

**Department of Computer Science
and Mathematics
Munich University of Applied Sciences (MUAS)**

In collaboration with

**Faculty of Informatics and Design
Cape Peninsula University of Technology (CPUT)**

The research as described in this document, was done by the first author as noted below. He/she was a student at the Munich University of Applied Sciences (MUAS) in 2014 - the topic was given as a student project. It has not been peer reviewed, but has been edited for basic grammar and accuracy. References have been standardized as far as possible according to the Harvard system. Consider this document as a working paper, to be used for basic referencing but not as seminal source for research work. It could be useful for research in the fields of Website Visibility, Information Retrieval and Search Engines.

**The degree of visibility to search engines enjoyed by
university digital libraries**

Author: Oliver Skorna (MUAS)

Student Number: 18

Mentor: Prof M Weideman (CPUT)

Submitted 29 October 2014 Edited November 2014
Information Systems Digital Libraries

Harvard Reference:

Skorna, O. & Weideman, M. 2014. The degree of visibility to search engines enjoyed by university digital libraries. *Working Paper*, Munich University of Applied Science, Munich. Available online: <http://www.willbeadded later.com>

ABSTRACT

Introduction. *With the rise of importance of information retrieval digital libraries become more and more sought after. University digital libraries serve as crucial information hubs for the public and researchers. With the abundance of information on the Internet findability of said information is key. A university digital library was investigated in order to find a lack of visibility towards search engines and possibilities of Search Engine Optimization (SEO) to improve visibility. The goal of this research was to take a deeper look at website visibility of university digital libraries.*

Method. *One open access university digital library (Cambridge Digital library) was identified and analysed by measurement of page rank with different key phrases. Additionally the page structure was analysed concerning SEO.*

Results and Analysis. *The page ranks for different key phrases implied good website visibility for two selected documents. Page analysis showed already good SEO with room of improvement.*

Conclusions. *It was concluded that findability through academic search engines still allows room of improvement and showed good website visibility in scope of this research.*

KEYWORDS

digital library, website visibility, internet, search engine optimization, seo, university

1. INTRODUCTION

In today's world the use of a computer is more and more leaning towards information retrieval than digital storage or word processing (Huang *et al*, 2003). As information retrieval is gaining importance so is availability of said informations. Throughout history libraries were the main holder of information. With the advancement of the age of information digital libraries become more important due to better accessibility over the Internet.

Historically digital libraries were often met with resistance. Arms (2012) concluded beneficial outcome for university libraries making the change to digitization. In the end it helps themselves to brace against financial pressure.

In the age of globalization fast and easy access to information is key for researchers. Digital libraries often can satisfy this need for information. In this context accessibility is one of the main restrictions for usability.

Finding information is only one of the problems with online research. Evaluation of found information is the other side of the coin (Smith, 2012). The question of trust often times leads to further research and prolongs academic work. Digital libraries of institutions like universities can

provide an inherent trust due to their reputation and therefore can shorten the academic research process.

With the rise of information retrieval SEO is helpful for successfully finding the information. Finding a document within a digital library is a two step problem. First the digital library containing the document has to be identified. Then the document has to be searched within the digital library. Search engines can reduce this problem to one step.

Additionally search engines offer support for multilingual searching for free while digital libraries often lack this feature (Soules, 2010). This is another argument for boosting the website visibility of a digital libraries through SEO.

Beel *et al* (2010) have identified reservations against SEO in academic circles. It is feared that SEO in a way dictates academic work. In spite of this Academic SEO was suggested as a common practise beneficial for everybody.

It is to note that digital libraries can provide limited access to its information due to licenzing. This research assumed public access to the information.

2. OTHER RESEARCH

As the usage of information retrieval through search engines takes over the more traditional approach through libraries (digital or analog) Onaifo *et al* (2013) proposed SEO as an imperative tool to boost a library's page rank and therefore its visibility. The research identified website characteristics influencing the page rank and came to the conclusion that most library websites do not rank high globally on Search Engine Results Pages (SERP).

Furthermore Arlitsch *et al* (2013) gave an overview of SEO within the context of institutional repositories. The researchers found low information retrieval possibilities and proposed practical advice on SEO techniques adresssing this problem.

3. METHODOLOGY

While Onaifo *et al* (2013) focused on the website visibility of digital libraries this research took a closer look at the visibility of documents contained in a university digital library.

This research used the Cambridge Digital Library <http://cudl.lib.cam.ac.uk/> as an example of a publicly accessible digital library. The University of Cambridge has build a substantial digitization of historical work throughout history readily available for public use. The documents are presented as digitized images of each page.

The websites (web pages of the Cambridge Digital Library and search engine result pages) used by this research were accessed on 25th of October 2014.

3.1 Analysis of page visibility

The visibility of two documents published on the Cambridge Digital Library was analyzed through search of different key phrases and the achieved page rank. Multiple key phrases were used:

- document title
- author
- name of one chapter
- description
- mix of document title and author surname
- some content of the document

Each key phrase was typed in three different search engines and the page rank of the document was recorded. Two major search engines (Davies 2013) and a representative of academic search engines were chosen:

- Google at <http://www.google.com>
- Bing at <http://www.bing.com>
- Google Scholar at <http://scholar.google.com>

The search engines were configured to only display english results.

Due to relevancy only the first result page was analyzed as Gay *et al* (2004) found a sharp drop off in time spent viewing the link after 10 results. This is probably due to search engine's default behaviour of only listing ten results per page.

One iteration was done with the pure key phrase and for the other iteration the key phrase „Cambridge Digital Library“ was added. This was done to simulate users who already know the location of the document and want to find it without accessing the library search engine.

The two web documents analyzed by this research were:

- Elementary Mathematics by Isaac Newton found at <http://cudl.lib.cam.ac.uk/view/MS-ADD-03959/3>
- Papers on the tides and trade winds by Matthew Flinders found at <http://cudl.lib.cam.ac.uk/view/MS-RGO-00014-00051/358>

3.2 Analysis of page structure

The analysis of the page structure concerning website visibility was done in conjunction with the Chambers model for website visibility (Chambers, 2005). It suggests multiple best practices used for SEO. This research analyzed inclusion of metatags, hypertext/anchor text, prominent domain names, prominent headings, prominent HTML naming conventions as positive factors and flash content, visible link spamming, frames and banner advertising as negative factors website visibility.

The web page <http://cudl.lib.cam.ac.uk/view/MS-ADD-03959/3> was analyzed for SEO possibilities.

4. RESULTS

4.1 Structure of DL

Elementary Mathematics by Isaac Newton found at http://cudl.lib.cam.ac.uk/view/MS-ADD-03959/3			
Key phrase	Page Rank		
	Google	Google Scholar	Bing
Elementary Mathematics	-	-	-
Elementary Mathematics Cambridge Digital Library	1	-	1
Isaac Newton	-	-	-
Isaac Newton Cambridge Digital Library	_*	-	_*
Observations on the algebra of Kinckhuysen	1	-	-
Observations on the algebra of Kinckhuysen Cambridge Digital Library	1	-	-
Notes and papers regarding algebra and geometry	1	-	-
Notes and papers regarding algebra and geometry Cambridge Digital Library	_*	-	1
Newton Elementary Mathematics	2	-	9
Newton Elementary Mathematics Cambridge Digital Library	1	-	1
Substractio in integris	-	-	-
Substractio in integris Cambridge Digital Library	-	-	-

Table 1: Page rank of Elementary Mathematics by Isaac Newton for

different key phrases.

Papers on the tides and trade winds by Matthew Flinders found at http://cudl.lib.cam.ac.uk/view/MS-RGO-00014-00051/358			
Key phrase	Page Rank		
	Google	Google Scholar	Bi ng
Papers on the tides and trade winds	1	-	1
Papers on the tides and trade winds Cambridge Digital Library	2	-	1
Matthew Flinders	-	-	-
Matthew Flinders Cambridge Digital Library	_*	-	_*
Nicholas May on the cause of the tides	9	-	-
Nicholas May on the cause of the tides Cambridge Digital Library	2	-	-
This mixed volume of correspondence stretches from the 1780s to 1820s	2	-	1
This mixed volume of correspondence stretches from the 1780s to 1820s Cambridge Digital Library	5	-	3
Flinders tides and trade winds	2	-	9
Flinders tides and trade winds Cambridge Digital Library	1	-	-
By the opportunity of a friend being released from this prison	-	-	-
By the opportunity of a friend being released from this prison Cambridge Digital Library	-	-	-

Table 2: Papers on the tides and trade winds by Matthew Flinders for different key phrases.

* While the direct link wasn't in the page result a link to an overview page with a link to the website was given.

Table 1 and table 2 show the results of page rank for the different key phrases regarding the three search engines.

The results show good visibility of the documents towards the two search engines Google and Bing although Bing showed a lack of findability regarding chapter titles.

The chapter titles are embedded in the web page through a hyperlink and it therefore suggests

that Bing doesn't rate in-site hyperlinks as much as Google.

The lack of visibility towards the academic search engine Google Scholar is conspicuous.

Furthermore no sign of in-text findability was found.

The lack of page rank for author name was not surprising due to the generality of the key phrase.

4.2 Web page structure concerning SEO

– Inclusion of meta tags

Metatags for title, description and keywords were included but rather short. Weideman (2007) suggested longer and keyword rich content for these tags.

– Hypertext/anchor text

The web page was found to make use of hypertext. Most notably links to each chapter of the document. These links could be improved to contain keywords like the title and author of the document.

– Prominent headings

Headings were included as non-visible elements especially for search engine crawlers. Appernt headings were implemented as special formatted span-tags. These could be changed to h-tags to improve visibility.

– Prominent HTML naming conventions

For this web page a generic URL was used. This could be changed to include the author and or document title.

– Flash content

No flash content was found.

– Visible link spamming

No visible link spamming was detected.

– Frames

No frames were used by this web page.

– Banner advertising

No banner advertising was used.

5. CONCLUSION

It was concluded by Mayr *et al* (2007) that coverage and up-to-dateness of specific web servers with Google Scholar varies greatly. Many years later this study still showed a lack of coverage. So it was not possible to find the chosen documents of Cambridge Digital Library with the use of Google Scholar.

It is suggested that Google Scholar can be used as a tool for information retrieval but the sole use of it could limit the reach concerning academic information.

With the importance of findability of information this research showed a good visibility of the Cambridge Digital Library to the two used search engines with room of improvement. It suggested some techniques in order to optimize search engine page rank.

This study was in line with the good website visibility concluded by Weideman (2013).

Optimization for in-text searchability of images could be improved if the text content of these images is manually transcribed and featured on the website or automatically extracted using image processing software. While Ashley *et al* (1995) put this out of reach with then available technology, current technology advances and projects are in place to solve this issue (Anonymous 2014).

It is to note that the small sample size of only one digital library puts a limit on deducibility of the findings in this research.

6. REFERENCES

Arlitsch, K. & OBrien P. S. 2013. *Improving the Visibility and Use of Digital Repositories through SEO*. Chicago: ALA TechSource.

Ashley, J., Dom, B., Flickner, M., Gorkani, M., Hafner, J., Huang, Q., Niblack, W., Petkovic, D., Sawhney, H., Steele, D., Yanker, P. 1995. Query by image and video content: the QBIC system. *Computer*, 28(9): 23-32.

Arms, W. Y. 2012. The 1990s: the formative years of digital libraries. *Library Hi Tech*, 30(4): 579-591.

Beel, J., Gipp, B. & Wilde, E. 2010. Academic Search Engine Optimization (ASEO). *Journal of Scholarly Publishing*, 41(2): 176-190.

Chambers, R. 2005. Search engine strategies: a model to improve website visibility for SMME

websites. Masters Thesis, Cape Peninsula University of Technology. Cape Town.

Anonymous. 2014. *Methoden der Bildverarbeitung und Digitalen Medien zur Unterstützung der paläografischen Forschung bei Transkription, Schreiberidentifikation und Schreiberprofilvergleich in historischen Dokumenten -Diptychon-*.

<http://gepris.dfg.de/gepris/projekt/203606267> [27 October 2014].

Davies, D. 2013. *Major Search Engines and Directories*.

<http://searchenginewatch.com/article/2048976/Major-Search-Engines-and-Directories> [27 October 2014].

Gay, G., Granka, L. A. & Joachims, T. 2004. Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2004)*, United Kingdom, Sheffield, 25 – 29 July 2004, 478-479.

Huang, H. & Liaw, S. 2003. An investigation of user attitudes toward search engines as an information retrieval tool. *Computers in Human Behavior*, 19(6): 751-765.

Mayr, P. & Walter, A. 2007. An exploratory study of Google Scholar. *Online Information Review*, 31(6): 814-830.

Onaifo, D. & Rasmussen, D. 2013. Increasing libraries' content findability on the web with search engine optimization. *Library Hi Tech*, 3(1): 87-108.

Smith, A. G. 2012. Internet search tactics. *Online Information Review*, 36(1): 7-20.

Soules, A. 2010. New e-sources, new models: reinventing library approaches to providing access. *Library Hi Tech News*, 27(2): 10-14.

Weideman, M. 2007. Use of Ethical SEO Methodologies to Achieve Top Rankings in Top Search Engines. *Proceedings of the 2007 Computer Science and IT Education Conference*, Republic of Mauritius, November 2007, 717-727.

Weideman, M. 2013. Comparative analysis of homepage Website visibility and academic rankings for UK universities. *Information Research*, 18(4).

<http://www.informationr.net/ir/index.html> [25 October 2014].