income. The use of the Internet by organizations opens up numerous business opportunities. These include overcoming geographical and cost barriers to new markets, improving service to customers, access to world-wide communication, streamlining internal processes, restructuring relationships and sharing scarce data (Simpande & Jakovljevic, 2003).

2. LITERATURE REVIEW

2.1 Factors Affecting Website Visibility

A number of authors (Rowley, 2001, Srinivasan, et al, 1996, Thurow, 2003, Nobles & O'Neil, 2000, Weideman, 2004)) in the field of website visibility, collectively claim that:
- Visibility is a sustained process.
- Content must be accurate and compelling.
- Frames should be avoided where possible.
- Meta-tags should be included, even though they are not often used.
- Macromedia Flash should be avoided as it increases download time of content.
- Graphics should be minimised to improve download speed.

- Care should be taken using JavaScript, as it is unreadable by crawlers.
- A website should be listed with the most popular search engines.

2.2 Search Engines and Crawlers

Green (2000:126) defines a web directory as, "a pre-defined list of websites, compiled selectively by human editors through categorisation, according to subject or topic". A web directory provides a service which allows a user to navigate through several listings and an option to search the entire directory. According to this author, a search engine is a search service that uses retrieval software called spiders or crawlers that examine websites and then index them into a database of website listings according to their relevancy.

3. RESEARCH METHODOLOGY

It is evident that there is a need to assist SMMEs in utilising the benefits of presenting a visible website ranking (Baard, 2004). It was therefore a requirement to determine how visible SMME websites are on the major search engines. A set of empirical experiments was used to produce results.

3.1 Phase 1

The purpose of this phase was to prove that SMMEs are in need of assistance in utilising the full potential of a visible website. The first step was to identify a list of SMMEs which operate within the Western Cape. Over 4000 websites were gathered from the CCofC website (Cape Chamber of Commerce, 2005), and judgement and random sampling was used to reduce it to 300. The ranking position of the 300 website samples within six search engines, was analysed by making use of keyword verification software provided by Marketleap (Marketleap, 2005). The results proved that only 6.3% of the SMMEs analysed, appeared in any of the top 30 results of the six search engines, when searching for keywords extracted from the SMMEs' descriptions. It was thus confirmed that a reason for this research to be conducted exists.

3.2 Phase 2

In this phase, the 300 websites were analysed to identify the elements used or avoided by website owners to achieve top positions. This study was conducted by making use of 23 keywords from the category list, six learners and a set of criteria. The five trained learners were scheduled for the analysis process while the sixth learner was scheduled to record the results. Of the 300 expected outcomes, only 144 were achieved in the allocated timeframe. Results recorded by the sixth learner were then further refined into separate sections containing the questions and the average result as listed in the in paragraphs to follow.

4. RESEARCH RESULTS AND ANALYIS

4.1 Results: Domain Names and HTML Naming Conventions

Domain names from the criteria list were used to determine how the domain names of top webpages relate to the content of that page and how it was structured. The majority (68.8%) of websites analysed, had a domain name containing words related to the company, products, services or website. Certain domain names were short (83.3%), pronounceable (79.2%), descriptive (65.3%) and memorable (68.8%). These results are strongly supported in the literature.

HTML naming conventions from the criteria list were used to determine how the HTML naming conventions of top webpages related to the content of that page and also how it was structured. The majority (63.9%) of websites analysed had file names containing words related to the company, products, services or website. Of the file names analysed, the majority were short (67.4%), pronounceable (70.1%), descriptive (68.8%) and memorable (54.9%). These results are supported in the literature.

4.2 Results: Frames, Hypertext/Anchor Text and HTML

Galon (1999) claimed: "I wouldn't recommend using frames on the entrance page…". Of all the websites analysed, only 5.6% contained frames - none made use of the <noframes> Meta tag. These results confirm the notion that the use of frames should be avoided if possible. Hypertext / Anchor Text was used to determine how hypertext relates to the content of that page and how it was structured. Statistical results confirm the importance of hypertext. Results show that 100% of the websites analysed made use of hyperlinks. Of these hyperlinks, 91% had an accurate resemblance between the content of the current page and the page they link to. Search engines often rank HTML keywords listed in headers higher than the keywords listed in the body text (Kritzinger & Weideman, 2005, Nobles & O'Neil, 2000). The results returned strongly supported the use of information-rich headings. A total 79.9% of the websites analysed contained headings with keyword-rich sentences.

4.3 Results: Dublin Core, Meta tags

The results indicated that only a small portion (5.6%) of websites ranking in the top positions still make use of Dublin Core Meta tags. These could provide content rich information about the site, but according to the conducted literature survey, they carry little value in the search engine indexing process.

The main focus of Meta tags is to provide optional or additional information about a webpage and its content, quality, condition and other characteristics. The title tag was used to determine to what extent top webpages make use of title tags, how well these tags relate to the content of the site, the structure of the title, as well as the length of the title. The title tag has been proven by Craven (2003) to carry the highest weight of all Meta tags in webpage indexing. According to Galon (1999), search engines initially look at the title to find important keywords. A total of 99.3% of top webpages analysed, made use of a title tag. Of these webpages, 94.4% provided keywords related to products, services or company name, while 85.3% of keywords were meaningful and understood by the evaluators. Furthermore, the majority (92.3%) of webpages made accurate use of the title tag, by providing a title without stop words, fewer than ten words (73.4%) and keywords that were written in lowercase (96.5%).

The Meta-description tag was inspected to determine to what extent top webpages make use of description tags, how well these tags relate to the content of the site, as well as the structure of the tag. Statistical results from the analysis process reflected a strong presence (63.9%) of the description tag. Of the analysed websites containing the description tag, 95.7% provided keywords related to products, services or company name, while 79.4% were meaningful and understood by the evaluators. There was also a small percentage (14.1%) where description tags were used incorrectly. Examples include cases where keywords were repeated, or where the description was left blank. There were also a number of instances (35.9%) where the title tag was duplicated in the description tag.

Statistical results from the analysis process returned a presence (63.2%) of the keyword tag. Of the analysed websites containing the keyword tag, 96.7% included keywords related to products, services or company name, while 78% were meaningful and understood by the evaluators. Results indicated that only a few webpages (23.6%) made use of the header tag. Of the pages which made use of the header tag, 88.2% had header descriptions which included keywords related to products, services or company name, while 91.2% were meaningful and understood by the evaluators.

Robot tags are used in one of two ways. One being the inclusion of a robot Meta tag within the Meta tag section, and the other the use of robot tags within a separate file (called robots.txt). Both these methods were tested. The results proved that a total of 43.8% of websites analysed, make use of robot tags to redirect search engine crawlers.

Konia (2002) stated that "…certain engines look for keywords in the alt tags…". Statements made by this and other authors were supported by the research results, which showed that 69.4% of all sites analysed, made use of alt tags. Of these pages, 68.0% contained descriptions which included keywords related to products, services or company name, while 77.0% of webpages contained keywords that were meaningful and which were understood by the evaluators.

4.4 Results: Spamming, Flash and Banner advertising

Results from the statistical analysis indicated that only 3.5% of websites analysed, which listed in top results, contained possible signs of link spamming features. This result strengthens the notion that spamming should

be avoided at all times. Websites containing Flash were analysed to determine if the pages that contained large sections of Flash, and provided the user with an option to load the site without the use of Flash. Furthermore, sites were also analysed to determine if they contained Flash within their navigational structure. Survey results strongly supported the fact that Flash should currently be kept to a minimum. Results indicate that none of the sites analysed make use of Flash which take up more than 50% of the webpage's content.

Banner advertising was used to determine to what extent top webpages make use of banner advertising methods and how well these advertising methods relate to the content of the site. The statistical analysis process indicated that developers were divided on the use of banner ads. A total of 26.4% of the websites listed in top results, make use of the banner-advertising medium. Of those websites which made use of banner ads, 76.3% contained banner ads which reflected the content of the website.

4.5 Results: Link Popularity, JavaScript and Personal Interview

Survey results returned that all of the top websites analysed contained links pointing to webpages within the website, or links pointing to other pages. Furthermore, 84.0% of the websites which contained links, had links which related to the content of the site analysed.

The webpages were also scanned for the use of JavaScript files and the <noscript > tag. Results from the statistical analysis process indicated that 71.5% of websites ranking in the top positions made use of JavaScript. Of these websites, the majority (65.1%) enclose part or all of their JavaScript code in a separate file. A further 28.2% of the analysed websites listed in to top positions make use of the <noscript> tag. These results proved that a website could achieve relatively high listings with the use of JavaScript. The last phase involved an interview with Vorster (2005), a representative of a leading South African search engine (Ananzi). Vorster was asked to rank the elements discussed according to importance on a scale from one to five, where one had the highest importance. Of the elements discussed, Vorster rated the inclusion of all Meta tags as the most important, due to the fact that the inclusion of these tags is an absolute requirement by Ananzi, even though they do not really increase the visibility of the site. Secondly, Vorster rated hypertext as a two, due to the fact that the majority of search engines consider back links.

5. CONCLUSION

A total of 144 websites were analysed to identify the elements used and avoided by website owners to achieve top positions. This study was conducted by making use of nine keywords and four search engines. Visibility elements gathered through academic literature were then ranked according to the usage of these visibility elements within the websites ranking in the top positions when searching for predetermined keywords. The elements present in these websites were recorded. Furthermore, an expert interview was conducted with a representative of a leading South African search engine. The goal of this interview was to provide clarification on the importance of elements identified through the academic literature and quantitative research. The top five elements gathered from the expert interview were recorded.

To construct the final specification, the authors assigned a ranking to the top elements identified through the different research approaches. The ten elements were ranked to add up to 55 (1+2+3…+10). Since the first three elements occurred the same number of times, a ranking of 2 was assigned to all three elements. This ensured that the sum of the ten rankings remains 55 (1+2+3 = 6) $\rightarrow$ 6/3 = 2. The rankings mentioned above were combined in Table 1 to provide a final ranking for the compilation of a top ten list of elements. Table 1 is the specification which could improve the visibility of a website. A lower figure in the Rank column indicates a higher ranking, i.e. a more important element. Rankings of elements appearing in both research approaches were added and divided by two to produce an average ranking. Only one expert interview was used in the qualitative part of the research, while a high number of websites were included in the quantitative part of the research. The ranking system however, stabilizes the effect of this apparent imbalance. Quantitative results range from one to ten, while the qualitative results cover only one to five.

Further research is currently underway, where the model of Figure 1 is being implemented in the commercial world, using detailed before-after measurements on a business site, to determine its value and applicability.

Table 1. Leading visibility elements – SPECIFICATION

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

ACKNOWLEDGEMENT

REFERENCES

Baard, V.C, 2004. Interactive information consulting system for South African small businesses – Part 1. In South African Journal of Information Management, Vol. 6, No. 2, June. http://www.sajim.co.za [20 February 2005].

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

Craven, T.C, 2003. HTML tags as extraction cues for webpage description construction. In Information Science Journal, Vol. 6, pp. 1-12.

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

Green, D. 2000. The evolution of Web searching. In Online Information Review, Vol. 24, No. 2, pp. 124-137.

Konia B.S, 2002. Search Engine optimization with WebPosition Gold 2. Wordware Publishing, Texas.

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

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

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

Rowley, J, 2001. Remodelling marketing communications in an Internet environment. In Internet Research: Electronic Networking Applications and Policy, Vol. 11, No. 3, pp. 203-212.

Simpande, J. & Jakovljevic, M, 2003. Integrating the Internet and enterprise resource planning (ERP) systems in South Africa electricity utility companies. In South African Journal of Information Management, Vol. 5, No. 1. http://www.sajim.co.za [20 April 2005].

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

Thurow, S, 2003. Search Engine Visibility. New Riders Publishing, IN, USA.

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

Weideman, M, 2004. Ethical issues on content distribution to digital consumers via paid placements, as opposed to website visibility in search engine results. Proceedings of ETHICOMP 2004. Syros, Greece, Vol. 2, pp. 904-915.