

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/257718722>

# Academic Content – A Valuable Resource to Establish your Presence on the Web

Article in *Procedia - Social and Behavioral Sciences* · February 2013

DOI: 10.1016/j.sbspro.2013.02.036

---

CITATIONS

0

READS

70

1 author:



[Melius Weideman](#)

Cape Peninsula University of Technology

181 PUBLICATIONS 599 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Development of a search engine marketing model using the application of a dual strategy [View project](#)



Semantic annotation through data serialization for improving relevant and personalized voice search results on smartphones [View project](#)



The 2nd International Conference on Integrated Information

## Academic Content - a valuable resource to establish your presence on the Web

Weideman, M\*

*HoD, Cape Peninsula University of Technology, PO Box 652, Cape Town, 8000, South Africa*

---

### Plenary Paper

#### Abstract

It has been proven beyond any doubt that high rankings on search engine result pages are non-negotiable for commercial websites. These high rankings can be achieved through a variety of methods, ethical and unethical. Both are being used extensively to impress crawlers, either in-house or by external search engine optimization experts. Arguably one of the most important components needed to achieve this high visibility, is the use of textual content honeycombed with the concentrated use of relevant keywords. This type of content is time-consuming and therefore expensive to create.

It is generally accepted that any active academic should publish his/her research results regularly in a variety of formats including books, book chapters, journal articles and conference papers. These publications are used to measure the value of a researcher's work, and citations of these works have been used as a measure of success.

Over a period of time an active academic amasses a large volume of descriptive, keyword-rich text on one or more closely related topics. This body of text is a valuable resource which should be used as a tool to establish a strong Web presence for each active academic, adding to the traditional exposure achieved through paper and online publications.

In this keynote, a critical evaluation is done of the usage of large volumes of research text coupled with website visibility principles, to establish a strong presence with search engine crawlers.

© 2013 The Authors. Published by Elsevier Ltd.

Selection and peer-review under responsibility of The 2nd International Conference on Integrated Information

*Keywords:* website visibility; academic content; search engines; body text

---

\* Corresponding author. Tel.: +27 21 9135515; fax:+27 21 9134801.

*E-mail address:* [weidemanm@cput.ac.za](mailto:weidemanm@cput.ac.za), CPUT, Cape Town, South Africa

## **1. Introduction**

It has been proven in prior research that the ranking of websites on search engine result pages (SERPs) is of importance to website owners. This is even more so when the website is commercial by nature, i.e. selling products or services for example. High rankings on SERPs imply more human traffic, which in some cases convert to more paying clients and higher return on investments [1].

The competition for the top spots on SERPs of mainline search engines (like Google, Yahoo! and Bing) is fierce, and some clients are spending large amounts of money with search engine optimization (SEO) companies to ensure that they remain on top or close to the top of the rankings. The large expenditure on pay-per-click (PPC) schemes confirms this trend, even though more than half of Internet users prefer not to click on PPC results [2].

Prior research has shown that a number of factors influence the ranking of natural results on search engines, but that the detail of the algorithms and exactly how ranking is calculated, are unknown outside the search engines themselves. Some work has been done on identifying and/or ranking these factors, and two factors appear to feature strongly towards the top of this list, if not occupying the top two positions: quality and quantity of inlinks, and keyword rich, relevant, original content [3]. In one case, empirical research has put these two in the top two positions, above 15 other factors [1].

## **2. Cost of human time**

Against this backdrop one problem associated with high rankings on the SERPs become evident: both factors above require (expensive) human time and specific expertise [4]. Creating real, live, relevant inlinks to a website requires human link builders to create content on various platforms across the Internet, with links to the central site and each other. Writing good content also requires a specific skill - high quality (normally English) technical copy-writing skills, plus an intimate knowledge of the business represented by the website. The text created thus must be keyword rich which will not alarm search engine algorithms as being an attempt at keyword stuffing.

Two main types of content on most websites are relevant at this point. The first is well-written, keyword rich natural English text, which convey a strong message without irritating a human reader with repetitive wording or nonsensical, strangely constructed sentences. Secondly, text exist which have either a too high percentage of keywords, or consist of seemingly meaningless sentences strung together, or a combination of both. The first type has probably been created at high cost by a human expert, while the second is possibly the output of an automatic content generator program.

## **3. Markov versus natural content**

As a result of the high cost associated with human expertise, many programs exist which automatically generate English content, given some keywords and/or phrases to indicate the type of content to be generated. One example is that of the MIT students who wrote a Markov generator to generate typical academic conference papers with the correct structure and layout, but nonsensical content [5]. See Figure 1 for an example.

# The Relationship Between Journaling File Systems and Virtual Machines with Optics

Weideman and Cronin

## Abstract

Randomized algorithms and flip-flop gates, while structured in theory, have not until recently been considered typical. Given the current status of wearable communication, scholars daringly desire the structured unification of the memory bus and redundancy, which embodies the appropriate principles of artificial intelligence. We describe a method for extreme programming, which we call Optics.

## Table of Contents

[1\) Introduction](#)

[2\) Principles](#)

[3\) Distributed Models](#)

[4\) Results](#)

- [4.1\) Hardware and Software Configuration](#)
- [4.2\) Dogfooding Our Approach](#)

[5\) Related Work](#)

[6\) Conclusion](#)

## 1 Introduction

In recent years, much research has been devoted to the development of RAID; on the other hand, few have synthesized the intuitive unification of the Ethernet and agents. The usual methods for the construction of hash tables do not apply in this area. The notion that cyberneticists agree with spreadsheets is generally considered key [20]. Nevertheless, linked lists [10] alone can fulfill the need for active networks.

In our research, we validate that although the location-identity split [1, 24, 28, 22] and kernels are never incompatible, the foremost stochastic algorithm for the analysis of link-level acknowledgements by David Johnson [28] is NP-complete. We view networking as following a cycle of four phases: exploration, development, location, and refinement. For example, many algorithms manage constant-time technology. The effect on e-voting technology of this has been well-received. Optics provides cacheable epistemologies. Thusly, we see no reason not to use virtual models to emulate Lamport clocks.

Fig. 1. Example of the output of a Markov content generator [5]

In principle it is easy to use this kind of content generator, feed it some relevant keywords and hyperlinks, and then simply plug in reams of Markov content on hundreds or even thousands of webpages in an attempt to increase rankings. While human readers should discern that the content does not make sense, search engine crawlers might accept it, and rank the page well for the supplied keywords. It is up to algorithm development by the search engines to keep their crawlers one step ahead to enable them to separate natural from Markov content. Similarly, some unethical Web authors consider it time-saving and productive to find existing, well-written and relevant Web content, then copy and paste it into their own webpages, in an attempt to please search engine crawlers even more. This action, termed content scraping, creates duplicate content which might not only infringe on copyright, but also dilutes Web content and create more and more similar or identical webpages, with resultant user frustration.

On the other hand, academics have the opposite situation existing in their academic environment. Any active academic is expected to do research, write it up and have it published on any one or more of many possible platforms: journal articles, conference proceedings, books and chapters, technical reports, and many more. These academic documents, by nature of the way they were created, all have the following characteristics in common:

- E they are well-written and error free (partially owing to the peer-review and editorial academic processes),
- E they are characterized by multiple, keyword-rich abstracts, phrases, summaries and body text (regardless of the field - be it engineering, medical science or the humanities),
- E they contain content of value which is highly relevant to a select audience.

Every academic has therefore already created (or have been part of the process) large volumes of natural content, with no duplication. These characteristics are in fact the same as those required from any piece of well-written WWW content to be hosted under a domain name. It follows that academics have (in some cases, partial) authorship to unique, valuable content of the kind that any industry is prepared to pay large sums of money to have written for them.

Traditionally academic publication platforms have been paper-based only, but are currently parallelizing more and more into dual-medium, with online media becoming a default in some cases [6].

The question is now: how can academics make use of this opportunity to market their field of expertise? This is the crux of this paper.

#### **4. Hyperlink structure**

The concept of indicating an interest in a piece of information, which will transport the reader to another, relevant location, as embodied in the WWW hyperlink structure, has been described decades ago [7]. Some authors consider Bush and his description of the "Memex" machine to be the father of the Internet, although he passed away well before the Internet was born.

The hyperlink concept closely echoes the academic principle of citing other, established authors to support or guide one's own research. In both cases, the fact that one concept links to another is an expression of trust in the value of the information found at the destination point [8]. It is the quantity and quality of these hyperlinks in part, on which Google has built the PageRank concept and the value it assigns to hyperlinks.

#### **5. Website visibility**

Website visibility is a measurement of how easily a website can be indexed, and how well the algorithms rank the content for (a) particular keyword(s)/key phrase(s). As such website visibility is not a single variable which can be plotted on a linear scale, but rather a collection of many factors and attributes of a given website. Similarly, the elements of a website which contribute to or detract from its visibility are many and varied [1]. Before website visibility can be achieved or measured, a given webpage must be indexed by a search engine. This status can easily be checked by using the Google "site:" operator (eg: site:www.mysite.com), or one of many programs that exist with this feature (eg: www.ranks.nl).

However, it is clear that there are two processes which could be used, separately or in tandem, to contribute to the increase of a website's degree of visibility. The one, SEO, involves changing the content and structure of a webpage to be more search engine friendly, and is in theory a once-off process which could be of benefit to the website for a long period of time, and involves no payment to a search engine. Secondly, a website owner can choose to set up a PPC campaign based on one or more keywords acquired through a bidding process with a search engine. This process involves a continuous funding stream (the volume chosen by the owner), but it

ensures immediate and targeted exposure. Ideally, both these should be used in tandem, but the choice of how to spend a certain budget on this kind of marketing lies with the website owner - see Figure 2.



Fig. 2. Roles of SEO and PPC in Website Visibility

The act of increasing the level of exposure of textual content to search engine crawlers is an ongoing and sometimes time-consuming process.

### 5.1. Creating original content

What is potentially an expensive, time-consuming process in industry, has actually already been achieved for an academic. As stated before, most academics have access to a large volume of specific text, written around a central research theme or set of themes. Even though copyright on some of this content probably exists, there are many ways to still expose parts of it without infringing any copyright laws.

### 5.2. Hosting content

A way must be found to host as much of this content as is legally possible on a variety of Web-based platforms. These include:

- E a webpage/site created, maintained and hosted by the researcher him/herself (this is no longer a technical or programming-based activity - many platforms exist whereby webpages can easily be created and hosted at no charge, or for a very low fee)
- E academic platforms, where it is possible to host content like journal article titles, author surnames, abstracts, etc - these include: [www.academia.edu](http://www.academia.edu), [www.mendeley.com](http://www.mendeley.com),
- E a microsite on the home institution's website, and
- E hosting on the institution's digital library.

Other forms of exposure will occur outside the control of the academic, including journals and conference organizers hosting papers presented/accepted, or even just abstracts, on their own websites.

### 5.3. *Manual search engine submission*

Each one of these content containers, once populated with natural content, should now be submitted to Google, Bing and DMOZ. Google is the first choice since it is by far the biggest search engine, and, regardless of one's own view of the company, omission from the Google index basically defines any website as invisible to searchers. Bing is the second biggest search engine and they also produce all Yahoo's answers. DMOZ is the only directory amongst the three, and uses human (volunteer) editors to decide which submissions should be indexed and which not. This implies that the quality of the webpages in the index should be higher than that of a crawler-generated index.

For Google and Bing, simply start at the homepage and follow the menus to get to the URL submission page. Alternatively, use the direct domain names:

- E [www.google.com/addurl](http://www.google.com/addurl)
- E [www.bing.com/toolbox/submit-site-url](http://www.bing.com/toolbox/submit-site-url)

DMOZ ([www.dmoz.org](http://www.dmoz.org)) will take a little longer, since there is no single URL submission page. The user has to first drill down through the directory structure to find the most relevant directory where his/her site belongs. This is an important step, since it prevents delays if the editor decides that the author's choice of sub-directory was irrelevant, and the submission will join the back of another queue. Once the most relevant sub-directory has been found, follow the "suggest URL" at the top of the category page, and respond to all the information requests following.

Users should refrain from using online automatic submission to search engines. Often a small fee is charged, in exchange for "submission to 100's of search engines". These services sometimes sell the authors' email address as a live lead to other services, and/or they might include the URL on webpages along with thousands of other, unrelated addresses. As a result the webpage might find itself in a spamdexing environment, which could even lead to banning of the domain from the search engines.

### 5.4. *Monitoring website visibility*

As final step, the degree of visibility should be monitored on an ongoing basis. Specialized programs could be used for this purpose - both freely available and paid systems. Examples include:

- [www.ranks.nl](http://www.ranks.nl)
- [www.seomoz.org/tools](http://www.seomoz.org/tools)

- [www.visible.net/tools/analyzer/](http://www.visible.net/tools/analyzer/)
- [www.spydermate.com](http://www.spydermate.com)

An extract of a typical report of such a program is given in Figure 3.

Web Page Analysis	
Factor	Page Analysis Results
HTTP Code	<b>GOOD:</b> No problem with the web page http status. <b>NOTICE:</b> The web page http code is 200 OK status.
Load Time	<b>GOOD:</b> No problem with the web page load time. <b>NOTICE:</b> The web page load time is 1.33 seconds.
Page Size	<b>GOOD:</b> No problem with the web page size. <b>NOTICE:</b> The size of the web page is 7084 bytes.
Page Text	<b>GOOD:</b> No problem with the page body text. <b>NOTICE:</b> The page body text has 310 total words.
Keywords	<b>GOOD:</b> No problem with the keywords found on the page. <b>NOTICE:</b> There are 51 targeted keywords.
Anchor Tags	<b>GOOD:</b> No problem with the anchor tags found in text links. <b>NOTICE:</b> There are 14 href anchor tags.
Nofollows	<b>WARNING:</b> No nofollow anchor tags were discovered. <b>NOTICE:</b> Nofollow tags are NOT a requirement, however your links may be passing popularity to less important pages. <b>ADVISORY:</b> Consider adding nofollow tags to less important text links or website sections.
Bold Tags	<b>WARNING:</b> No bold or strong tags were discovered. <b>NOTICE:</b> Bold or strong tags are NOT a requirement, however search engines do use them to determine the topic(s) for a given page. <b>ADVISORY:</b> Consider adding bold or strong tags to at least one main keyword or keyword phrase on this page.
Headings	<b>GOOD:</b> No problem with the heading tags found on the page. <b>NOTICE:</b> There are 4 heading tags.
Alt Tags	<b>GOOD:</b> No problem with the alt tags found in images. <b>NOTICE:</b> There are 3 image alt tags.
Out Links	<b>GOOD:</b> No problem with the outbound links on this page. <b>NOTICE:</b> There are 14 outbound links.
In Links	<b>WARNING:</b> No inbound links were discovered. <b>NOTICE:</b> Inbound links are NOT a requirement, however search engines may not give as much weight to this page. <b>ADVISORY:</b> Consider building the number of inbound links that point this page from other popular related websites.

Fig. 3. Sample extract of Website Visibility factors reported.

## 6. Conclusion

The act of increasing the level of exposure of textual content to search engine crawlers is an ongoing and sometimes time-consuming process. Academic content provides an opportunity for wide exposure of a research topic to search engine crawlers [9]. The website visibility thus created can be monitored, and should be increased over time by applying some simple SEO techniques.

## References

- [1] Weideman, M. (2009). *Website Visibility: The Theory and Practice of Improving Rankings*. Oxford: Chandos.
- [2] Neethling, R. (2008). Search engine optimisation or paid placement systems - user preference. Masters Thesis, Cape Peninsula, University of Technology, Cape Town.
- [3] Sullivan, D. (2011). Introducing: The Periodic Table Of SEO Ranking Factors. Available WWW: <http://searchengineland.com/introducing-the-periodic-table-of-seo-ranking-factors-77181> (accessed 30 August 2012).
- [4] Halvorson, K. (2011). Understanding the Discipline of Web Content Strategy. *Bulletin of the American Society for Information Science and Technology*, 37(2), 23-25.
- [5] SCIGen - An Automatic CS Paper Generator. Available WWW: <http://pdos.csail.mit.edu/scigen/> (accessed 29 August 2012).
- [6] Gould, T.H.P. (2010). Scholar as E-Publisher. *Journal of Scholarly Publishing*, 41(4), 428-448.
- [7] Bush, V. (1945). As we may think. *Atlantic Monthly*, 176(1), 101-108.
- [8] Thelwall, M. & Sud, P. (2011). A comparison of methods for collecting web citation data for academic organisations. *Journal of the American Society for Information Science and Technology*, 62(8), 1488–1497.
- [9] Weideman, M. 2012. Abstracts of research publications on Website Visibility, Search Engines and Information Retrieval. Available WWW: [www.web-visibility.co.za/website-visibility-abstracts-seo.htm](http://www.web-visibility.co.za/website-visibility-abstracts-seo.htm) (accessed 30 August 2012).