

**SAICSIT 2002**  
**Enablement through**  
**Technology**  
Annual Conference of the  
**South African Institute of Computer**  
**Scientists & Information Technologists**  
The Boardwalk, Port Elizabeth, South Africa  
16-18 September 2002



# FOIOTI: Successful Internet Searching for the Average User

Melius Weideman

*Faculty of Business Informatics, Cape Technikon, Cape Town*  
**meliusw@yahoo.com**

## Abstract

The sheer size of the Internet and lack of categorization of the information available makes finding relevant information a daunting task. A number of methods do exist to enable users to be successful in this venture. Most search engines offer phrases, Boolean, Inclusion, Exclusion and other operators. The average user seems to exhibit a measure of resistance in using most of these operators, however. Other problems noted include a lack of clear search specification formulation and low productivity due to serial use of search engines.

A total of 1109 learners from three continents were involved 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. 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 the topic.

**Keywords:** *Internet, searching, interface, search engine*

**Computing Review Categories:** *H.3.3*

## 1 Background And Fundamental Concepts

The Internet offers an ocean of data to the user, comparable in size to the grains of sand on the seashore, and changing as regularly as the waves breaking on the beaches of the world. Although many programs exist to enable the average user to sail the data seas (friendly browsers, free search engines), general consensus exists that navigating the Internet is not a straightforward task [9]. One specific skill, which eludes many of these average users, is the finding of relevant information on the Internet in a short time [1].

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 data. Both approaches should lead the user to those elusive sites, which contain the exact information required. Commonly used search engines include AltaVista, Excite, Hotbot, Infoseek and Lycos [2]. 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 and Looksmart) 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 [8].

Since human collectors grow tired and need lunch and sleep breaks, they cover fewer websites (thus producing smaller indices) than spiders. Owing to their large numbers and 24/7 mode of operation, spiders produce massive indices: Google's index currently contains a claimed 2073 million webpages [4]! However, since they cannot evaluate websites for relevance, authenticity, etc, spiders' indices are often of lower quality than those produced by their human counterparts.

## 2 Alternative Internet Searching Approaches

Most computer users believe that higher numbers normally indicate 'better' in the computer world: faster processors, more memory and disk space, highest numbered version of software and more pages per minute of printer output.

All search engines offer a number of features to enable the user to find information with ease. However, it has been proved that most users do not avail themselves of these features [10]. 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.

An Internet search should start with the definition of an information need, followed by a search methodology, manifesting in the construction of a search query, which should lead to an acceptable answer (see Figure 1). An information need could be: 'I need more detail about the life of Osama Bin Laden,' while the corresponding search query could be: 'Osama Bin Laden' +Afghanistan +Qaeda – 'World Trade Center'.

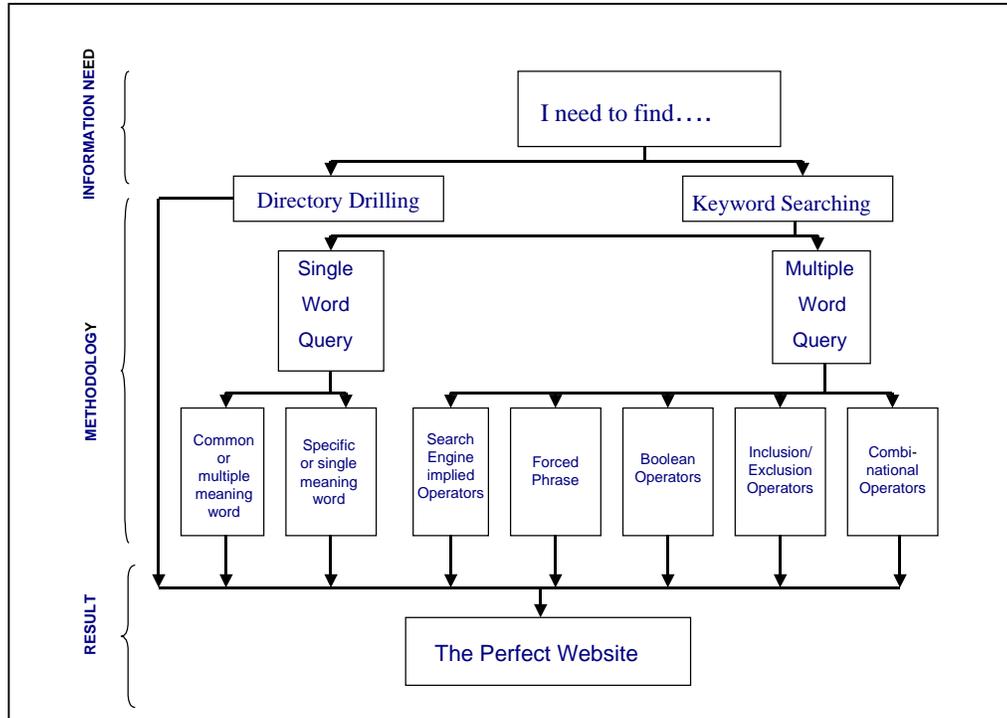


Figure 1: Alternative searching approaches

Probably the most basic form of 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 [7]. 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. The probability of finding a specific site among millions of answer is very low.

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.

In (forced) phrase searching, the user has to choose two or more words with a special relation between them, qualifying them as being a phrase. For example, ‘world cup final’ is a phrase, since it would fit into a grammatically correct English sentence. By this definition however, ‘cup final world’ is not a phrase, although the same words have been used. Most search engines require the user to enclose a phrase in quotes, thereby forcing the search engine to preserve the positional location of each word.

The use of Boolean operators, although simple in concept, creates major problems for many searchers in practice [6]. It is based on the linking of two or more words with one or more types of operators. To complicate matters further, some search engines add implied Boolean operators between words, without telling the user.

For example, a search where the user specifies ‘Tony AND Blair’ using Altavista would not yield the same number of results as ‘Tony Blair’. In the first case, both words must appear on the same web page before that page is returned as a possible answer. In the second case, the search engine itself decides how to link the two keywords.

Similarly, a search specification of ‘whales OR sharks OR dolphins’ will produce a list of answers where any one (or more) of these words appear on a website. Google adds an implied ‘AND’ between unconnected words typed by a user.

A method similar to certain Boolean operators, 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 moves 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, the average user seldom combines operators in this way.

### 3 How Do Users Actually Search?

Extensive research was done to establish how, where and for how long users search for information, given a free hand, 30 minutes and one topic they could choose themselves. A total of 1109 students, spread across three continents and 20 Higher Education Institutes were involved in 46 separate searching sessions in this endeavor. In the first phase, some pilot tests were done on 105 students to weed out any basic problems in the methodology, and a test instrument was designed to measure searching success. Secondly, 293 other students used this instrument, and the results were recorded. At this point, a total of only 39.1% of the participants claimed success in finding their information. Some of the most common problems that were found include:

- Users were not capable of using the operators at their disposal
- Users could not formulate their search specification
- Users wasted time by waiting for only one search at a time to produce answers.

As a result, it was considered viable to design a model to assist users in finding relevant information.

During the third phase, another 90 students were tested, and the results were used to further refine the model used to assist them in searching. At this stage, it was decided to code the final model as a web-based program. This had the advantage of being usable at remote sites, and also made it more easily accessible to other audiences. This program was called FOIOTI – Finder Of Information On The Internet.

A total of 554 students participated in the final phase of testing, which proved FOIOTI's success: 71.0% claimed that this program enabled them to find relevant information, whereas they were unsuccessful without it.

### 4 FOIOTI Design

FOIOTI's design is based on Fidel's definition of two distinct modes of searching: the operationalist versus the conceptualist approach [3].

The operationalist decides on an information need, and then concentrates on trying to find the best mix of search engine operators to achieve success. He/she could even use different search engines (a good idea in principle) in the search for that ideal website. This approach is clearly favored by the technically minded searcher.

However, the conceptualist attempts to think around the information need, and changes it as the search progresses. This person realizes that the formulation and subsequent fine-tuning of the search query is crucial to searching success.

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 even simplifies the choice of search engine (which in itself could be a major topic of discussion) to the click of a button. The user simply types in the information need, extracts a few 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.

## 5 FOIOTI Usage

FOIOTI is available at [www.mwe.co.za](http://www.mwe.co.za). Follow the links from the homepage (you will need Java to be enabled on your browser to see FOIOTI).

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 large gray FOIOTI box (see Figure 2) requests you to complete all five rectangular white boxes, before you can click on any one (or more) of three search engine buttons at the bottom. It will then construct a query, using the information you have just supplied, load the search engine(s) you have chosen, and feed the query to it. The result screens are presented to the user in the same way had you simply accessed the search engine the normal way. A complete example of how you could effectively use FOIOTI is given at the end of this section.

The first box asks you to type out your information need in a full English sentence. This is a crucial first step, since it forces you to think carefully about what it is that you are really looking for. You should use very specific words and phrases, for example 'BMW' rather than 'motorcar,' 'standard marathon' rather than 'jogging,' etc.

Also attempt to use words which are difficult to spell – this obviously depends on the topic of your search (e.g. oesophagus, derailleur, Czechoslovakia, etc). If you make a spelling mistake, you will receive answers from website authors who made the same mistake – probably not what you want! If unsure about spelling, use a dictionary or simply type the word(s) in a word-processing document and use the spellchecker feature. If you have trouble in completing the information requested in the remaining four boxes, it normally implies that your information need in the first box is incomplete. You 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 you to fill in a word (it could be more than one) which is normally spelt using (a) capital letter(s). First names and abbreviations immediately come to mind. Be careful 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. Capitalization is obviously preserved before the query is submitted.

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.' Steer away from 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.

If you complete all the boxes, start a search and receive a few thousand answers, it probably indicates that you have not focussed the search enough. Add one or two words to the phrase and try again – this should bring down the number of answers, indicating better focussing. Similarly, if you receive zero answers 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'

Again your capitalization is preserved – by all means use capitals (correctly) in this box, as in the last three phrases listed.

The final two boxes require any two words related to your topic to be typed in. Again, attempt to refrain from using general terms (such as 'bicycle' or 'President') – substitute with specific words (such as 'Trek' and 'Bush'). As before, if you have run out of ideas, return to the first box, rethink your search, and broaden the information need in this first box. Capitalization is not preserved in these last two boxes – most search engines convert whatever you type here to lowercase. Once all five boxes have been completed, you should now click on one or more of the search engine buttons at the bottom of the gray box.

Research has proved that the use of multiple databases increases the chances of success [5], so you are strongly advised to use all three for the same search. Simply click on all of them (you might have to minimize some windows to get back to the FOIOTI screen).

Do not waste time waiting for all the answers to be returned – switch between the windows until you see the first answer screen. While you are reading this first answer set, the other search engines have time to complete their tasks.

**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 good enough!)**

I need to find the name of the island close to Cape Town where Nelson Mandela was kept prisoner before his release **Spec box**

**Type one word from the spec box which is normally spelt with (a) capital(s)**  **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)**  **Phrase box**

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

**Word boxes**

**ALL FIVE BOXES ABOVE MUST BE COMPLETED BEFORE PROCEEDING!!**

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

Figure 2: FOIOTI screen

3. [Excite Travel: Destinations: South Africa](#)  
South Africa South Africa Monday, March 5 Yellow Pages Maps Today's Weather About South Africa The hottest tourist attraction in South Africa these days is Robben **Island**, where President Nelson...  
[http://www.excite.com/travel/countries/south\\_africa/](http://www.excite.com/travel/countries/south_africa/) [\[more from this site\]](#)
4. [Excite Travel: Destinations: South Africa](#)  
South Africa South Africa Sunday, March 4 Yellow Pages Maps Today's Weather About South Africa The hottest tourist attraction in South Africa these days is Robben **Island**, where President Nelson...  
[http://netcaster.excite.com/travel/countries/south\\_africa](http://netcaster.excite.com/travel/countries/south_africa) [\[more from this site\]](#)
5. [Daily Mail & Guardian Mandela](#)  
Daily Mail & Guardian Nelson **Mandela** site  
<http://www.mg.co.za/mg/mandela/b-secret.htm> [\[more from this site\]](#)
6. [Cosby to lighten up jail island](#)  
Comprehensive package of the most recent television news and information.  
[http://www.sunmedia.ca/TelevisionCosby/mar20\\_cosby.html](http://www.sunmedia.ca/TelevisionCosby/mar20_cosby.html) [\[more from this site\]](#)
7. [Black Power!](#)  
BYLINE: RUSSELL WATSON with JOSEPH CONTRERAS in Johannesburg and JOSHUA HAMMER in Durban Guided by Nelson **Mandela** and F. W. de Klerk, South Africans complete an amazing, and largely peaceful, tra...  
<http://aace.virginia.edu/go/capetown/B-black.html> [\[more from this site\]](#)
8. [CNN - Clinton, Mandela agree to disagree - March 27, 1998](#)  
SITE SOURCES Clinton and **Mandela** walk hand in hand before a news conference in **Cape Town** Friday In this story: **CAPE TOWN**, South Africa (CNN) -- Despite praising President Clinton for his "co...  
<http://www.cnn.com/WORLD/9803/27/clinton.mandela/index.html> [\[more from this site\]](#)
9. [Bruce Damer's Robben Island Prison Photo Gallery](#)  
South African Journeys Robben **Island** picture gallery, May 1998 About Robben **Island** Prison A visit to the infamous prison **island** off the coast of South Africa.

Figure 3: Results from Excite

If you receive fewer than about 50 answers on any given screen, you are probably close to success. It is worthwhile inspecting some of these websites. When you do find a screen with a number of potentially useful website summaries, make a few copies of this screen, and run each search in parallel by loading one website per window.

If the number of website summaries runs into the hundreds, thousands or more, you need to focus your search, possibly by twiddling with the phrase length. It is also possible that you have used too many general terms – rethink your first box, and update all the others. Close all the windows except FOIOTI, and resubmit your specification.

If you receive zero answers, your phrase could be too long – adjust as described. If the phrase is already only two or three words long, the probability is high that there is a spelling mistake in a crucial word in one of your boxes. Check the spelling, close all unnecessary windows, and resubmit.

As an example, a complete, successful FOIOTI searching session with results is shown. Sites number 3, 4 and 9 in Figure 3 contain the answer (Robben Island) to the information need in the website summary, without even having visited the website itself.

## 6 The Future

It is doubtful whether there will ever be an up-to-date, easily navigable database, to allow Internet users orderly and consistent access to all the information on the Internet. The freedom which users have to add and modify information on websites, makes it virtually impossible to manage and control this information.

Search engine databases are forever changing, meaning that the same search on the same search engine today and tomorrow 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 searching methodologies outlined earlier is essential for survival in the data seas. Carefully focused searches using more advanced operators and concepts are no longer reserved for nerds only.

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 [11]. 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 a range of dedicated databases only, allowing the user easier access to a broader set of information.

## References

- [1] E. A. Brewer. When everything is searchable. *Communications of the ACM* 44(3): 54, 2001.
- [2] M. P. Courtois and M. W. Berry. Results ranking in web search engines. *Online*, page 44, 1999.
- [3] R. Fidel. Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science* 42(7): 515 –517, 1991.
- [4] Google; see [www.google.com](http://www.google.com)
- [5] P. Jacso. Database Selection Tools. *Online & CD-ROM Review* 23(4): 227, 1999.
- [6] B. J. Jansen and U. Pooch. 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, 2001.
- [7] S. Siegfried, M. J. Bates and D. N. Wilde. 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, 1993.
- [8] D. Sullivan. *How Search Engines Work*; see <http://searchenginewatch.com/webmasters/work.html>

- [9] H. J. Voorbij. Searching scientific information on the Internet: a Dutch academic user survey. *Journal of the American Society for Information Science* 50(7): 604-605, 1999.
- [10] M. Weideman. *Internet searching as a study aid for information technology and information systems learners at a tertiary level*. Unpublished PhD thesis, page 197. University of Cape Town, 2001.
- [11] M. Weideman. *Foioti faq Page*; see [www.mwe.co.za/home/seafaqtips.htm](http://www.mwe.co.za/home/seafaqtips.htm), 2002.