FOIOTI: An implementation of the conceptualist approach to Internet Information Retrieval

M. Weideman[1]
Cape Peninsula University of Technology, Cape Town, South Africa
meliusw@yahoo.com

The objective of this research project was to evaluate searching methodologies used by undergraduate learners in searching for academic information, and to design an aid if required. Literature surveys indicated that the sheer size of the Internet and lack of categorization of the information available makes finding relevant information a daunting task. Other problems include a lack of clear search specification formulation and inefficient usage of time and computing power by loading and using one search engine at a time. It was also clear from the literature that Internet searchers have difficulty in general to locate relevant information. The methodology used included empirical experiments involving a total of 1109 learners in a series of empirical experiments to address this situation. Their failure/success, methodology and a number of other factors were measured, and an instrument was designed to overcome these problems. The main conclusion was that the use of this instrument (called FOIOTI: Finder Of Information On The Internet) increased the chances of success under controlled circumstances dramatically. This was achieved by hiding the operationalist detail from the user, allowing him/her to concentrate on conceptualizing the topic.

Keywords: Internet searching, Information retrieval, information interface, search engine, FOIOTI.

1 Introduction

The Internet offers its users the world's largest and most complex, chaotic and unstructured search space (Sherman 1999: 54). Although programs such as friendly browsers and free search engines exist to assist the user to find his/her way around this unknown data repository, general consensus exists that navigating the Internet is not a straightforward task (Voorbij 1999: 598). One specific skill, which eludes many of these average users, is the finding of relevant information on the Internet in a short time (Brewer 2001: 54).

An Internet search engine is a program, which allows the user to either specify (a) keyword(s), or drill down into a topic through various levels of directories. Both approaches should lead the user to 'the perfect website', which will contain the exact information required. Commonly used search engines include Google, AllTheWeb, Yahoo! and MSN Search (Sullivan 2003). These search engines each contain a front end, an index and a set of collectors. The front end is the human interface, i.e., what the user interacts with on the screen during the searching process. The index is a large file which contains detail about millions of websites, and it is this file (not the Internet itself) which is queried when a user searches for information via a search engine. The collectors are either human editors (as with Yahoo) or automated programs called spiders, bots or crawlers (as used by Google and AltaVista). Both types of collectors gather information about available websites, and build it into the index file (Sullivan 1999: 34, Sullivan 2002).

All search engines offer a number of features to enable the user to find information with ease. However, it has been proven that most users do not avail themselves of these features (Weideman 2001: 197). The purpose of this article is to introduce the reader to an alternative approach to Internet searching - one that does not presuppose any knowledge of search engines or their syntax and operators. This approach is aimed at the typical IS/IT (Information Systems / Information Technology) university or technikon student who is not familiar with the detailed syntax of search engines.

The most basic form of Internet searching (inherited from pre-Internet systems) is to load a search engine, type in a single word, and instruct the search engine to find the information. Not surprisingly, this method seldom delivers relevant answers, especially if the word has many different interpretations (Siegfried *et al.* 1993: 273). As an example, common words like 'religion,' 'computers,' 'sport' and 'weather' produced the following approximate numbers of answers in four separate searches on Google: 12 000 000, 20 200 000, 23 100 000 and 28 000 000.

A logical next step would be to use either a very specific single word, or more than one word as a search query. Multiple-word queries fall into one of a number of categories: phrase searching, Boolean operators, inclusion/exclusion, combinational operators, and other methods.

---

1. Prof. Melius Weideman is Head: Research Planning and Capacity Building, Cape Peninsula University of Technology, Cape Town, South Africa.

Another method which focuses results, is the use of the inclusion (+) and exclusion (-) operators. By using these operators, the presence or absence of certain words can be enforced. The information need of a user is clear from the following search query:

> +surf +board +specs -Internet -screen -website.

As a user progresses from simple to more advanced searching, operators can be combined to form very effective filters to produce a strongly focused search specification, for example:

> (+Mathematics +formula) OR (+science +formula) AND 'volume of a sphere'

However, a recent study claimed to have found that '... most people use few search terms, few modified queries, view few Web pages, and rarely use advanced search features' (Spink *et al.* 2001: 226).

2 Literature review

2.1 General searching
Lancaster made a prediction about the way in which humans and computers would interact in 1978:

> 'The scientist of the year 2000 will use a terminal in many different ways: to receive, transmit, compose and search for text, seek answers to factual questions, build files and converse with colleagues. The terminal provides a single entry to a wide range of capabilities that will substitute, wholly or in part, for many activities now handled in different ways' (Lancaster 1978: 176).

Lancaster's 'scientist' is today's average PC (Personal Computer) user, his 'terminal' has become the desktop computer and 'conversing' has evolved into e-mail. The 'different ways' could be the shift from using an intermediate searching expert, as was done in the traditional library environment, to doing the search oneself. This moves the focus of this project to Internet search engines.

Schwartz stated that there could be a change towards smaller but better resources operating in a richer environment in the future (Schwartz 1998: 973), while Feldman claimed that the quality of results appeared to increase from one year to the next (Feldman 2000). However, a claim is made that search engines are complex, trusted without being understood and that users simply deal with their answers without understanding why they receive those answers (Lynch 2001: 17). These disparate views require further investigation.

2.2 Search engines and operators
The design of search engine interfaces was criticised by Large and Beheshti, after having done some searching experiments with 50 grade six schoolchildren. They believed that long lists of irrelevant answers could be suppressed to enhance a child's experience of searching the WWW (World Wide Web) for relevant information (Large and Beheshti 2000: 1079).

The syntax of search engines differs widely, and a wide variety of operators exists to allow focusing of a search query. Lancaster referred to natural language searching, where no operators are used, in an early reference (Lancaster 1978: 279). As an alternate, commonly used operators include phrases (using quotes), the inclusion operator (+) and the exclusion operator (-). Other, more advanced and less often used ones include Boolean operators (AND, OR, NOT), proximity operators (NEAR), stemming (*) and field limiters (DOMAIN, TITLE). These operators are discussed by a variety of authors (Boulton 2002, Habib and Balliot 1999, Hock 1999:25, Notess 1997: 65, Notess 1999a: 64,65, Sullivan 1999: 38).

However, Jansen and Pooch quote a number of other studies who claim a low percentage of real life searches having included Boolean operators: the Fireball study (less than 3% - done in 1998), the Webcrawler study (12% - done in 1997), Infoseek (10% - done in 1998) and Excite (29% - done in 2000) (Jansen and Pooch 2001: 237, 238).

During a survey of 316 Excite users, Spink, Bateman and Jansen stated that few users employed logical operators and even fewer used the syntax correctly. They also had problems with search phrases and with the construction of 'good search terms' and 'complex search queries' (Spink *et al.* 1999: 123, 125).

It can thus be concluded that the best way to construct a search query during an Internet search is not obvious to the average searcher. Where some search engines would add an implied OR operator between words (Infoseek), others will insert an AND (AltaVista and Google in their basic searching modes), and some might even treat it as a phrase (Looksmart). These three widely differing approaches would produce very different results, which would confuse rather than enlighten an average searcher. An Internet search where five words with AND operators between them are used, will produce x answers from a given database. The same search using the same five terms and the same database, but using OR operators between the words, will produce y answers, where $x << y$.

2.3 Multiple databases

Spink has confirmed that some users (without any advice or training) are doing more than one search on a variety of sources on OPACs (Online Public Access Catalog) and CD-ROM databases for better search results: 57% of a sample of 200 academic end-users conducted multiple search sessions (Spink 1996: 603).

The value of this richness of information is confirmed by Chun quoting Reid in saying that the more pieces of information available to choose from, the higher the chance of finding the correct one (Chun 1999: 137). The fact that those pieces of information could be spread over more than one database, assuming that the databases are all of the same quality, does not affect this basic truth.

However, a breed of search engine called a meta search engine makes this multiple searching possible in one Internet searching session. A meta search engine simply takes the user's query, and submits it to the indices of a number of search engines simultaneously. Once the results have been received, website summaries are combined and one set of answers is produced.

Garman in fact built a case for the use of meta search engines by strongly advising users to search multiple databases to satisfy the same query (Garman 1999: 75). Other authors confirmed the validity of this advice (Jacso 1999a: 227; Jacso 1999b: 99; Mickey 1999: 79; Williams 1996: 246).

2.4 Searching success

Searching success in the context used here can be described as finding information, which satisfies the need, within a reasonable time. Lancaster discussed some factors which influence success in online searching many years before the advent of Internet search engines (Lancaster 1978: 331). However, lately many authors agree that learners should know how to retrieve information from electronic sources (Cronje and Clarke 1999: 215; De Jager and Sayed 1998: 197; Edling 2000: 11; Gan 1998: 29; Marais and Marais 1999: 85; Wong 1998: 5). IS managers consider the process of letting users (i.e. what the learners of today will become) find information for themselves as liberating to IS staff (Foley 1996: 45).

Some authors claim that, in general, finding relevant information in an information source is not a difficult task:

- '... the majority of Australian academics have a high expectation of success as they engage in information seeking on the Internet' (Bruce 1998: 555).
- '... Australian academic users generally ... are satisfied with information seeking ...' (Bruce 1998: 553).
- '... even the novice information searcher is able to extract relevant documents in rank order of calculated relevance.' (Ross and Wolfram 2000: 949).

However, other authors claim that the opposite is true:

- 'The very size of digital libraries begs the question of how average users will continue to find context and meaning.' (Huwe 1999: 68).
- 'Using Internet search engines to retrieve information can sometimes be a slow process ... and can result in far too many "hits" so that picking out relevant ones can be difficult' (Large et al. 1999: 5).
- 'Our customers (where customers are library users) understand their information needs, but they don't understand the tools that can satisfy them' (O'Leary 2000: 22).
- '...some respondents seemed confused about what they were to report when asked to list query terms for their search' (Spink et al. 1999: 122).
- '...the user's ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low.' (Spink et al. 1999: 129).
- 'Only 33% of the Internet users agree or strongly agree with the statement "It is easy to perform subject searches on the Internet." ' (Voorbij 1999: 604).
- 'I find it difficult to search information on the Internet...' (Voorbij 1999: 605).
- '...information seeking is a complex and difficult process for these students, who seek to reduce the task to finding an obvious answer or finding a good website...' (Wallace et al. 2000: 75).
- '... both novice and experienced searchers were overconfident in their performance' (Wolfram and Dimitroff 1997: 1145).

In conclusion, it became clear from the literature that many users found the formulation of a search query difficult. It also appears as if the general success rate of Internet searching is low. Lancaster proved as far back as 1968 that the low quality of query formulations was the main reason for the failure of a search (Frants and Shapiro 1991: 17).

The following points have emerged as fundamental issues from the literature survey:

- the sheer amount of information available on the Internet is overwhelming,
- many features exist in search engines to help users find relevant data,
- users find it difficult to construct a search query from their information need,
- the use of more than one database seems to be advisable, and
- many authors agree that learners do not appear to be capable searchers.

Definite positive aspects characterize both approaches. An in-depth understanding of the search operators of a few well-known search engines is a formidable weapon in the constant drive to find relevant information on the Internet. Similarly, the ability to manipulate the definition of an information need in order to filter out unwanted answers also goes a long way towards finding the exact information that is required.

FOIOTI was designed to hide the technicalities of search engine operators and the construction of the search query from the user. This 'construction' process is actually a basic form of computer programming - both involve instructions in a non-natural language given to a computer. FOIOTI further simplifies the choice of search engine (which in itself could be a topic for research) to the click of a button.

The user types in the information need, extracts the required components from it, and initiates the search. Since no 'programming' is involved, the user is freed from the technicalities (FOIOTI takes over the operationalist component), and finds it easier to concentrate on the information need (the conceptualist approach).

As an example, assume your information need is: 'I need to find the final score in the 1995 World Cup rugby final.' The operationalist could submit a variety of queries, such as: 'World AND rugby AND final AND 1995 +score,' or '+rugby +score +1995 +final -football,' etc. If unsuccessful, he/she could change to another search engine.

However, the conceptualist might remember a photograph on a newspaper front page, showing Nelson Mandela handing over the cup to Francois Pienaar. These two persons have very little in common: the first is an ex-political prisoner, then became a much-honored (now retired) President, while the second is a rugby captain. By changing the information need to 'I need an article where the names Nelson Mandela and Francois Pienaar appear in close proximity,' the chances of success are increased. A resultant search query like 'Nelson Mandela' NEAR 'Francois Pienaar' AND 1995 could be successful.

Singh further discussed these two concepts, and highlighted the fact that the operational searcher concentrates on high precision searches, by manipulating the system and its operators. The conceptualist chases a higher recall figure by spending more time on concepts and terminology, and generates subsets of his/her results (Singh 2003).

4 Research results

4.1 Population groups

As indicated earlier, an attempt was made to ensure that the major South African population groups would all be represented in the sample. It was practically impossible to mirror the exact distribution of the South African population across the four generally accepted categories in this experiment. However, Graph 1 indicates the actual figures of the race groups as covered by this experiment.
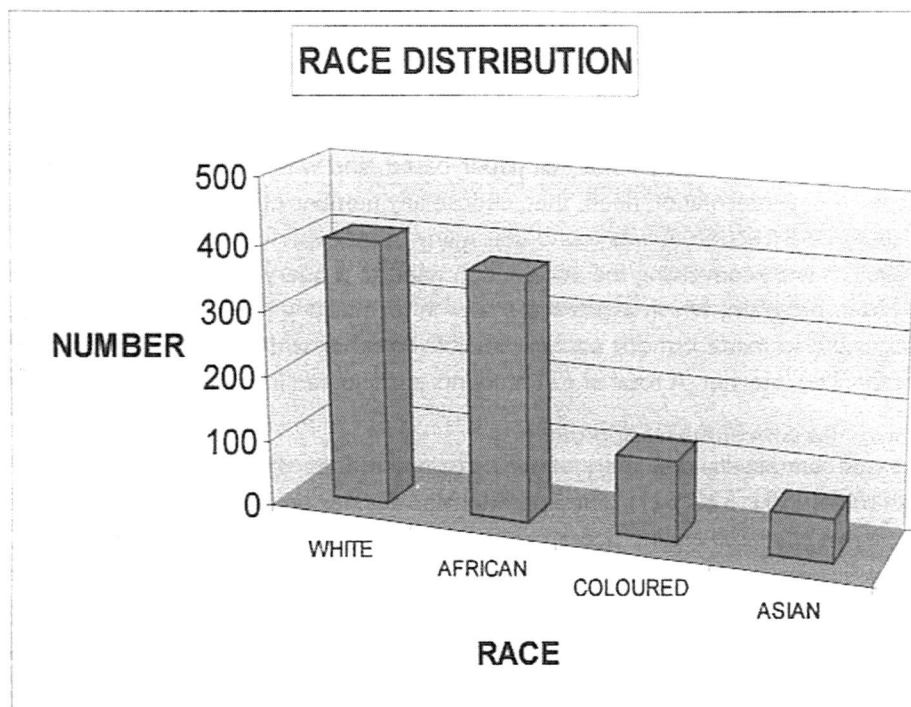
Graph 1. Race Distribution

4.2 Search Engine choice

A decision had to be taken on which search engines were to be included in the final version of the model. A weighting system was used to combine search engines identified as being 'good' choices based on the literature survey with those chosen by the learners.

A total of 20 search engines were selected, based on positive comments made by a variety of authors listed in the literature survey. A score of 10.5 points was allocated to all of them (10.5 is the average of 20 + 19 + 18 + ... + 2 + 1). These scores are based on the assumption that scores between 1 and 20 would have been allocated to them if they were selected on preference, and is called Score 1 (see Table 2).

Secondly, all the search engines chosen by the learners received a score, ranked according to the number of learners choosing them (column marked # USERS 2 in Table 2). Thirdly, since there are 22 search engines in this group, the search engine being chosen by the most learners received a score of 22, then 21, etc. Those with an identical number of votes received identical scores. This figure was called Score 2 in Table 2. The last column indicates the sum of Score 1 and Score 2, where a higher figure indicates a better choice of search engine.

Table 2 Search Engine Choice

SEARCH ENGINE	SCORE 1	# USERS 2	SCORE 2	SC.1 + SC.2
Aardvark	0	1	1	1
All-in-One	10.5	0	0	10.5
AltaVista	10.5	45	21	31.5
Ananzi	0	10	18	18
Ask Jeeves	10.5	1	1	11.5
Dogpile	10.5	0	0	10.5
Excite	10.5	8	16	26.5
Google	0	2	9	9
Goto	10.5	1	1	11.5
Hotbot	10.5	8	16	26.5
IBM	0	1	1	1
Infoseek	10.5	20	19	29.5
Looksmart	10.5	0	0	10.5
Lycos	10.5	6	15	25.5
Magellan	10.5	0	0	10.5
Mamma	10.5	1	1	11.5
Metacrawler	10.5	2	9	19.5
Metasearch	10.5	0	0	10.5
Microsoft	0	3	12	12
Netscape Search	10.5	2	9	19.5
Northern Light	10.5	0	0	10.5
Profusion	10.5	0	0	10.5
Savvysearch	10.5	0	0	10.5
Search	0	5	14	14
Snap	10.5	1	1	11.5
Webcrawler	0	26	20	20
Webferret	0	1	1	1
Webopedia	0	1	1	1
Yahoo!	10.5	153	22	32.5
Zazoo	0	3	12	12

From Table 2 above, all search engines with a total (Score 1 + Score 2) of 20 and higher were chosen to be subjected to further tests. The search engines identified this way are listed in Table 3.

Table 3 Highest Search Engine Scores

SEARCH ENGINE	TOTAL
Yahoo! (YA)	32.5
AltaVista (AV)	31.5
Infoseek (IS)	29.5
Excite (EX)	26.5
Hotbot (HB)	26.5
Lycos (LY)	25.5
Webcrawler (WC)	20

4.3 Search Engine operator choice

A summary was made of the methodology used by all those learners who did not find any information without a model, but did find it by using the model. Since these learners used some of the operators suggested in the initial versions of the model to achieve success, it was assumed that these operators have a higher probability of generating successful results for other searchers as well. Their queries were inspected, and the operators used by these successful searchers are summarized below.

Table 4 Operators used by Searchers

OPERATOR	# OF SEARCHERS
+ (Inclusion)	28
' ' (Phrase)	14
* (Stemming)	4
- (Exclusion)	3
AND (Logical)	1
DOMAIN: (Field)	1
URL: (Field)	1

These findings confirmed those of Spink, Bateman and Jansen as well as Voorbij, who stated that searchers seem to favour simple operators and avoid more complex ones (Spink *et al.* 1999: 123; Voorbij 1999: 603). It was decided to do further tests on these operators, to determine which one(s) produced the most relevant results when used in combination with the search engines listed in Table 3.

4.4 Search Engines/operator experiments

The next step was to combine the use of the search engines identified in Table 3, with the operators identified in Table 4 above. Since this research project is aimed at an audience of real-world users, a real-world testing scenario was preferred above a theoretical one. The author did over 300 detailed Internet searches, recorded the results and a final decision about the operation of FOIOTI was taken. One example taken from these searches is given below. The searches start with a REQUEST each, which is based on a perceived information need (PIN) of a typical IT/IS learner (Mizzaro 1998: 306). A variety of QUERIES were then to be constructed, based on the operators identified in Table 4 above. Half of the queries done were constructed from the popular operators (the first two rows of Table 4: Inclusion and Phrases), and the other half from the least popular operators (the last three rows of Table 4: Logical and Field).

Rows number three and four (Stemming and Exclusion) were considered useful mostly on repeated successive searches on the same topic. For example, a user does a search on Intel processors, using the query: 'Intel' +processors. Only after receiving some incomplete answers, then realizes that the last 's' in the query could filter out wanted answers, then changes the query to 'Intel' +processor*. Similarly, a search for password rules in Windows 2000 using the query: Windows AND 2000 AND password could produce superfluous answers including all other Windows versions. Afterwards the user might adjust the query to: Windows AND 2000 AND password -NT -Me -XP -98 -95 -3.* to exclude unwanted version numbers. For this reason these two operators were excluded from the experiments.

Although the first ten queries were generated by the author, each one's topic was based on query topics of some of the participants up to this point. The author constructed each query to be generic in nature, and did not consider the exact syntax of any specific search engine. This was done to mimic the final version of the model, which would treat the components of the query as submitted by the user the same. Each query would be submitted to each one of the search engines identified in Table 3 above.

The first 10 results of each of the searches would be inspected in each case, since various authors noted that most users view only the first few results (Courtois and Berry 1999: 44; Notes 1999b: 84). A decision would then be taken about whether or not a relevant answer to the query was found amongst these ten websites. The author's own relevance judgement was considered reliable enough to make this judgement, since all the REQUESTS are in his field of knowledge.

Details about the interpretation of the table and QUERIES below follow.

- The first row in each table contains an abbreviation for each search engine listed in Table 3 above.
- The first column in each table lists the queries described directly above that table.
- The numbers in the other columns indicate the number of answers received. A lowercase t indicates a time-out error given by the browser, with no figure as a result. A lowercase x indicates an error message from the search engine concerning the query syntax, with no figure as a result. The figures in bold include at least one relevant site under the first 10 (i.e. success), as judged by the author.
- Queries based on the Inclusion and Phrase operators are:
  Q1.3, 1.5, 2.3, 2.5, 3.3, 3.5, 4.3, 4.5, 5.3, 5.5, 6.1, 6.3, 7.1, 7.3, 8.1, 8.3, 9.1, 9.3, 10.1, 10.3 plus all queries from 11 to 20 both included.
- Queries based on Logical and Field operators are all those not listed above.

Only one of the 20 queries done is listed as an example below.

REQUEST 4

I need a description of the operation of the Melissa computer virus.

QUERY 4.1
TITLE:Virus AND Melissa +description AND 'Melissa'

QUERY 4.2
TITLE:Virus OR Melissa +description OR 'Melissa'

QUERY 4.3
'Melissa' +computer +virus +description

QUERY 4.4
(TITLE:Virus OR URL:Virus) AND (+Melissa +computer +virus +description)

QUERY 4.5
+'Melissa' +'computer virus' +description

Table 5 Results from Request 4

	AV	IS	YA	LY	HB	WC	EX
Q4.1	**720 909**	0	**1**	0	**290**	x	x
Q4.2	254 893	95 969	**1**	20	82	x	x
Q4.3	26	**1 539**	418	0	**1 040**	**308**	**9 438**
Q4.4	**374**	0	**9**	0	0	x	x
Q4.5	0	**18**	**11**	0	**1 040**	0	**331**

The two least successful search engines (Yahoo! and Lycos) were subsequently dropped from the last stage of experimentation. Furthermore, the successes (bold figures) per search engine were added, producing the following final number of successes:

Table 6 Search Engine Success

	AV	IS	HB	WC	EX
# OF SUCCESSES	3	3	7	4	7

The results of these 20 sets of searches showed that the use of simple phrases combined with the inclusion operator, for the three chosen search engines (identified below), consistently provided a higher success rate. For example, Q4.3 and Q4.5 (see Table 5) each has a relatively high number of Bold figures, which is an indication of success rate. This trend was also evident in the remaining queries. It was therefore concluded that the operation of FOIOTI would be based on the following:

- overcoming common errors identified in searching strategy,
- the most successful operators, i.e. inclusion and phrases, and
- the three most successful search engines, when used in combination with these operators, i.e. Hotbot, Webcrawler

and Excite (Table 6).

NOTE: Subsequent to the completion of these experiments, the Webcrawler search engine has been replaced in FOIOTI by Google, as part of a further refinement experiment.

FOIOTI's actual operation consists of the following simple algorithm:

- ignore contents of the specification box (it served its purpose to force the user to think clearly about the information need)
- surround contents of caps and phrase boxes with quotes to preserve capitalization and sequence
- concatenate the last four boxes in a sequence of keywords
- precede each one of the four components with an inclusion sign.

For example, assume a user completes the FOIOTI boxes as follows:

CAPS BOX:	Turkey
PHRASE BOX:	death toll
WORD BOX:	earthquake
WORD BOX:	1999

FOIOTI will then produce the query, and submit it to the search engine, from these boxes as follows:

+'Turkey' +'death toll' +earthquake +1999

FOIOTI's design is therefore based on Fidel's statement that '… analyzing users' seeking and searching behavior as it occurs in actual situations is a promising method for … suggesting improvements in system design and in search environments' (Fidel et al. 1999: 36).

4.5 FOIOTI usage

FOIOTI is available at *www.mwe.co.za*. Follow the links from the homepage (Java must be enabled on the browser to make FOIOTI visible, and Ctrl should be held down while clicking to disable any pop-up blockers).

The success of this model is based on the fact that the user is shielded from search engine syntax and other technicalities. The only two prerequisites are a clear understanding of exactly what the search topic is, and basic mastery of the English language. The reader is reminded that FOIOTI was designed for 'the average' undergraduate IS/IT learner.

The large gray FOIOTI box (see Figure 1) requests that the user must complete all five rectangular white boxes, before clicking on any one (or more) of three search engine buttons at the bottom. It will then construct a query, using the information just supplied, load the chosen search engine(s), and feed the query to it. The result screens are presented to the user in the same way had he/she simply accessed the search engine the normal way (Weideman 2002a).

The first box requests the user to type out the information need in a full English sentence. This is a crucial first step, since it forces him/her to think carefully about what the information need actually is. Specific words and phrases should be used, for example 'BMW' rather than 'motorcar,' 'standard marathon' rather than 'jogging,' etc. Although this first box plays no part in FOIOTI's final construction of the search query, it does force the user to think about the topic formulation. It also produces the word(s) required by the other four boxes. If the user has trouble in completing the information requested in the remaining four boxes, it normally implies that the information need in the first box is incomplete. The user may at any time simply come back to this box, add more specific terms/phrases, update the other four boxes, and try again.

The second box requires that a word be entered (it could be more than one) which is normally spelt using (a) capital letter(s). First names and abbreviations immediately come to mind. Users should take care to type this word the way it is used in everyday terms, i.e. 'MICROSOFT' is correctly spelt, but 'Microsoft' is a more common way of using this word. Other examples: FBI, DVD, Monaco and Charles. FOIOTI preserves the capitalization of this box before the query is submitted. If no caps words can be deduced from the spec. box, any lowercase meaningful word could be used (stop words such as: the, that, in, information, computer, etc should be avoided).

The third box requests a phrase. FOIOTI's definition of a phrase is: 'Two or more words which have been copied from a grammatically correct, English sentence (in the Spec box) without disturbing the order.' Users should avoid general phrases such as 'information about computers,' 'John Smith,' 'world record,' etc. Better examples are 'lower Manhattan,' 'Tour de France' and 'UK economic indicators.' An important feature of the contents of this box is that it can be used effectively to focus the scope of a search. FOIOTI preserves the sequence, capitalization and spelling of this phrase in the search query.

The phrase should be used to widen or narrow the scope of the search. If too many answers are received, the phrase should be lengthened. Similarly, if zero answers are produced while using a six-word phrase, cut off one word at a time to broaden the focus, which should produce more answers. The following four phrases, as an example, could be used as part of a complete FOIOTI specification to successively apply a sharper focus to a search:

- 'teaching materials'
- 'teaching materials for English'
- 'teaching materials for English literature'

- 'teaching materials for Shakespearean English literature'

The final two boxes require any two words related to the topic to be typed in. Again, an attempt should be made to refrain from using general terms (such as 'bicycle' or 'President') - substitute with specific words (such as 'Cannondale' and 'Putin'). Once all five boxes have been completed, the user should now click on one or more of the search engine buttons at the bottom of the gray box.

Describe EXACTLY what you are looking for by typing one sentence in the spec box (a single word or even a short phrase is not goor enough!)

I need a description of the 7 layer OSI networking model — **Spec box**

Type one word from the spec box which is)
normally spelt with () capital(s)
OSI — **Caps box**

Type two or three words from the spec box,
which have a special meaning if they
are used together. (one space between each word)
networking model — **Phrase box**

Type two other words from the spec box (one word per box below), which describe your topic

description layer — **Word boxes**

ALL FIVE BOXES ABOVE MUST BE COMPLETED BEFORE PROCEEDING!!

Select any Search Engine by clicking on one of the buttons below

| Google | Excite | Hotbot |

Figure 1. FOIOTI

Omission of information in any one or more of the last four boxes will produce inconsistent results. If the number of website summaries runs into the hundreds, thousands or more, the search needs to be focussed, possibly by adjusting the phrase length. It is also possible that too many general terms were used - rethink the information need, and update all the others. Close all the windows except FOIOTI, and resubmit the specification. As an example, a complete, successful FOIOTI searching session with results is shown in Figure 1. One of the three search engines produced the result summary screen shown in Figure 2.

When visiting the second site from the result summaries, a document is found which contains a full explanation of the original information need - see Figure 3 for an excerpt.

HOTBOT   If they book rooms   City
ST ⏷ GO

Results for *·"OSI"·"networking model"·descriptio (SEARCH) ☐ Search within these results

NEW SEARCH | REVISE OPTIONS | ADVANCED SEARCH

WEB RESULTS   (Showing Results **1 - 10** of more than 200)   next »
1. Chapter 2: Networking: 2.1 OSI Networking Model
   UNIX Hints & Hacks Contents Index Chapter 2: Networking
   Previous Chapter Next Chapter Sections in this Chapter: 2.1
   OSI Networking Model 2.4 Remote Network Configurations
   2.7 Unmounting and Busy Devic
   12/27/1999 http://lovecraft.die.udec.cl/OS/UNIX/unix-hints-and-
   hacks/19270033.htm
   See results from this site only.

2. Lesson 1: Open Systems Interconnection (OSI)
   Reference Model
   CIS370 HOME click for audio related to this page Lesson 1:
   Open Systems Interconnection (OSI) Reference Model This
   lesson describes the Open Systems Interconnection (OSI)
   reference model. The OSI refe
   2/26/2002 http://pluto.ksi.edu/~cyh/cis370/ebook/ch05b.htm
   See results from this site only.

3. Related Links
   The following websites are related to the current topic "OSI
   Reference Model". Feel free to click on them for more

Figure 2. Result Summary screen

A Layered Architecture

The OSI reference model architecture divides network communication into seven layers. Each layer covers different activities, equipment, or protocols. Figure 5.1 represents the layered architecture of the OSI reference model. (*Layer* different functions and services as data moves from one computer through the network cabling to another computer. reference model defines how each layer communicates and works with the layers immediately above and below it. Fc session layer communicates and works with the presentation and transport layers.

7. Application layer
6. Presentation layer
5. Session layer
4. Transport layer
3. Network layer
2. Data-link layer
1. Physical layer

Figure 5.1 *The seven-layer OSI reference model*

Figure 3. Excerpt from explanation

4.6 Final results

The most basic purpose of this research project was to enable a learner to find relevant information on the Internet. One way to measure FOIOTI's success in achieving this goal, is to compare the number of Yes answers (i.e. 'Yes', I have found relevant information within the time and other constraints of the experiment) before FOIOTI with the number of Yes answers thereafter. This led to the definition of four possible classes, depending on the final answer given for each experiment (before/after): No/No, No/Yes, Yes/No and Yes/Yes.

However, since the two related experiments were often done on separate days at the same institution, it could happen that a learner attended only one of the two sessions. If one of the two answers were absent, that participant's contribution would become invalid, producing a fifth class. Other reasons for classifying a response as invalid included incomplete forms and invalid Yes/No answers, pointed out by other information supplied on the same form. These five classes were now formalised. For ease of interpretation, they were arranged from most wanted result (Class 1) to least wanted result (Class 4).

Table 7 FOIOTI Class Definition

CLASS	ANSWER (BEFORE/AFTER)	IMPLICATION
1	No / Yes	FOIOTI had a positive effect on searching ability
2	Yes / Yes	FOIOTI could not have a positive effect on searching ability - learner found information without it
3	No / No	FOIOTI had no positive effect on searching ability
4	Yes / No	FOIOTI had a negative effect on searching ability
5	Not applicable	Data was ignored

Finally, if class 5 was indicated, a reason was added.

Table 8 Reasons for Invalid definition

NO.	DESCRIPTION
R1	Documentation incomplete - only did one experiment
R2	Information on Form incomplete - cannot decide if answer is Yes or No
R3	Information on Form strongly suggests that wrong Yes/No answer was given
R4	Other reasons

However, only class 1, 3 and 4 provide a true indication of FOIOTI's performance. In the case of class 2, FOIOTI could not have any positive effect on the learner's searching ability, and it did not have a negative effect. Similarly, class 5 is invalid for reasons outside the control of the author, and it has no bearing on FOIOTI's performance. Therefore both

classes 2 and 5 are omitted for the final evaluation.

FOIOTI's performance-based results of the last two sets of experiments are listed in Table 9.

Table 9 FOIOTI Class Results

	PHASE 11	PHASE 13	AVERAGE %
CLASS 1	76	49	71.0
CLASS 3	31	14	25.6
CLASS 4	3	3	3.4
TOTAL	110	66	100

Overall, FOIOTI thus improved the searching ability of 71.0% of learners, it had no positive effect on 25.6% of learners, and it had a negative effect on 3.4% of learners.

5 Conclusion

Considering the limits of the study in terms of defining the population, the practical issues of working in computer laboratories, etc, the results showed that FOIOTI can be successful in increasing the success rate of the population sample.

It is doubtful whether there will ever be an up-to-date, central index from where users can access all the information on the Internet in an orderly and consistent way. The freedom which users have to add and modify information on websites, makes it virtually impossible to manage and control an index of this kind.

Search engine databases are in a constant state of change, meaning that the same search on the same search engine on two consecutive days often yields different results. Interfaces and even searching algorithms are updated, forcing the user to alter his/her searching strategy.

The result is that it is up to the user to acquire a certain level of navigational skill, and specifically searching competence. A basic understanding of the alternative search engine operators outlined earlier is essential for survival in a world where information overload commonly occurs. Carefully focussed searches using more advanced operators and concepts are becoming almost a necessity and not the exception.

However, extensive searching experience and consultation by the author have shown that most users have a strong resistance to: the technicalities of searching, changing to new search engines and the learning curve of search strategies. Products like FOIOTI could solve this problem.

A number of dedicated data bases and/or search engines have been in existence for some time, servicing some of the major disciplines such as law, medicine, finance, etc. (Weideman 2002b). These search engines focus on one specific area only, and offers the user a selection of links in a very specific and narrow area. However, most of them still slow the user down by offering unfriendly interfaces, technical procedures and limited depth databases. The FOIOTI concept could be applied to one of these search engines, allowing the user easier access to a set of selected, focused set of information.

It is finally concluded that, if used by an audience as defined, there is a strong possibility that FOIOTI will increase the success rate of Internet searching.

A new research direction which the author has embarked on, flowing from the research described here, is website visibility. This article (and other work done at this point) looks at the website from the user's side: how can the user easily find the website? Website visibility looks at the same website from a different angle: how can the website owner ensure that, once a search engine has found the website, it will be indexed properly? Work is currently in progress towards defining a best practice approach to designing websites for search engines.

6 Acknowledgements

7 References

Alreck, P.L. and Settle, R.B. 1985. *The research survey handbook.* Homewood, IL: Irwin, 63.

Boulton, R. 2002. *An introduction to information retrieval.* [Online]. Available: *http://www.omsee.com/developer/docs/intro_ir.html* [Site visited on 10/04/2002].

Brewer, E.A. 2001. When everything is searchable. *Communications of the ACM.* 44(3): 54.

Bruce, H. 1998. User satisfaction with information seeking on the Internet. *Journal of the American Society for Information Science*. 49(6): 541-556.

Chun, T.Y. 1999. World wide web robots: an overview. *Online & CD-ROM Review*. 23(3): 135-142.

Courtois, M.P. and Berry, M.W. 1999. Results ranking in web search engines. *Online*, 44.

Cronje, J.C. and Clarke, P.A. 1999. Teaching 'teaching on the Internet' on the Internet. *South African Journal of Higher Education*. 13(1): 213-226.

De Jager, K. and Sayed, Y. 1998. Aspects of information literacy at five institutions of higher education in the Western Cape. *South African Journal of Higher Education*. 12(2): 197-203.

Edling, R.J. 2000. Information technology in the classroom: experiences and recommendations. *Campus-Wide Information Systems*. 17(1): 10-15.

Feldman, S. 2000. *Web search services in 1998: trends and challenges*. [Online]. Available: *http://www.infotoday.com/searcher/jun98/story2.htm* [Site visited on 08/12/2000].

Fidel, R. 1991. Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science*. 42(7): 515, 517, 520, 521.

Fidel, R. *et al*. 1999. A visit to the information mall: web searching behavior of high school students. *Journal of the American Society for Information Science*. 50(1): 24-37.

Foley, M.J. 1996. Hot new database technologies. *Datamation*. September: 44-49.

Forcese, D.P. and Richer, S. 1973. *Social research methods*. Englewood Cliffs, NJ: Prentice-Hall, 121.

Frants, V.I. and Shapiro O, J. 1991. Algorithm for automatic construction of query formulations in boolean form. *Journal of the American Society for Information Science*. 42(1): 16-26.

Gan, S.L. 1998. An overview of information technology and education in Malaysia. *Journal of Global Information Management*. 6(1): 27-32.

Garman, N. 1999. Meta search engines. *Online*. 23(3): 74-78.

Habib, D.P. and Balliot, R.L. 1999. *How to search the world wide web: a tutorial for beginners and non-experts*. [Online]. Available: *http://www.ultranet.com/~egrlib/totur.htm* [Site visited on 6/3/1999].

Heinstrom, J. 2002. *Fast surfers, Broad scanners and Deep divers - personality and information-seeking behaviour*. [Online]. Available: *http://www.abo.fi/~jheinstr/text.pdf*. [Site visited on 19/11/2003].

Hock, R. 1999. Web search engines - features and commands. *Online*. May/June: 24-28.

Huwe, T.K. 1999. New search tools for multidisciplinary digital libraries. *Online*. March/April: 67-74.

Jacso, P. 1999a. Database selection tools. *Online & CD-ROM Review*. 23(4): 227-229.

Jacso, P. 1999b. Savvy searchers do ask for directions. *Online & CD-ROM Review*. 23(2): 99-102.

Janes, J. 2000. Sampling: you don't have to ask everybody. *Library Hi Tech*. 18(1): 97-101.

Jansen, B.J., and Pooch, U. 2001. A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*. 52(3): 237-239.

Kothari, C.R. 1991. *Research methodology methods & techniques*. New Delhi: Wiley Eastern, 121.

Lancaster, F.W. 1978. *Information retrieval systems: characteristics, testing and evaluation*. New York, NY: John Wiley, 176, 279, 331.

Large, A., Tedd, L.A. and Hartley, R.J. 1999. *Information seeking in the online age: principles and practice*. London: Bowker-Saur, 5.

Large, A. and Beheshti, J. 2000. The web as a classroom resource: reactions from the users. *Journal of the American Society for Information Science*. 51(12): 1069-1080.

Lynch, C.A. 2001. When documents deceive: trust and provenance as new factors for information retrieval in a tangled web. *Journal of the American Society for Information Science and Technology*. 52(1): 12-17.

Marais, P.J.J.G. and Marais, A.F. 1999. Science education for the new millennium: an exciting challenge. *South African Journal of Higher Education*. 13(2): 82-89.

Mickey, B. 1999. A web search trifecta. *Online*. May/June: 79-82.

Mizzaro, S. 1998. How many relevances in information retrieval? *Interacting with Computers*. 16: 303-320.

Notess, G.R. 1997. Internet search techniques and strategies. *Online*. July/August: 63-66.

Notess, G.R. 1999a. On-the-fly search engine analysis. *Online*. September/October: 63-66.

Notess, G.R. 1999b. Rising relevance in search engines. *Online*. May/June: 84-86.

O'Leary, M. 2000. New roles come of age. *Online*. 24(2): 21-25.

Ross, N.C.M. and Wolfram, D. 2000. End user searching on the Internet: an analysis of term pair topics submitted to the Excite search engine. *Journal of the American Society for Information Science*. 51(10): 949-958.

Sapsford, R. and Jupp, V. 1996. *Data collection and analysis*. London: Sage, 28.

Schwartz, C. 1998. Web search engines. *Journal of the American Society for Information Science*. 49(11): 973-982.

Sherman, C. 1999. The future of web search. *Online*. May/June: 54-61.

Siegfried, S., Bates, M.J. and Wilde, D.N. 1993. A profile of end-user searching behavior by humanities scholars: The Getty Online Searching Project Report No. 2. *Journal of the American Society for Information Science*. 44(5): 273.

Singh, P. 2003. *Information seeking in Electronic Environments*. [Online]. Available: *http://drtc.isibang.ac.in/~prachi/cool o1.html* [Site visited on 18/11/2003].

Spink, A. 1996. Multiple search sessions model of end-user behavior: an exploratory study. *Journal of the American Society for Information Science*. 47(8): 603-609.

Spink, A., Batesman, J. and Jansen, B.J. 1999. Searching the web: a survey of Excite users. *Internet Research: Electronic Networking Applications and Policy*. 9(2): 117-128.

Spink, A. *et al*. 2001. Searching the web: the public and their queries. *Journal of the American Society for Information Science and Technology*. 52(3): 226-234.

Sullivan, D. 1999. Crawling under the hood. *Online*. May/June: 30-38.

Sullivan, D. 2002. *How Search Engines Work*. [Online]. Available: *http://searchenginewatch.com/webmasters/work.html* [Site visited on 08/11/2002].

Sullivan, D. 2003. *Major Search Engines and Directories*. [Online]. Available: *http://searchenginewatch.com/links/article.php/2156221* [Site visited on 05/11/2003].

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.

Wallace, R.M., Kupperman, J. and Krajcik, J. 2000. Science on the web: students online in a sixth-grade classroom. *The Journal of the Learning Sciences*. 9(1): 75-104.

Weideman, M. 2001. *Internet searching as a study aid for information technology and information systems learners at a tertiary level*. PhD thesis, p197. University of Cape Town.

Weideman, M. 2002a. *FOIOTI FAQ Page*. [Online]. Available: *www.mwe.co.za/home/seafaqtips.htm* [Site visited on 20/11/2002].

Weideman, M. 2002b. *Search Engine URL Page*. [Online]. Available: *www.mwe.co.za/home/seaenginesurl.htm* [Site visited on 20/11/2002].

Williams, J. 1996. *Bots and other internet beasties*. Indianapolis, IN: Sams, 246.

Wolfram, D. and Dimitroff, A. 1997. Preliminary findings on searcher performance and perceptions of performance in a hypertext bibliographic retrieval system. *Journal of the American Society for Information Science*. 48(12): 1142-1145.

Wong, P. 1998. Leapfrogging across the millenium: information technology in Singapore schools. *Journal of Global Information Management*. 6(1): 5-13.