30th June - 2nd July 1999

To be held at: **The BMW Pavilion, Victoria & Alfred Waterfront, Cape Town, South Africa**

Just want to register? Click here
Check out the programme? Click here

General Host Sponsors Contributions Participation Programme Travel & Accommodation Committees FAQs

Hosted by: **The Department of Information Systems at the University of the Western Cape**

Update notices:

14th June 1999	Provisional programme added
11th June 1999	More abstracts added and index updated
18th May 1999	Index of papers and individual abstracts added
25th March 1999	Details of registration fees and procedures updated

---

Mail us: Admininistration Committee

INTERNET TOPIC SEARCHING MATCHED WITH SEARCH ENGINES TO PROVIDE RELEVANT INFORMATION TO IT/IS STUDENTS – A SOUTH AFRICAN APPROACH.

AUTHOR: Melius Weideman, Cape Technikon, Cape Town, South Africa
melius@ctech.ac.za
P.O. Box 3109, TYGERPARK 7536 SOUTH AFRICA

1 ABSTRACT

Some early research has proven that students show a higher level of participation in a class if they are involved in their own evaluation. It has also been proven that they are keen to use the internet to find information relating to their studies.

A pilot study was used to build a measuring instrument to be used to determine student's ability to search for information. A country-wide research tour followed. The instrument was then used to measure various parameters involving the degree of success experienced by students during internet searching.

It was found that only 38% of the subjects found the information they were looking for within 30 minutes in a controlled environment. There was also strong evidence that historically disadvantaged students had a lower searching success rate than others.

Early versions of a model to assist students in the searching process have shown promise of success.

2 INTRODUCTION

2.1 AIM

The aim of this paper is to disseminate the results of research done in South Africa, which involved computer study students searching for subject information on the internet.

2.2 BACKGROUND

Information is critical in any organisation (Evans & Wurster 1997:72). IT/IS professionals depend for their survival on the use of information in their daily lives. IS managers consider the process of letting users find information for themselves as liberating to IS staff (Foley 1996:45).

The research described in this paper focuses on the retrieval of timely, accurate and relevant information – clearly a crucial component in the lives of would-be IT/IS professionals.

The use of a library has always been fundamental in any study or research, and will probably remain so. However, rapid advancements in the IT/IS field have made it difficult for libraries to keep up with the constant demand for new texts in these fields. Budget constraints and updating logistics compounded this problem. It was a natural move to merge the function of a library and that of the internet as resource providers. Radcliff (1993:17) viewed the incorporation of the internet into the service offered by a library as a challenge. Furner-Hines (1995:31) found that the largest application for hypertext systems in academic libraries in the UK is for networked document retrieval. This indicates that this merge has been taking place.

However, it was found that librarians, traditionally a certain source of answers for the information-seeking learner, experience problems in dealing with the internet as medium. Subject librarians do not always find what they need using the searching facilities on the internet. Basu (1995:38) has shown that 29,6% of a study group of reference librarians found the information they needed between 21% and 30% of the time, while only 14,8% of them found it more than 50% of the time. Problems associated with the use of this method to find information include lack of time (Basu (1995:38)) and the quality and quantity of data returned (Basu (1995:39)). Other studies have shown that some librarians are unable to supply answers to the questions put to them due to a lack of sophistication on their side (Devlin & Burke 1997:104).

The author is involved in full-time lecturing of subjects including Systems Software, Operating Systems and Computer Security at the Cape Technikon. Over the years he has attempted to adjust lecturing methods and implement new ideas in an attempt to make the learning process of the student more successful. Some of these ideas involve alternate resources to libraries, and are described below.

2.3 EXPERIMENT 1

An early experiment was launched in 1993, where students were given some advice on how to and how not to set multiple choice questions. Three weeks before an official test in the subject, they were asked to set up and submit their own multiple choice questions, covering all work to be evaluated in that test.

The students were promised that the lecturer would evaluate these questions, and include the ones he considered to be "good" questions in the test. This implied that if a student took part in this experiment, he would stand a chance of having to answer some questions in a test, which he has set himself!

The response was overwhelming, compared to traditional class participation. Although only approximately 50% of the submitted questions were eventually included in the test (after some minor grammatical and other changes), it proved that more students partake in a class when their own evaluation is at stake.

2.4 THE INTERNET

The internet is a large source of information. Searching for information on the internet has grown into an art which allows the user to run complex state-based searches for information (Lloyd 1996:23). However, Tate (1996:49) states that the quality of information on the Web is suspect due to the absence of editors and fact checkers. It is also claimed that some programs do make it easy for the user to find relevant data (Dragan 1997:192), but their real value will only be known with time.

The lack of structure of resources on the internet is noted (Devlin & Burke 1997:102, Sha 1995:467). Furthermore, the presence of information on the medium of the internet does not guarantee success in finding it.

As the internet became more accessible in the Cape Technikon's computer labs, a new problem cropped up: aimless surfing, chatting and e-mail. The author found that many students did not do the practical work expected of them in the labs, even during "supervised practicals". However, they appeared to be experts on the use of the internet without ever having received official tuition on the use of the internet. The attraction of colourful websites with animation and sound, screens full of interesting information and the ever-present pornography, made programming and other necessary pastimes disappear in the background. The information age has arrived, bringing with it another educational problem to be solved.

It was considered necessary to bring student's attention back to the subject matter using intrinsic motivation. Can we harness the continuous motion of the waves in the sea to provide useful energy to mankind? Can we harness the continuous aimless surfing in the computer lab to provide useful information to the student?

2.5 EXPERIMENT 2

Another experiment was done, based on the positive outcomes of the earlier one. The objective was to direct the time and energy students spend on aimless surfing into providing reference material for the subject they were studying.

Students were challenged during 1998 with a scheme, which required them to "officially" use the internet to find information. If successful, a reward in the form of recognition of their skills was offered implicitly. Initially the students were given basic background on the internet, and on searching techniques and URL's (approximately 1 hour each). After having covered certain topics of a subject using traditional resources (handbooks, notes, transparencies, etc), a 2-hour session would be arranged in a laboratory. A list of topics was given (based on the work covered the preceding few weeks), and students had to search using any search engine and any technique to find relevant information.

The author circulated through the class, and was called to a workstation every time a student found a website, which he considered to contain relevant information. The author then evaluated the site. If it was considered to be unacceptable, a short explanation of why not was given, and the student would carry on searching. If the site did contain relevant information, the lecturer copied the URL onto a diskette, and added a short description about how the contents are to be interpreted (afterwards).

The student then carried on with the next topic, having enjoyed the reward of his website being recognized as acceptable by the lecturer. This implied that the student has mastered a certain level of searching skills and general internet awareness.

After approximately 1,5 hours, the author copied the file containing the URL's and comments onto the student's network, where all students could access it. They were instructed to copy this file, and visit each site listed before the end of the lab session. Although this sounds to be impossible to do in the time allowed, each URL simply had to be copied and not typed into the browser's address field (using the clipboard.) Furthermore, some of these sites were still in the proxy server's cache, resulting in faster download times than the original access to that site. The information found at the relevant website then had to be copied to their own diskette or network directory, or printed, thereby ensuring that the same information was found by all concerned.

The lecturer also printed this information directly after class, to ensure that the correct version was to be used in subsequent evaluations. It was made clear at the beginning that a student had to be present in the class. They could not expect the information having been identified as relevant to be available days, weeks and certainly not months after the date of identification.

In general this experiment yielded positive results in terms of student learning. It was found that students enjoyed applying their (otherwise ignored by lecturers) internet skills to do something considered to be useful by the lecturer! The information found by the students eventually formed part of the study material, and subsequent tests and the final examination included references to it.

2.6 SUMMARY

A summary of what has transpired so far follows.

2.6.1 Students approve of being part of decisions which involve their evaluation.

2.6.2 The presence of internet access exercises a very strong force on the attention of students, away from the work at hand.

2.6.3 Students readily uses the internet for academic purposes if a reward is given which enhances their standing in a group.

3 PILOT STUDY

It was considered necessary to investigate the possibility of "officially" using the internet as a study aid. In the process the results of the previous informal experiments could be used to the student's advantage.

3.1 MEASURING INSTRUMENT

The author designed an instrument to measure the amount of success achieved by a student when searching for subject-relevant information on the internet.

This instrument consisted of a form, which each student had to fill in, indicating background data such as date, name, course, etc.

They also had to identify a traditional resource (handbook, lecture, notes, etc) which did not supply them with all the information they needed to master a part of the work they were studying. They then had to describe the information, which, if they could find it, would solve this problem. Finally, they had to search the internet for this information and indicate whether or not it was found within a set time.

A pilot study was done during July and August 1998, whereby 109 students from 2 institutes were used to test and refine this measuring instrument. The detail of the instrument was changed a number of times during this phase, as its weak points became apparent. The actual searching results from Phase 1 were not used in subsequent statistics, since it only served the purpose of providing initial feedback to refine the measuring instrument.

3.2 RESULTS

The results from this pilot study proved the following:

3.2.1 Students generally do not read instructions, and sometimes give irrelevant answers as a result.

3.2.2 The halo effect causes students to give an answer, which they think is correct, rather than the correct answer. For example, some students answered "Yes, I found the information" when the fields they filled in on the form indicated that they did not find any relevant websites. This could possibly be due to the fact that they want to please rather than to admit that they did not find the information.

3.2.3 Bandwidth limitations and slow PC's are major problems in typical computer labs, even when just 20 students search for information simultaneously

3.2.4 Some students failed to describe, in their home language, what they wanted to search for. As a result, they were unsure about whether or not they found it!

These results were used to refine the instrument, eg by omitting the subject's name to counter the halo effect, and by stressing the specification more, before applying it on the nationwide experiment.

4 STUDY

4.1 RESEARCH TOUR

The actual study was done by visiting a number of institutes across the country, and doing the experiment with as many students as possible (Phase 2). A countrywide trip was undertaken between 5 October and 17 November 1998. The prerequisites for students taking part in this experiment were clearly stated to the coordinator at each institute beforehand as being:

The participating students must be doing a course that involves IT/IS subjects, they must have internet access, and they must have worked on the internet before. The year of study and the skill level on the use of the internet was not specified.

4.2 INSTITUTES

The following institutes were visited:

University of Durban-Westville (Durban)
University of Natal (Pietermaritzburg)
Vista University (Pretoria)
Rand Afrikaans University (Johannesburg)
Vista University (Port Elizabeth)
Vista University (Bloemfontein)
University of the Orange Free State (Bloemfontein)

Technikon Natal (Durban)
Technikon Witwatersrand (Johannesburg)
PE Technikon (Port Elizabeth)
Technikon Free State (Bloemfontein)
Cape Technikon (Cape Town)

Universities and Technikons together form the major force in the post-schooling education in South Africa. Universities provide theoretical, in-depth background (typically 3-year Computer Science degrees) while Technikons offer practice-orientated courses (typically 3-year Information Technology Diplomas).

Both traditionally disadvantaged institutes (eg. Vista Universities, Durban-Westville) and historically strong, well funded institutes (eg. Rand Afrikaans University, University of the Orange Free State) were included.

4.3 TIMING

Since work was done on the internet, containing everchanging information, the experiments were done over a relatively short period of time. The actual time of day of the experiment was also kept to within a few hours of being the same, thereby not having a large influence internet search times.

The experiment was done in 2 separate sessions. The first one was untimed, and students were allowed as much time as they needed to complete it. A typical time was about 20 minutes. During this session the purpose of the experiment was explained, and they were guided through filling in the first page of the 2-page document. Information required here included date, city, course and subjects enrolled for. They also had to identify a traditional resource (handbook, lecture, notes, etc) which did not supply them with all the information they needed to master a part of the work they were studying. They then had to describe the information which, if they could find it, would solve this problem.

The second session was timed at 30 minutes, and involved the students doing an internet search for the information they have just specified. No background on the internet or searching was given. Finally, they had to indicate whether or not the information was found, as well as the time it took.

There are a number of motivations why the time limit of 30 minutes was chosen. Firstly, in a typical educational institute, lecturing times are between 40 and 60 minutes each (this was confirmed during the various visits to the other institutes). It does, however, take a number of minutes to arrive in a classroom, log on, take register, etc. Secondly, it was considered to be the maximum time a person would be prepared to spend on finding one piece of information. Educational libraries are mostly on campus, and given say 45 minutes to find information, walking to a library and asking somebody who is an expert on finding information could be a more attractive choice.

4.4 STUDENTS

A total of 293 students were involved in the experiment. They were from both sexes, and from all the major races in the country. A number of different courses were covered.

5 RESULTS

A total of 293 students (excluding the 109 of Phase 1) from four race groups (both genders) at 12 different institutes following 13 different courses were involved in the research tour.

Student ages varied between 17 and 42 years, with an average of 21.56 years. The spread between genders were 32.3% female, 67.7% male. A question was posed: “How many hours of formal instruction have you received on the use of the internet?” More than 75% answered “Less than one hour”.

A total of 46% of the students use the internet at least once per day. Furthermore, 82% have used a search engine to find information before. The subject: “Computer and Information Technology” appears to be creating most problems for students –74% of the respondents who are enrolled for it have identified it to pose a problem. Figures for other subjects include 55.2% (Technical Programming), 48.3% (Computer Science), 36.3% (Information Systems), 31% (Informatics), 27.3% (Business Information Systems), 26.3% (Psychology), 25% (Chemistry), 24.2% (System Software) and 23.1% (Education).

A total of 46.4% of the participants indicated that their handbooks had the most gaps in the information they provide, while traditional lectures came second with 30.3%. Videos and libraries posed the smallest problem. When asked to choose any search engine to use in the final step of the experiment, four out of every ten students chose Yahoo – a far greater figure than any other search engine.

The number of hits listed before students started visiting websites varied between 1 and 53 469 078. The number of sites visited within the allotted 30 minutes varied between 1 and 42. Almost 25% of the students spent the full 30 minutes on the search, including successful and unsuccessful candidates. The average time spent before finding the information or giving up before it was found was 20.7 minutes.

If the Unknown group's figure is omitted, only 38% of the students found the information they identified initially, within the allotted time.

6 SOUTH AFRICAN PERSPECTIVE

Although some of the issues below could be be regarded as being sensitive (eg race vs. performance), this paper is about the South African perspective, and the issues are considered to be important and worth inspecting.

6.1 RACE vs SEARCHING SUCCESS

The objective here is to see if there is a relation between the race of an individual and the success achieved during searching. The three possible answers to the question: "Did you find the information you were looking for?" are summarized in the table below.

	NO	YES	UNKNOWN
African	**76**	**28**	**34**
Coloured	**4**	**4**	
Asian	**18**	**12**	
Unknown	**5**	**2**	
White	**55**	**51**	**4**
TOTAL		**293**	

Table 1 Race vs Searching Success (Total)

Since the population size of the Coloured group was insignificant (8), and since the Unknown group does not contribute any insight, they were both omitted for statistical purposes.

In the next table, the number of YES answers per remaining race group is expressed as a percentage of the size of the group. Therefore a Success % figure lower than 38% indicates below average searching abilities for that group, and a figure higher than 38% indicates above average searching capabilities for that group.

	NO	YES	SUCCESS %
African	**76**	**28**	**26.9**
Asian	**18**	**12**	**40**
White	**55**	**51**	**48.1**

Table 2 Race vs Searching Success (Summary)

It is clear that the African group has a YES percentage well below the average, the Asian group is approximately on average and the White group is well above the average. These results suggest that the African group, probably as a result of racial discrimination in education in the past, is still suffering from the results. The White group, on the other hand, indicates an abnormally high number of YES answers, most likely as a result of having received a better education based on race.

6.2 CHOICE OF SEARCH ENIGINE

Some authors have inspected the use of different types of search engines. Brandt (1996:35) has shown that comprehensive search engines yield a larger degree of recall, while subject-based categories tend to have a large degree of precision. This fact could affect the way in which learners approach the process of searching for relevant information in their subject field.

Experiments have been done (Clarke 1997:187) to compare the recall performance, precision and coverage of three established search engines. It was found that a noticeable degree of overlap did exist amongst the output of the different programs. However, a method was found to perform an accurate estimation of recall and precision of Web search engines.

In contrast, it has been proven in earlier works (Su 1992:514) that the success of a search as judged by the user depends more on absolute recall than on precision.

Tate (1996:55) proposes a strategy to help Web users to critically evaluate Web resources for accuracy, authority, objectivity, currency and coverage.

It has been shown (Taubes 1995:1355) that the processing load imposed by spiders on internet file servers can be damaging. Some spiders attempt to download the entire website being inspected, onto the machine attempting to categorise the site. If the site is a typical research site, with large amounts of data stored, it imposes an unmanageable load on the server being inspected, and on the communication links.

A closer inspection on the use of search engines by the participants in this study will now be made.

6.2.1 The objective here is to see if there is a relation between the search engine chosen by the individual and the success he achieved in the searching process. The number of individuals choosing a search engine, plus the number of successes found with this choice are given in the table below.

SEARCH ENGINE	TOTAL CHOSEN	TOTAL FOUND
AARDVARK	1	1
ALTAVISTA	23	8
ANANZI	10	6
EXCITE	2	1
GOTO	1	0
HOTBOT	4	3
IBM	1	0
INFOSEEK	22	9
LYCOS	3	0
MAMMA	1	0
METACRAWLER	5	2
MICROSOFT	2	0
NETSCAPE	9	4
SEARCH	5	2
WEBCRAWLER	13	4
WEBFERRET	1	1
WEBOPEDIA	1	1
YAHOO	121	47

Table 3 Search Engine choice vs Searching Success (Total)

All the search engines with a population size of less than ten in the table above were omitted, and the % success rate for each remaining search engine is given in the next table.

SEARCH ENGINE	TOTAL CHOSEN	TOTAL FOUND	% SUCCESS
ALTAVISTA	23	8	34.8
ANANZI	10	6	60
INFOSEEK	22	9	40.9
WEBCRAWLER	13	4	30.8
YAHOO	121	47	38.8

Table 4 Search Engine choice vs Searching Success (Summary)

The average % success achieved by using the five search engines above (41%) is close to the overall success rate (38%). However, there is a significant variation between the individual success rates. It is interesting to note that the only South African search engine in the list (Ananzi), has by far the highest success rate, although only 10 subjects used it - the lowest number of all the remaining search engines! It would be naïve to suggest that these figures indicate search engine performance. Altavista is considered to have the largest index of all search engines (approximately 150 million web pages), but it produced a below average success rate.

These results seem to suggest that subjects are more familiar with local products than with overseas ones.

The abnormally high number of subjects (121, ie one out of every 4 subjects, which is far above the second most popular choice) choosing Yahoo as a search engine warrants closer inspection.

6.2.2 It was found that the average age of the Yahoo users was 20.8 years, as opposed to 22.1 of the non-Yahoo users. This would indicate that the “catchy” name and image of Yahoo appeals more to younger people.

6.2.3 Furthermore, 39,5% of female subjects chose Yahoo, while 42,9% of male subjects did so. It is considered that gender did not play a significant role in the choice of this search engine.

6.2.4 Considering the race of Yahooers: 100% of the Coloureds, 55% of the Whites, 33% of the Asians and 28% of the African subjects chose it.

6.3 INSTITUTE vs SEARCHING SUCCESS

An inspection was done into the % success achieved by the searchers of the various institutes. It was considered necessary to protect their identities, and it will thus not be divulged.

The two top performers with 43% and 40% success respectively, were both traditionally white and old Universities. The two worst performers (13% and 5% respectively) were both Technikons with mostly Asian and Black students respectively. Once again, the effects of the racial discrimination in South African education are evident in the ability or lack thereof, of students searching for information.

7 CONCLUSION

7.1 The low searching success rate of 38% has proven that without a researched, tested and proven model to guide them, students cannot be expected to make use of the internet as an alternative information source to traditional resources.

7.2 Furthermore: the bandwidth and the general level of computing facilities in laboratories will both have to increase to make the successful implementation of serious internet searching possible.

7.3 The South African history of racial discrimination has left evidence of discrepancies in the abilities of some learners to use the internet as a study aid.

At the time of reading this paper (June/July 1999), the model referred to above will be 90% complete. Testing early versions of this model on a worldwide audience of students has proven that it shows promise of success. Final figures to indicate its success or failure should be available before the end of 1999. The model will also be available on the internet either as a standalone product or as part of an existing search engine.

8 ACKNOWLDEGEMENTS

8.1 Cape Technikon (study leave, financial assistance)
8.2 Acer Africa (Pty) Ltd (financial assistance)
8.3 Hewlett Packard SA (financial assistance)
8.4 Council for Scientific Development (financial assistance)
8.5 Various colleagues at the institutes visited for arrangements and time made available to the author.

9 REFERENCES

Basu, G. 1995. **Using Internet for Reference: Myths vs. Realities.** Computers in Libraries. 15(2):38-39.

Brandt, D.S. 1996. **Relevancy and Searching the Internet**. Computers in Libraries. Volume 16(8).

Clarke,S.J., Willett, P. 1997. **Estimating the recall performance of Web search engines.** Aslib Proceedings, Volume 49 No 7. August. 184 – 189.

Collins, S. 1997. **Seek and ye shall find.** .net January. 50-66.

Devlin, B, Burke, M. 1997. **Internet: the ultimate reference tool?** Internet Research. 7(2):101-108.

Dragan, R.V., Boscardin, A. 1997. **Taming the wild, wild Web.** PC Magazine. March 25. 192.

Evans, P.B., Wurster, T.S. 1997. **Strategy and the new Economics of Information.** Harvard Business Review. September/October. 71-82.

Foley, M.J. 1996. **Hot new Database Technologies.** Datamation. September. 44-49.

Furner-Hines, J, Willett, P. 1995. **The use of the World-Wide Web in UK academis libraries.** Aslib proceedings. Volume 47 No. 1. 23 – 32.

Lloyd, L. 1996. **The Ovid Web Gateway creates sophisticated search opportunities on the World Wide Web.** Campus-Wide Information Systems 13(3):23-24.

Radcliff, C, Du Mont, M, Gatten, J. 1993. **Internet and Reference Services: Implications for Academic Libraries.** Library review. Volume 42 No 1. 15 – 19.

Sha, V.T. 1995. **Cataloguing Internet resources: the library approach.** The Electronic Library. Volume 13, No 5, October. 467 – 476.

Su, L.T. 1992. **Evaluation Measures for Interactive Information Retrieval.** Information Processing & Management. Volume 24 No. 4. 503 – 516.

Tate, M, Alexander, J. **Teaching Critical Evaluation Skills for World Wide Web Resources.** Computers in Libraries. Volume 16. 49 – 55.

Taubes, G. 1995. **Indexing the Internet.** Science. Volume 269. September 8. 1354 - 1356.

Venditto, G. 1996. **Search Engine Showdown.** Internet World. May. 79-80.

Wilson, DL. 1995. **Internet Users Face Lines at Computers.** The Chronicle of Higer Education. December 1. A31-A36.

http://www.altavista.com. Alta Vista Search Engine.

http://www.excite.com. Excite Search Engine.

http://www.hotbot.com. Hotbot Search Engine.

http://www.infoseek.com. Infoseek Search Engine.

http://www.yahoo.com. Yahoo! Search Engine.

http://www.searchenginewatch.com. Search Engine Reports.