PROCEEDINGS OF THE 12th ANNUAL CONFERENCE ON WORLD WIDE WEB APPLICATIONS

21-23 September 2010
Durban
South Africa

Editors:

A. Koch
P.A. van Brakel

Publisher:

Cape Peninsula University of Technology
PO Box 652
Cape Town
8000

Proceedings published at
http://www.zaw3.co.za

ISBN: 978-0-620-48797-9

TO WHOM IT MAY CONCERN

The full papers were refereed by a double-blind reviewing process according to South Africa's Department of Higher Education and Training (DHET) refereeing standards. Before accepting a paper, authors were to include the corrections as stated by the peer-reviewers. Of the 73 full papers received, 55 were accepted for the Proceedings (acceptance rate: 75.3%).

Papers were reviewed according to the following criteria:

- Relevancy of the paper to Web-based applications
- Explanation of the research problem & investigative questions
- Quality of the literature analysis
- Appropriateness of the research method(s)
- Adequacy of the evidence (findings) presented in the paper
- Technical (eg language editing; reference style.

The following reviewers took part in the process of evaluating the full papers of the 12th Annual Conference on World Wide Web Applications, held from 21 to 23 September 2010 in Durban, South Africa:

Mr U Averweg
eThekwini Municipality / University of KwaZulu-Natal
Durban

Prof J Brits
Faculty of Information Sciences (Dean)
University of Wisconsin
Milwaukee
USA

Prof A Bytheway
Research Unit
Faculty of Informatics and Design
Cape Peninsula University of Technology

Dr W Chigona
Department of Information Systems
University of Cape Town
Cape Town

Dr AM El-Sobky
Educational Sciences
RITSEC
Cairo
Egypt

ISBN: 978-0-620-48797-9

Prof M Herselman
Living Labs
Meraka Institute
CSIR
Pretoria

Prof S Mutula
Department of Information Science
University of Botswana
Botswana

Dr Z Mitrovic
Department of Information Systems
Faculty of Management
University of the Western Cape
Cape Town

Dr David Raitt
European Space Agency
The Hague
The Netherlands

Prof A Singh
Department of Business Information Systems
University of KwaZulu-Natal
Durban

Prof P Weimann
Economics and Social Sciences
University of Applied Sciences
Berlin
Germany

Further enquiries:

Prof PA van Brakel
Conference Chair: Annual Conference on WWW Applications
Cape Peninsula University of Technology
Cape Town
+27 21 469 1015 (landline)
+27 82 966 0789 (mobile)

Search engine query generation for effective retrieval of known academic publications

M Weideman
Faculty of Informatics & Design
Cape Peninsula University of Technology
Cape Town, South Africa
weidemanm@cput.ac.za

Abstract

Academics often need access to publications of other academic work for reference. Sometimes however, some detail of a specific reference is known, but a copy of the full document is required. The objective of this research project was to compare different methods of query generation to successfully retrieve academic documents, while assuming that the user has the basic document details.

In a series of empirical experiments, 20 universities with digital libraries (not requiring logins) were identified. Five academic documents stored in each ones' digital library were further found and inspected. Searches were done, using three types of query for each one of these documents. Subsequently, rankings on search engine result pages were recorded. The current visibility of these documents was then calculated. After submission to Google, a waiting period was allowed for crawler visitation, and the searches and calculations repeated.

The resultant data was used to measure the success of the three different types of queries over 300 searches. This was done both before and after manually submitting each document's URL to Google.

Results indicate that using keywords from the document title produces the most efficient query, with much improvement after submission. Secondly, using a text sequence from the body produces the second-most efficient query, but with a small reduction in visibility. Finally, using author surnames produced a much less efficient query, although with slightly increased visibility.

It was concluded that academic searchers should concentrate on using a concatenation of weight-carrying keywords from the title of a known academic document as search query for most efficient document retrieval.

Keywords: Search engine, keyword, query generation academic publication.

1. Introduction

General information retrieval was first practiced centuries ago by the Sumerians (3rd millennium BC) (Weideman, 2001:21). However, searching for information on the Internet is as old as the Internet itself, dating back to the mid 1990's. More than a decade ago authors claimed that 71% of Internet users attempt to access information through search engines (CommerceNet/Nielsen Media, 1997).

During these early days, various programs were written to allow basic information retrieval, which later spawned more sophisticated interfaces and algorithms, know today as search engines. More recently it was claimed that Internet searching through the use of search engines is second only to the use of e-mail in the list of top Internet applications (Vinson, 2007; Ashley, 2007).

However, the starting point of any search for information remains the query generated and input by the human searcher. The quality of this query plays a large role in the success of the searching operation (Weideman, 2009:5). Sadly, research has proven that the general rate of success for Internet searches is rather low – typical figures quoted are between 30% and 40% (Voorbij, 1999: 604; Weideman, 2001:207).

2. Other research

2.1 Search engines

Many wars for dominance are ongoing in the information technology world. One of them is the battle for honours in the world of search. During the earlier days, AltaVista was the biggest search engine, boasting an index size of 250 000 000 webpages. At the time index size was used as the most important measure of success. Yahoo! took over at a later stage, in terms of user popularity. However, since then Google appeared and has firmly established itself as the leader of search. One author claims that Google generated more than 70% of all searches taking place in the USA (Sullivan, 2008).

Google has also established itself as the leader in paid search (Blankenbaker, 2009). However, lately Google seems to be losing part of its market share to other big competitors: it is claimed that the US stakes are currently: Google (62.6%), Yahoo! (18.9%) and Microsoft (12.7%) (Marshall, 2010). This still implies that about two out of every three searches done in the USA is funnelled through Google.

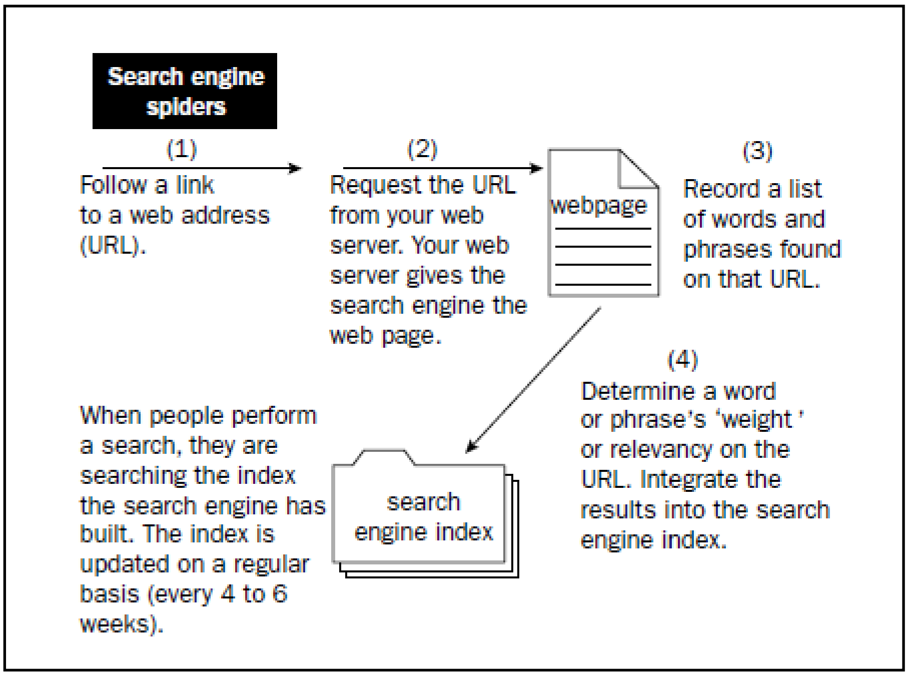
The components and operation of search engines can be described or viewed in many different ways (see Figure 1). It is generally accepted that:

- automated programs called *crawlers* trawl the live Internet for webpage content,
- this content is loaded into an *index*, which is conceptually a single massive data file,
- a program called an *interface* accepts human input and

ISBN: 978-0-620-48797-9

- an *algorithm* then attempts to find the best matches between this human input and the data stored in the index, then passes the results back to the *interface* (Weideman, 2009:22; Thurow, 2003:15).

Figure 1: Search engine components (Thurow, 2003:15)

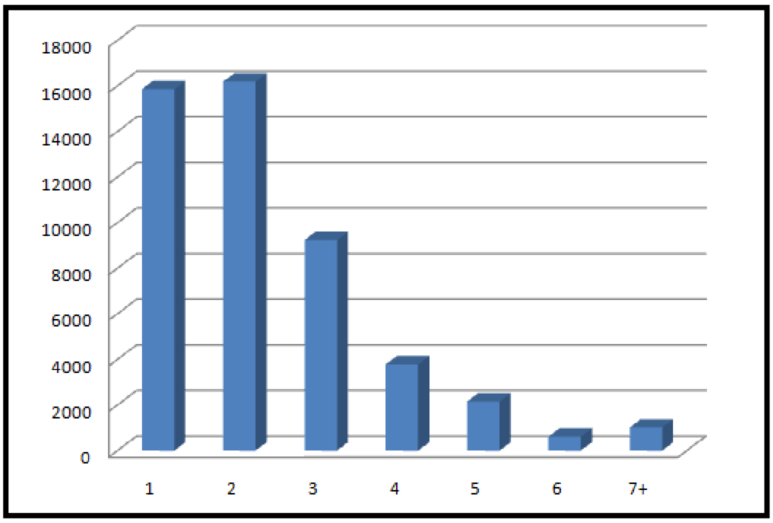


2.2 Query generation and success rate

Early research indicates that the number of words in a search query is strongly related to searching success. Most users seem to use around two words of text per query. Spink found an average query length of between 1.5 and 2.8 words to be used per query (Spink *et al*, 2000). At the same time another author proved that the peak frequencies are for single- and two-word searching (Jansen, 2000). See Figure 2.

ISBN: 978-0-620-48797-9

Figure 2: User's choice of number of keywords (Jansen, 2002)



Other authors have proven that there appears to be a link between query length and searching success – it is claimed that single-word queries can have a success rate of 30.3%, while two or more word queries could be as high as 42.3% (Weideman *et al,* 2004).

2.3 Academic publications

Academics active in research publish their research and results in a variety of ways – on paper in peer-reviewed journals, academic conference proceedings, books or book chapters and technical reports, to name the most common ones. Also, these documents could be published electronically - in databases, digital repositories and on other websites. Academics may make reference to these published works, to confirm their own methodologies, results or conclusions.

On the one hand, an overload of information (academic and otherwise) seems to be one of the problems brought on by the Internet and its free access to a virtually unlimited source of information (Baez *et al*, 2010). Filtering through masses of information in an attempt to retain only the relevant has become a skill necessary for academic survival. The first port of call seems to be free access to full text materials (Martin, 2010). However, finding such materials poses a continuous challenge to researchers. On the other hand, recent research has been done to enhance specifically the process of finding academic literature, dubbed "Academic Search Engine Optimization" by the authors (Beel, 2010). Although the research is general in nature, it does also focus on optimizing for Google Scholar specifically. The risk of creating spamdexing is also covered.

ISBN: 978-0-620-48797-9

Once found, the next challenge is to source the full reference for the document – author detail, name and date of publication, volume and number of journal, or title, date and location of conference and the actual full text. Often academics have a hard copy of an academic reference, but do not have the full document or reference detail available. It is this problem that this research project addresses.

3. Methodology

The objective of this research was to identify the best method to find academic material, through the standard Google search engine interface, on the Internet. Research was done on university digital libraries and quantitative data gathered to address this issue.

Although many institutions host digital libraries, the focus of this research was on academics searching for information, hence the decision to target university digital libraries. There are thousands of universities the world over, and a choice had to be made on which one's digital libraries to use in this research. It was considered prudent to use the digital libraries of those universities considered to be top ranked in world. Various research groups are working on the ranking of universities according to their own pre-determined scale. A number of these were identified, and each one was inspected to see if it used an academic method on which they based their ranking. Three were found (see Table 1), and each one's top 30 universities were listed.

Table 1: Three world university scoring systems

Name	Ranking URL	Method URL
ARWU	http://www.arwu.org/ARWU2009.jsp	http://www.arwu.org/FieldMethodology2009.jsp
Top Universities	http://www.topuniversities.com/university-rankings/world-university-rankings/2009/results	http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/simple-overview
Webometrics Ranking Of Universities	http://www.webometrics.info/top6000.asp	http://www.webometrics.info/methodology.html

A score of 30 for the top university in each of these three lists, 29 for the second from the top, etc was allocated, then added. Thus the minimum and maximum score any university could earn in the final list was: one (i.e. university appears only on one list, in position 30), and 90 (university appears in the top position on all three lists) respectively. The list was sorted from highest to lowest score, with a total of 57 universities making it into the final group - see Table 2.

Table 2: Combined top world university list

University	Score	University	Score
Harvard University	89	Texas A&M University	14
Massachusetts Institute of Technology	79	University of California, San Francisco	14
Stanford University	72	Mcgill University	13
California Institute of Technology	70	University of Maryland	13
University of Chicago	65	ETH Zurich (Swiss Federal Institute of T...	11
Cornell University	62	Purdue University	11
Columbia University	60	University of Edinburgh	10
University of California, Berkeley	55	University of Minnesota, Twin Cities	9
University of Pennsylvania	52	University of Tokyo	9
Yale University	49	KING'S College London	8
Princeton University	47	Pennsylvania State University	8
University of Oxford	47	Washington University in St. Louis	8
University of Michigan	45	Northwestern University	7
Johns Hopkins University	41	University of Hong Kong	7
University of Wisconsin - Madison	40	Kyoto University	6
University of Cambridge	38	University of Arizona	6
University of Washington	35	National Taiwan University	5
University of Illinois at Urbana-Champaign	33	New York University	5

University of Minnesota	33	University of Manchester	5
University of California Los Angeles	31	Rockefeller University	4
UCL (University College London)	27	University of Florida	4
Imperial College London	26	Ecole Normale Superieure, PARIS	3
Duke University	23	University of Colorado at Boulder	3
Carnegie Mellon University	21	University of California, Santa Barbara	2
University of Texas Austin	20	Virginia Polytechnic Institute and State University	2
University of California, San Diego	18	National University of Singapore (NUS)	1
University of Tokyo	16	University of British Columbia	1
University of Toronto	15	University of North Carolina Chapel Hill	1
Australian National University	14	Cape Peninsula University of Technology	-

However, every university in Table 2 did not have a digital library. Furthermore, it is known that search engine crawlers cannot find and/or index documents which are hidden behind a required human login. Thus only those universities offering a digital library without requiring that a human must log in were used. Each university domain was then found through ordinary Internet searching and URL guessing. Every domain was inspected in terms of its content, logo, sub-domain names and contact detail, to confirm that it was the official university website.

Furthermore, every official university website was inspected to determine whether or not it offered a digital library service to its users (reference was also made to www.opendoar.org/). Both those that did not and those which did, but required a human login, were omitted. The remaining top 19 names on this list were retained for the experiment, and the name of the author's own university digital library was added. These 20 university names are shaded in Table 2. The next step was to select five documents already indexed by each library, producing 100 test documents.

Documents were chosen from "Recent Submissions" or similar headings if it were offered. If not, links to Faculties and/or departments were followed, and a random selection of five documents was made. All 100 documents identified in this way were stored in PDF format.

Previous research has proven that 67% of Internet searchers on the average read only the first page of search engine results (Weideman, 2009:32). Google shows 10 results per page by default, so a test document was recorded as being listed by Google only if its exact URL is being shown within the first 10 results on a Google search result page.

A decision had to be taken on how to search for each document, before its presence or absence in Google search results could be recorded. It has been proven that phrase searching produces relevant results (Weideman, 2001:148), as well as simply stringing together a number of weight-carrying keywords, separated by spaces (Weideman, 2009:10). As a result, the following three rules for search query construction were formulated:

- filter out the first five keywords from the title of the test document, by omitting all stop words (the, that, a, etc) - the search query will then be these five (or four, or three, etc) keywords, separated by spaces,
- generate a search query by linking the surnames of all authors - the search query will then be a series of words (surnames) separated by spaces and
- generate a search query, in quotes, by copying the first words of the first sentence of the body of the document, up to but not including the first punctuation sign (comma, full stop, etc).

The three terms: “KE” (keywords), “SU” (surnames) and “SE” (sentence) will be used for these three types of searching respectively.

As an example, one of the test documents had the following attributes:

TITLE: Production of silicon carbide Al 2124 alloy functionally graded materials by mechanical powder metallurgy technique

AUTHORS: C.-Y. Lin, C. Bathias, H. B. McShane, and R. D. Rawlings

FIRST SENTENCE OF BODY TEXT:

" A mechanical powder metallurgy process involving a vibration stage before cold and hot compaction has been developed to manufacture functionally graded materials (FGMs).“

Table 3 lists the three search queries, based on the rules above, run on this test document.

Table 3: Three search queries based on one test document

Type	Query
KE	Silicon carbide AL 2124 alloy
SU	Lin Bathias McShane Rawlings
SE	"A mechanical powder metallurgy process involving a vibration stage before cold and hot compaction has been developed to manufacture functionally graded materials"

The results of each one of the 300 searches (20 digital libraries x 5 documents each x 3 search terms each) were recorded. "NF" ("Not Found") was used if the exact URL was not listed in the first 10 Google results, and a figure (eg "3") was used to indicate that the exact URL was found in position 3 of the results.

In parallel with this work, three separate text-only HTML webpages were created on three different domains, and the three pages were submitted to Google on the same day. The purpose was to determine what a reasonable waiting period for crawler visitation would be, and to see if the crawler could index these text-only pages with ease. All three text documents were indexed on day 15, and were easily findable by using small parts of their text content as a Google search query.

After the first phase searching and result recording, each one of the 300 document URLs were manually submitted to Google, using their standard website submission page (http://www.google.com/addurl/?continue=/addurl). A waiting period of 24 days was allowed for crawler visitation, and the second phase searching and result recording was done. Both the two search phases and the submission phase were each fitted into one day each, to ensure that all documents had the same exposure time to the Google crawler.

A system was implemented where a score was allocated depending on a document's ranking. A ranking of NF earned zero, rank 10 (the worst possible on a screen of 10 results) earned one, up to a ranking of one (score = 10). In this process a more natural system (higher is better) was incorporated.

4. Results

Due to the large volume of data, all results cannot be displayed. A summary of all the results for one university only is given in Table 4. An explanation of some of the table headings follows.

DOC1 QUERY1: The query generated for the first document, using method KE.

DOC1 ANS1 BE4: The ranking for the query submitted before the URL was manually submitted to Google: 1=first position, i.e. the best possible. NF=Not Found under the first 10 results on the search engine result page (SERP).

DOC1 SCORE1 BE4: The ranking converted to a score: rank 1=score 10, rank 2 =score 9, … rank NF=score 0, etc.

DOC1 ANS1 AFTR: The ranking for the query submitted after URL was manually submitted to Google.

………

DOC1 QUERY2: The query generated for the first document, using method SU.

………

DOC1 QUERY3: The query generated for the first document, using method SE.

………

ISBN: 978-0-620-48797-9

Table 4: Example results: Cambridge University

UNIVERSITY NAME	**University of Cambridge**	**DOC3 ANS2 AFTR**	**NF**
RANKING	38	DOC3 SCORE2 AFTR	0
HOME PAGE URL	http://www.cam.ac.uk/	DOC3 QUERY 3	"The performance of a recursive data structure"
DIGITAL LIBRARY URL	http://www.cl.cam.ac.uk/techreports/	DOC3 ANS3 BE4	NF
PASSWORD ACCESS	N	DOC3 SCORE3 BE4	0
DOC1 URL	http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-766.pdf	DOC3 ANS3 AFTR	NF
DOC1 QUERY1	Robust inexact geometric computation	DOC3 SCORE3 AFTR	0
DOC1 ANS1 BE4	1		
DOC1 SCORE1 BE4	10	VISIB SCORE BE4	3.33
DOC1 ANS1 AFTR	1	VISIB SCORE AFTR	3.33
DOC1 SCORE1 AFTR	10	CLASS	3
DOC1 QUERY 2	Smith	DOC 4 URL	http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-756.pdf
DOC1 ANS2 BE4	NF	DOC4 QUERY 1	Analysis Internet's structural evolution
DOC1 SCORE2 B4	0	DOC4 ANS1 BE4	NF
DOC1 ANS2 AFTR	NF	DOC4 SCORE1 BE4	0
DOC1 SCORE2 AFTR	0	DOC4 ANS1 AFTR	NF
DOC1 QUERY 3	"Geometric algorithms implemented using rounded arithmetic are prone to robustness problems"	DOC4 SCORE1 AFTR	0

DOC1 ANS3 BE4	NF	DOC4 QUERY 2	Haddadi Fay Uhlig Moore Mortier Jamakovic
DOC1 SCORE3 BE4	0	DOC4 ANS2 BE4	NF
DOC1 ANS3 AFTR	NF	DOC4 SCORE2 BE4	0
DOC1 SCORE3 AFTR	0	DOC4 ANS2 AFTR	NF
VISIB SCORE BE4	3.33	DOC4 SCORE2 AFTR	0
VISIB SCORE AFTR	3.33	DOC4 QUERY 3	"In this paper we study the structural evolution of the AS topology as inferred from two different datasets over a period of seven years"
CLASS	3	DOC4 ANS3 BE4	NF
DOC2 URL	http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-764.pdf	DOC4 SCORE3 BE4	0
DOC2 QUERY 1	Statistical anaphora resolution biomedical texts	DOC4 ANS3 AFTR	NF
DOC2 ANS1 BE4	3	DOC4 SCORE3 AFTR	0
DOC2 SCORE1 BE4	8	VISIB SCORE BE4	0.00
DOC2 ANS1 AFTR	2	VISIB SCORE AFTR	0.00
DOC2 SCORE1 AFTR	9	CLASS	3
DOC2 QUERY 2	Gasperin	DOC5 URL	http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-755.pdf
DOC2 ANS2 BE4	NF	DOC5 QUERY 1	Skin-detached surface interactive mesh editing
DOC2 SCORE2 BE4	0	DOC5 ANS1 BE4	1
DOC2 ANS2 AFTR	NF	DOC5 SCORE1 BE4	10
DOC2 SCORE2 AFTR	0	DOC5 ANS1 AFTR	1

DOC2 QUERY 3	"This thesis presents a study of anaphora on biomedical scientific literature and tackles the problem of anaphora resolution in this domain"	DOC5 SCORE1 AFTR	10
DOC2 ANS3 BE4	NF	DOC5 QUERY 2	Gao Hao Zhao Dodgson
DOC2 SCORE3 BE4	0	DOC5 ANS2 BE4	1
DOC2 ANS3 AFTR	NF	DOC5 SCORE2 BE4	10
DOC2 SCORE3 AFTR	0	DOC5 ANS2 AFTR	1
VISIB SCORE BE4	2.67	DOC5 SCORE2 AFTR	10
VISIB SCORE AFTR	3.00	DOC5 QUERY 3	"We propose a method for interactive deformation of large detail meshes"
CLASS	1	DOC5 ANS3 BE4	1
DOC3 URL	http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-757.pdf	DOC5 SCORE3 BE4	10
DOC3 QUERY 1	Improving cache performance runtime movement	DOC5 ANS3 AFTR	1
DOC3 ANS1 BE4	1	DOC5 SCORE3 AFTR	10
DOC3 SCORE1 BE4	10	VISIB SCORE BE4	10
DOC3 ANS1 AFTR	1	VISIB SCORE AFTR	10.00
DOC3 SCORE1 AFTR	10	CLASS	3
DOC3 QUERY 2	Adcock	AVER VISIB SCORE BE4	3.867
DOC3 ANS2 BE4	NF	AVER VISIB SCORE AFTR	3.933
DOC3 SCORE2 BE4	0	AVER CLASS	2.600

ISBN: 978-0-620-48797-9

Calculating the averages of the scores produces the results in Figures 5, 6 and 7. For example, the figures: "3.6 5.2" depict the average of "Score before" and "score after" for document 1 across the 20 digital libraries, using method KE. The pair "1.85 3.65" are the averages for document 2 across the 20 libraries, etc.

Table 5: Average scores for method KE

Ave Score Before Submission	Ave Score After Submission
3.6	5.2
1.85	3.65
5.6	5.15
2.4	3.14
1.95	3.25
3.08	**4.08**

Table 6: Average scores for method SU

Ave Score Before Submission	Ave Score After Submission
0.5	0.9
0	0
1	1.1
0.8	0.3
1	1.35
0.66	**0.73**

Table 7: Average scores for method SE

Ave Score Before Submission	Ave Score After Submission
1.45	2.4
2.9	1.4
2.9	2.9
2	1.85
3.45	3.8
2.54	**2.47**

ISBN: 978-0-620-48797-9

5. Analysis and conclusion

The crux of this research lies in the interpretation of Tables 5 to 7. In all three cases the "Score Before" column average (figure in bold) is an indication of the success rate of finding a document which is definitely stored in a digital library, exposed to searchers on the Internet. The scale ranges from "0", being worst, to "10", being the best result.

5.1 Method KE (Table 5)

The "before" and "after" figures are 3.08 and 4.08 respectively, and are the highest scores achieved out of the three methods. It also follows the expected trend of improving in visibility after the URL has been manually submitted to Google.

5.2 Method SU (Table 6)

The "before" and "after" figures are 0.66 and 0.73 respectively. Although the expected trend is present, these are the worst results out of the three searching methods.

5.3 Method SE (Table 7)

The "before" and "after" figures are 2.54 and 2.47 respectively, and fit in between the other two sets of results while being close to the best value. However, this is the only one of the three sets which has a trend opposite to the expected-the visibility has decreased instead of increased after search engine submission.

5.4 Conclusion

All the figures above are lower than reasonably expected. A figure of "10" would indicate that all twenty test documents occupied position number one (best) on the SERP for the given search query generation method. This ideal figure is unlikely to be achieved in real life, since no assumptions can be made about whether or not the owners of the 20 digital libraries have attempted to increase the visibility to crawlers of their documents. Using what has been proven in this research as the most efficient method, namely using the first five keywords from the title, a best score of 4.08 out of 10 was achieved.

The increase from "before" to "after" is also the highest of the three methods, which leads to the conclusion that **academic documents were most effectively retrieved when searched for using the KE method**. This concurs with the instinctive algorithm identified in earlier research:

" - Express the information need as a single, keyword-rich English sentence.
- Remove all the stop words from this sentence.
- Type the remaining string of keywords into a search engine search box." (Weideman, 2009:10).

ISBN: 978-0-620-48797-9

Secondly, the SU method was expected to have a reasonable success rate, since it was noticed from the sample of 100 test documents that many surnames were unique. This was a result of the wide variety of nationalities of authors in the sample set. Unique surnames as keywords (such as "Lin Bathias McShane Rawlings" from Table 3) have been proven to generate more effective search queries than commonly used ones (such as "Smith" from Table 4). However, the final results disproved this hunch. **Using author surnames is not an effective search query generation strategy for academic documents.**

Thirdly, phrase searching has been identified as an effective searching strategy in prior research, since it focuses the search by insisting that not only the keywords themselves but also the keyword sequence as specified be found on a webpage. The results indicate that it is only slightly less efficient that the KE method, with an average of 2.54 as opposed to 3.08. However, the "after submission" result decreased from 2.54 to 2.47. Although only marginal in value, the direction of change is against the expectation.

This anomaly can be ascribed to a number of factors. Google could have executed its regular Google dance during the 24 days (an updating of its indices, always resulting in some increases and decreases in ranking). Secondly, some of these documents could have been renamed, moved to a different subdirectory, or deleted - all these actions would have the same effect. The decrease could also be linked to the fact that all test documents were in PDF and not plain text format. Although Google has claimed that it does index PDF documents, other research has proven that it is not always the case. In conclusion, **phrase searching using the SE method could be effective,** but results will not be as reliable as with the KE method.

List of references

Ashley, M. 2007. Search – The Killer App in Vista and Office 2007. Available WWW: http://www.networkworld.com/community/node/22148 (accessed 3 August 2010).

Baez, M., Birukou, A., Casati, F., and Marchese, M. 2010. Addressing information overload in the scientific community. *IEEE Internet Computing*. Available WWW: http://doi.ieeecomputersociety.org/10.1109/MIC.2010.107 (accessed 3 August 2010).

Beel, J., Gipp, B. and Wilde, E. 2010. Academic search engine optimization (ASEO). Journal of Scholarly Publishing, *41(2):176-190.*

Blankenbaker, J. and Mishra, J. 2009. Paid search for online travel agencies: exploring strategies for search keywords. *Journal of Revenue & Pricing Management,* 8:155-165.

CommerceNet/Nielsen Media. 1997. Search Engines' Most Popular Method of Surfing the Web. Available WWW: http://www.commerce.net/news/press/0416.html (accessed 5 August 2010).

Jansen, B.J. 2000. The effect of query complexity on Web searching results. *Information Research*, (6)1. Available WWW: http://InformationR.net/ir/6-1/paper87.html (accessed 15 August 2005).

Marshall, J. 2010. Google loses Search Share to Yahoo, Microsoft in June. Available WWW: http://www.clickz.com/clickz/stats/1724692/google-loses-search-share-yahoo-microsoft-june (accessed 6 August 2010).

Martin, R.A. 2010. Finding free and open access resources: a value-added service for patrons. *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve,* 20(3):189-200.

Spink, A. and Xu, J.L. 2000. Selected results from a large study of Web searching: the Excite study. *Information Research*, (6)1. Available WWW: http://InformationR.net/ir/6-1/paper87.html (accessed 4 August 2010).

Sullivan, D. 2008. Hitwise: Google Again Hits New High; Microsoft & Yahoo Again New Lows. Available WWW: http://searchengineland.com/080617-173543.php (accessed 2 August 2010).

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

Vinson, J. 2007. Killer Apps in Knowledge Management. Available WWW: http://blog.jackvinson.com/archives/2007/01/17/killer_apps_in_knowledge_management.html (accessed 3 August 2010).

Voorbij, H.J. 1999. Searching scientific information on the Internet: a Dutch academic user survey. *Journal of the American Society for Information Science*, 50(7):598–615.

Weideman, M. 2009. *Website visibility: the theory and practice of improving rankings.* Oxford: Chandos Publishing.

Weideman, M. and Ngindana, M.W. 2004. Website navigation architectures and their effect on website visibility – can search engines deliver on the promise? *Proceedings of SAICSIT 2004 Conference: Fulfilling the Promise of ICT*, October 2004. Stellenbosch, South Africa:292–296.

Weideman, M. 2001. Internet searching as a study aid for Information Technology and Information Systems learners at a tertiary level. Unpublished PhD thesis, University of Cape Town, Cape Town.

Acknowledgement

The author would herewith like to acknowledge Ms Prashini Pillay, for her part with sourcing references and the empirical part of this research project.

ISBN: 978-0-620-48797-9