

# ***Internet Informed***

**Guidance for the Dedicated Searcher**

**(2-Chapter Version)**

**(FYI: The full version is online now)**



**Precision  
Prominence  
Quality  
Identity  
Haste  
Structure  
Attention  
Utopia  
Pursuit  
Choreography  
Author Bio**

**by David Novak**

## **NOTICE !!!**

The full book, Internet Informed : Guidance for the Dedicated Searcher,  
is now published under a Creative Commons License and online,  
available as either a single 332 page PDF file, or as 9.5 hours of  
Audiobook.

If you like this work, read on.

You will find a copy at [SpireProject.com](http://SpireProject.com)

- David Novak

FOR MY FAMILY AND FOR THOSE WHO SHARE THE UTOPIAN DREAM.

Internet Informed : Guidance for the Dedicated Searcher  
(Limited Internet Version)  
By David Novak

Copyright ©2006,2008 David Novak.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License. See [CreativeCommons.org/licenses/by-nc-nd/2.5/](http://CreativeCommons.org/licenses/by-nc-nd/2.5/) for details.

Attribution Required. NonCommercial use permitted. No Derivative works.

Translation assistance most welcome. Corrections welcome.

Preferred Font: Gentium/Arial. See [scripts.sil.org/gentium](http://scripts.sil.org/gentium) to download this free font.

Cover Art: NGC2440 “Cocoon of a New White Dwarf” acquired by Howard Bond (STScI) and Robin Ciardullo (Penn State) using the Hubble Space Telescope. Post-processing by Forrest Hamilton. Used with permission and courtesy of STScI/NASA. The glowing ball is a white dwarf, the remnant of an exploded star, speeding away from the centre of the initial explosion at tremendous speed.

### **This document resides at [SpireProject.com](http://SpireProject.com)**

This Limited Internet Edition consists of Chapter One and Chapter Two of:

Internet informed : guidance for the dedicated searcher.

ISBN: 0-9757299-1-8

Paperback 332 pages printed December 2008

Dimensions 158mm x 234mm

Price: A\$59.95

Internet Informed is published by The Spire Project. This limited internet version is taken from the printed book, Internet Informed which is shipping now. To purchase a copy, visit [SpireProject.com](http://SpireProject.com) or use the form at the end of this document.

The Spire Project  
PO Box 60  
Toorak, Victoria 3142,  
Australia

To arrange bulk purchase discounts, premiums or promotion for this or the full print edition, contact sales directly on +61 403055544 or preferably use the email form on the bottom of [SpireProject.com](http://SpireProject.com)

Disclaimer: While considerable precaution has been taken in the preparation of this book, the author and publisher assumes no responsibility for errors, omissions or for damages resulting from the use of the information herein.

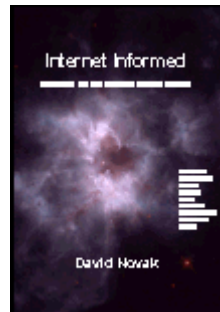
ALSO BY DAVID NOVAK...

## **The Spire Project** Better ways to find information.

### **The Information Research FAQ**



And now:



# ***Internet Informed***

**Guidance for the Dedicated Searcher**

**(Limited Internet Version)**

by **David Novak**

of



**The Spire Project**

'Tis true. There's magic in the web...

A sibyl... in her prophetic fury

Sewed the work.<sup>1</sup>

– William Shakespeare

# CONTENTS

PROLOGUE	· 5
<b>CHAPTER ONE: PRECISION</b>	<b>· 15</b>
<b>CHAPTER TWO: PROMINENCE</b>	<b>· 49</b>
QUALITY	· 67
IDENTITY	· 123
HASTE	· 155
STRUCTURE	· 175
ATTENTION	· 195
UTOPIA	· 221
PURSUIT	· 261
CHOREOGRAPHY	· 283
EPILOGUE	· 295
GLOSSARY/NOTES/INDEX	· 301
<b>AUTHOR BIO</b>	<b>· end</b>

(Notice – full book now online as pdf and mp3.  
See [SpireProject.com](http://SpireProject.com)

## PRECISION

If a search is war, then the global search engine is our sword. Grab this favoured weapon, march into battle and swing. Many a battle can be fought and won with this sword, especially if the enemy is a peasant, a simpleton. Occasionally we need finesse. Sometimes we need much much more.

Let us hold this sword of ours correctly. Let us address the punctuation accepted by the vast global search engines. Search engine punctuation consists of a set of tactics that allow us to insist search engines provide us with specific information. We will describe what ‘specific’ means later in this chapter but these tactics are widely used in library circles since they form a foundation for searching all computerized databases. From library book catalogues to the most expensive of patent databases, we use tactics with names like proximity indicators, Boolean operators and field search terms. It is all very complex.

On the internet, however, these tactics often behave differently than library science would suggest. Many tactics are abridged and severely limited.

We will look closely at quotes “ ”, the +/- symbols, the use of **OR** and three field searches: **TITLE**, **URL** and **LINK**. There are further tactics. You may know some of them. We will focus just on these since they provide almost all the tactical advantages we will need and since these tactics apply almost uniformly across the many search engines.

Toss a few words to a search engine. Type something and receive a list of a hundred thousand matching results. More accurately, we receive the first twenty search results from a list a hundred thousand long. We do not get a hundred thousand results. We cannot get a hundred thousand results.

We get only the top of this list. For many reasons we will address, this may not be the start of a good search.

We search in a more specific manner by adding punctuation. We can, for instance:

- insist two words appear next to each other on a webpage,
- insist a word appears in the title of a webpage,
- insist results have some element in the address of a webpage
- and remove from our attention anything with a particular word, title or element in its web address.

Punctuation allows us to be specific with our attention. Yes, search engines practice a kind of relevancy ranking. They invite us to let them select which information we should browse. This ranking becomes more sophisticated every year. Ranking already duplicates some of the tactics I am about to introduce. However, like the purist who asserts everyone should learn to cook an egg, I believe we should all learn to punctuate our searches. Only then will we have the option to reject this ranking assistance. On certain occasions, throwing a few keywords at a search engine works very much to our advantage – many occasions if we seek general overviews or if we phrase our questions well. Yet if we ask a challenging, specific or comprehensive question, throwing keywords fares rather badly indeed. Let us consider each tactics, each punctuation mark, in turn.

## QUOTES

**internet service provider** reveals webpages with these three words.

**“internet service provider”** reveals webpages with this phrase.

With quotes, we insist words appear together. In library-speak this is called basic proximity. When we place quotes “” around two or more words in our search query, we insist the results include these words, together, in order.

A search for **“internet service provider”** will match only pages with this phrase. As a search, this is enormously more specific than a search for **internet service provider** (without quotes), a search that asks only that these three words appear somewhere on the page, in any order, together or apart.

Thanks to ranking technology, the major search engines appear to render this tactic unnecessary. Search for a couple words, perhaps someone's name, and webpages where our words appear beside each other are preferentially lifted to the top of the list. Adding quotes to a search may not change anything on the first page of results. Simple searches, however, lack a specific nature. When we are not specific, the number of matches means little. We will come to value this number soon.

Including quotes in our search is the single simplest way to search more effectively. The use of quotes is a tactic that works on every search engine and most every search tool we will ever meet (though some search tools may require we select 'as a phrase' from a selection box instead). Occasionally, when we use quotes, we will retrieve results with our words separated by a comma, a period or perhaps a stop word. Stop words are simply words search engines usually ignore: words like **a/the/and**. Irrespective, using quotes will always generate a far smaller and far more focused list of results.

Search for a book title, a person's name, a phone number – especially search for a concept like “underground irrigation” or “unconditional love” – and we should use quotes. I use quotes in at least half of all my searches.

Suppose we seek information about an author; about me. A search for **“David Novak” research** will return a list of webpages about myself, and as it happens, another David Novak active in Jewish historical research. Such a search is specific. Search without quotes, search for **David Novak research**, and we generate a much longer list, fifty times longer, listing all webpages with these three words: **David** and **Novak** and **research**. Such a list is messy and unfocused. Muddy. Forty-nine in fifty of these references point to webpages by someone other than David Novak – perhaps by a David Brown and James Novak – since all we ask is that our three keywords appear on a page.

Use quotes for a more specific search. Remember this and we need never ask a friend for the address to their website. Just ask how to spell their name. With a name in quotes and a single word describing one of their most obvious interests, we should have little difficulty finding their website (unless the person is almost unknown to the internet).

Incidentally, we can also use quotes with all library catalogues and all commercial-quality databases. It works the same way. Secondly, we may not need to type the closing quotes since search engines will often close

quotes for us. A search for “**underground irrigation** [lacking the closing quote marks] gives the same results as “**underground irrigation**”

## THE PLUS AND MINUS SYMBOLS

**+love** reveals only those webpages with the word ‘love’.

**-love** reveals only those webpages without the word ‘love’.

A second tactic is to insist words appear or do not appear in the results. In library-speak this is called Boolean searching, after mathematician George Boole (1815 to 1864) who wrote a paper on the mathematics of logic. He described the mathematical use of the words AND, OR and NOT and their role in set theory. You may remember studying this topic in high school along with Venn diagrams. This Boolean was once known as the insurmountable molehill since older library surveys showed the use of Boolean dumbfounded the lay public. On the internet, Boolean is worse. Without standards, with several search engines only recently accepting the use of brackets and without knowing in advance how Boolean is applied on a particular search tool, Boolean falls apart at its seams. It becomes three different tactics: AND, OR, NOT.

Our first step is to replace the word AND with the plus symbol (+), NOT with the minus (-) symbol. Using the +/- symbols avoids some confusing results on certain search tools. While most search tools interpret AND and NOT correctly, I have yet to encounter a search tool that misinterprets the +/- symbols.

Plus/minus is simple. Place the plus symbol (+) immediately before a word to insist the word be present in each matching record. Place the minus symbol (-) immediately before a word to insist the word **MUST NOT** appear on the referenced document.

**+unconditional +love -medicine**

Send this query to a search engine and we generate a list of webpages or web documents that include the words **unconditional** and **love** but do not include the word **medicine**. It seems simple and it is. Furthermore, we can place a +/- before quotes and in front of the title tag and other tags we will introduce in a moment.

### +“David Novak” -title:spire

Notice the plus comes before each and every word or word group. Miss the leading + before ‘David’ and we will occasionally encounter search tools that treat our first word as optional.

We must address two simple changes to this picture at this time. The first requires a little history lesson.

About six years ago, the popular press hammered the large global search engines mercilessly for returning millions of pages any time we typed a few words. At that time, a search for **three blind mice** would retrieve a list of tens of millions of matches simply because search engines considered pages with any of our words, even just one word, as a match. The popular press had a field day with this confusion making it the catchphrase for the chaos of the internet.

Then, almost overnight, all the primary global search engines changed so as to presume that when we type several words, we want all these words. Today, global search engines assume a plus symbol (+) precedes each word.

We rarely need to use the + symbol now. Plus is assumed. But beware. Every so often, I encounter some search tool that still defaults to any word. There is also something called ‘Fuzzy And’, a search for three words that returns no matches, triggers a search for pages with two of the three words we seek. That is, a fuzzy search gives the best answer it can, always offering some suggestion even when nothing contains all the search words we request. AltaVista implemented ‘Fuzzy And’ for a time in 2002. In early 2006 I saw it again in Yahoo’s Video Search. While rare, ‘Fuzzy And’ is fairly typical of the subtle oddities we encounter time and time again among the many internet search tools.

Historically, the use of plus was tremendously helpful back when it was not assumed. Today, we leave it off and just assume our search tools understand we want all our words. However, if ever we receive a confusing response from a search tool – and more on what constitutes a confusing response shortly – then one possibility is we have stumbled upon a search tool that does not assume the plus symbol. Now that we know how to use the plus symbol, forget it.

The second change to the picture we have just painted involves the use of the minus symbol (the NOT function) that changes a basic tenet of library science. When searching a commercial database, researchers are strongly advised against using the Boolean NOT since a researcher is far

too likely to remove items of interest. This is good advice. Consider a search for **heartache NOT love** on a medical article database. The use of **NOT love** will remove that perfect article that just happens to read, “Many doctors love to treat heartache with Aspirin.” The word love is present so the reference is discarded. Yet this referenced article may be the only article in the database that connects Aspirin to heartache. Commercial databases are best searched in a very specific manner with very limited, cautious use of NOT. Many of the search features of commercial-quality databases, like a heavy use of descriptors and the refined use of fields, assist us to craft such specific searches.

The internet, however, is a different beast to the commercial database. Google indexed over 8 billion records as of late 2005 and a suggested greater than twenty million as of late 2006 - far more than any commercial database. Despite this, we miss great quantities of information when we reach for a global search engine. Unlike a commercial or library database, the global search engine delivers incomplete coverage. We search a fraction of the internet. It is hard to say with certainty, hard even to guess, but I expect Google’s index covers just 10% to 20% of the internet directly.

Any search misses more than it searches. We will look more closely at coverage in time but our struggle is not to sift carefully through a small quantity of articles from a carefully indexed, complete collection of literature. We shift rubble. If it’s not love, get rid of it. Use the minus symbol frequently and with little regard for what is removed. There is far too much information out there for us to be concerned about the few references we mistakenly discard along the way.

## OR (IN CAPITAL LETTERS)

**search OR research** reveals webpages with either word.

On to the next tactic: as of 2003, the top global search engines finally standardized their use of OR. As of 2006, the top three global search engines marry OR with brackets in a way standard among commercial databases for decades. I want **(pizza “home delivery”) OR (chinese “take away”)**. AltaVista started this trend but now Yahoo, Google and MSN Search accept OR and brackets. Other search tools accept OR but not brackets. **this\_word OR that\_word**.

Being precise here, OR means either/or/both. Just one word is required. If both appear, that is fine too. We use OR on the internet primarily to broaden a search by including synonyms and alternative spellings for our search words. We also use OR to allow for plurals.

**hello OR hi**

**reporter OR journalist**

**color OR colour**

**heartache OR “heart ache”**

**dog OR dogs**

This last example is surprisingly important since as I write, global search engines consider dog and dogs as different words. We rarely care and usually mean dog or dogs but we must convey this to a search engine each time by using **OR** (in capital letters).

OR works for phrases in quotes:

**“hello kitty” OR “kero keroppi”**

and also with fields:

**intitle:“hellokitty” OR inurl:hellokitty**

The first search returns pages with either the phrase “hello kitty” or “kero keroppi” (or both). The second search returns pages with hellokitty either in the title or the URL (or both). Allowing for alternative spelling, for synonyms and for plurals in this way is good searching. It is professional. This tactic may reveal relevant information that would otherwise lay hidden.

Personally, I do not often take the time to write properly inclusive searches rich in the use of OR. When a search of mine returns insufficient matches, I certainly reach for OR then. When an obvious synonym or comparable term arises, I certainly reach for OR. However, for simple searches I don’t envision as difficult, I don’t bother. I use OR in perhaps 10% of my searches.

## FIELD SEARCHES

Finally we reach field searching, in many ways the pinnacle of internet searching, indeed all good searching. Walk into a library, approach a

computer and seek a book by name. We are about to undertake a field search. The computer will search just the titles and authors of all the books within the library database. It searches the author/title field.

Alternatively, we could seek a book by looking first for a suitable subject. This is a different field. The field this time is a record of all subject headings. Dewey decimal number – yet another field. Title/author, subject and Dewey decimal number are each distinct fields. Each search is a field search. This is not a puzzling concept. It is just that field searches differ greatly from the generic search-everything kind of search; a search often called a keyword search.

My state library's online catalogue allows searching by author, title, call number (Dewey decimal number) and subject. Further fields include the year of publication, material type, serial type, language, publisher and physical location.

ERIC (the Education Resources Information Center) at **eric.ed.gov** is a prominent free commercial-quality database of education related literature. It has more fields: Author, Title, ERIC Number, Journal Citation, Major Descriptors, All Descriptors, Identifiers, Abstract, Geographic Source, Institution Name, Publication Type, Publication Date, ISBN, ISSN, Clearinghouse Number, Government, Availability, Note and Language. We can search for something in any of these categories, in any of these fields.

The US Library of Congress Online Catalogue (LOCOC) at **catalog.loc.gov** contains records to over twenty-nine million books and millions more manuscripts. It has many more fields. In addition to Author, Subject, Title, Corporate Author, Publication Date and Publisher, a further 30 fields exist! As you may suspect, field searching is very significant to library and commercial research.

A field has a strict meaning in computer science as the area of a database record into which a particular item of data is entered. The library science definition is more suited to searching since it insists on a logical category of bibliographic description. That is, fields concern a fact about the information.

For some commercial databases, a tremendous amount of work infuses the accurate development of fields such as abstracts and descriptors. A definitive list of descriptors called a thesaurus may be created to standardize classification and assist searching. To categorize inventions, the World Intellectual Property Organization (WIPO) creates a complex system

of patent classification updated and improved every three years. This International Patent Classification (IPC) is now on version eight and is pivotal to searching patents.

Field searching is a vital step in the effective use of any commercial-quality information database. However, the internet is most certainly not commercial-quality nor professionally organized. We have only three fields available of note: TITLE, URL and LINK. Yet searching these three fields is pivotal to using the internet effectively. Much of this book will tease out implications that arise directly or indirectly from two of these three fields. Brilliant searching start here.

## THE TITLE SEARCH

**intitle:jupiter** reveals webpages with ‘jupiter’ in the title.

On the internet, the TITLE is the few words that appear in the very top left bar of our web browser. For readers who understand Hypertext Markup Language (HTML), the title corresponds to the words found between the two title tags, <title> and </title>, positioned near the beginning of all webpages. The title usually describes the content of the webpage in just a few words. Most titles are just two to six words in length and many are completely unrelated to the topic of the webpage.

Consequently, title searches are rather clumsy. Few webpages about Afghanistan will include **Afghanistan** in their title. Yes, webpages with **Afghanistan** in the title most certainly discuss Afghanistan in some manner and this may sound promising. However, searching well involves being a little more specific than this. Even if we seek something general, it is better to undertake a search for prominence, the topic of Chapter Two, than a crude, brutish search by title. If we have reason to expect a word belongs in a title, then proceed. Otherwise, this very hit-or-miss approach will reveal perhaps only five percent of the relevant documents. Yes, a crude tactic indeed.

To request a title search, simply precede the search word or words with the term **intitle:.** Type this into the standard search box of a global search engine.

intitle:spire	Search
---------------	--------

If you prefer Yahoo, use the little textbox on **yahoo.com** – not their advanced search page. If you prefer Google, use the small search box on **google.com**; not their advanced search page located elsewhere. Type **intitle:spire** for webpages with ‘spire’ in the title. Type **intitle:“spire project”** retrieves webpages with the phrase ‘spire project’ in the title.

Be sure not to include a space between **intitle:** and the word or phrase to follow. **intitle: spire** (with a space) is a request for two words not one in the title. Also keep in mind field search terms like **intitle:** can change over time. Yahoo once used **title:** but now matches Google with **intitle:.** MSN search must have recently added **intitle:** as well. Other search tools may use the previous standard field search term of **title:.**

## THE URL SEARCH

**inurl:jupiter** reveals webpages with ‘jupiter’ within the web address.

URL stands for Uniform Resource Locator. It is the specific location on the internet for an item of information. Now that most everything has migrated to the web, URL roughly translates as web address. <http://something-or-another>. There is a subtle difference between the meaning of URL and web address since URLs can point to destinations that are not strictly webpages; perhaps a newsgroup, a telnet service or an ftp archive. Differences like these were once very important to searching the internet but no longer. Information moves between tools too easily and too much has migrated to the web. Format is a more effective concept, a topic for Chapter Four.

The web address is an elegant system to place each item of information at a unique location. With a URL search, we ask a search engine to reveal just the information that shares some element in its web address.

There is a great deal of information to be found within an address – more than a country code and type of organization (.edu/.gov/.com). For instance, information within the same directory shares most of the same address. Thus, being able to search for a specific URL allows us to ask for a list of all the webpages a search engine has found within a particular

directory. This is actually extremely empowering. We know of it as local context and will speak of it further in Chapter Three.

At this moment, however, let us just note that the URL field is a field we can search much like the title field. We can insist on or exclude information that has something particular in its URL. To request a URL field search, simply precede our URL-segment with the term **inurl:**. Thus, **inurl:wikipedia** is a search for webpages with ‘wikipedia’ somewhere in the web address.

Once again, search engines differ. The previous standard of **url:** has shifted now that Yahoo (but not AlltheWeb) uses **inurl:** like Google. MSN search as of early 2006 uses neither, not as described here. During July 2006, between June and July 2003 and again even earlier, Google dropped the support of **inurl:**. AlltheWeb allows **url:word** but not **url:address**. The list of complications go on and on.

Part of this difficulty rests with the very similar site field search (requested as **site:domain\_name**). Site is not nearly as flexible as inurl, so I rarely use site when inurl is available. Site does allow us to ask for a list of webpages in a particular domain, so **site:bbc.co.uk** will return a list of webpages from this website. However, site does not permit us to reach into a specific directory to find truly local information (except, in some ways, on Google) or to request a specific word in a web address.

Just as an aside, if you do use the site search, or the URL field search for that matter, see if you can drop the ‘www’, the hostname that precedes most addresses. **site:bbc.co.uk** and **inurl:wikipedia.org** will often lead to far more results than a search for **site:www.bbc.co.uk** and **inurl:www.wikipedia.org**. Much of the English material on the Wikipedia, for instance, resides on **en.wikipedia.org**. A search for **inurl:www.wikipedia.org** would miss it.

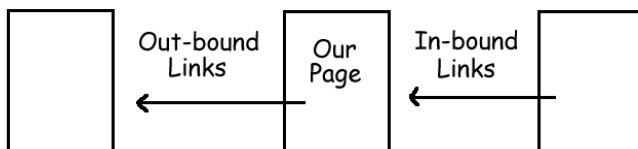
I really like how Google handles their URL field search. For many years their inurl field search term was not widely known. Their advanced search page used a clumsy **allinurl:** (which we can ignore) and now a **site:** search. Their help page only recently began to mention inurl at all. With the past standard as **url:**, which Google never supported, few people ever knew Google permitted inurl. It is Google’s hidden field search. I had a role in revealing inurl in the late 1990s I will tell you about later. Just keep in mind this field rocks! I use the URL field very frequently, perhaps 20% of the time. You must learn how to do a URL field search on your favourite search engine or adopt a search engine with a flexible URL field search.

## THE LINK SEARCH

**link:wikipedia.org** reveals webpages linking to **wikipedia.org**.

The link is a connection from one webpage to another. Essentially, a link directs attention towards another page. Click a link and we move the focus of our web browser to the newly referenced page. Links appear as images or text. For readers who understand HTML, the link comes from **href="web\_address"** usually found as **<a href="web\_address"> ... </a>**.

The link field refers to only the in-bound links: links originating on webpages elsewhere on the internet. Do not confuse this with a list of links on the page itself; links going elsewhere; links we shall call out-bound links. The link field search enables us to ask a search engine to list the links pointing at the webpage we specify. We provide the web address and the search engine replies with pages that link to that web address.



In a superficial way, the more in-bound links a webpage has, the more popular and more recognized the webpage. This is why search engines use the number and presumed significance of in-bound links in their ranking technologies. References that appear at the top of a search engine results page usually point to the webpages with the most in-bound links. We will explore this further in Chapter Two.

Once again, there is more to this link field. We can use link field searches to discover further related resources. Simply work backwards, then forwards again to the link companions. We can use the link search in quality assessment as one of several types of endorsements. We can even triangulate our way to information resources with the link search. At this moment, just note the link field is a field we can search just like the title and URL fields.

To request a link search, simply type a web address and precede it with the field search term **link:**. Do not include the **http://** since some search engines will not like it. Once again, no space between **link:** and the address that follows. Thus **link:wikipedia.org** is a search for all webpages with links to **wikipedia.org**.

**Link:** has long been the standard search term for a link search. I can recall no search engines with another term. Google does not appear to show all the linking pages it knows about – perhaps only those with a decent PageRank (a topic for Chapter Two). Yahoo and MSN search have a specialized **linkdomain:** field search term I occasionally find helpful. The linkdomain field uncovers links pointing at any page within a given domain. Lastly, as I write, Google and Yahoo both do not permit searches for multiple links as in **link:google.com link:yahoo.com**. For this kind of triangulation, I use AlltheWeb.

## FURTHER FIELDS AND COMPLEXITY

That about completes how we ask for a title, URL and link field search. Remember, this is just the internet version of the library's author/title, subject and Dewey decimal number search. We will shortly see the URL and link field searches lead to some very sophisticated search techniques indeed.

Global search engines offer a range of fields beyond these three including perhaps language, filetype, topic, anchor text, update date and adult content. These are all valid search tactics and may be important for certain search occasions. I have heard from teachers who find it rewarding to search for .ppt files (Microsoft's power point) because such searches often provide a good overview to a topic. I have heard from lawyers who limit their searches to .pdf files (Adobe's portable document format) because pdf documents tend to be more authoritative. Topic searches like Google's US Government Search, Yahoo's product search and Technorati's Blog search will come back to us in Chapter Six.

Some of these searches we can avoid. Set aside filetype because we will learn in Chapter Four that format is a more powerful concept. Searching by language is simple but usually less helpful than searching a regional search engine. Other fields may be absolutely critical to accomplishing some unique and rare task but we will not need them often.

The real complexity comes when we step beyond our favourite global search engines and follow closely the subtle movements of each global search engine. There is a lot of movement. Words typed twice in a search query means something to Google. MSN search seems to have difficulties

with its URL field search. In July 2006, Google suddenly lost the ability to combine a word and a URL field search in a single search. A search like **jupiter inurl:wikipedia** gave a false answer. A week later, I see it is working again. So arises a complicated chore of teasing out the many distinct differences between search engines; watching as these differences change.

We can deal with this complexity in two ways. Firstly, strive to understand these subtle changes in depth. ResearchBuzz,<sup>2</sup> a weekly newsletter by internet research expert Tara Calishain, addresses this kind of search engine minutiae well. Her newsletter covers such topics as Google's strange date-of-indexing field based on the Julian calendar. (You and I use the Gregorian calendar.) Consider also watching the print publications *Online* and *Searcher by Information Today*<sup>3</sup> as well as *Online Currents*<sup>4</sup> here in Australia. We can search **SearchEngineWatch.com** for this sort of information too.

Alternatively, curtail our avid enthusiasm for all things searchable and reach just for the established search techniques and tactics. If we need something special, only then learn how a particular search tactic works on our favourite search engine.

I doubt many of us need to know more about search engines than what I have just shared with you. They promise to grow more complex with time. Stepping away from some of this complexity makes good sense. I recommend you retreat to the established search tactics of "", -, OR as well as title, URL and link field searches. Recognize there are further fields and specific idiosyncrasies to the many search engines. Now get some practice.

## PRACTICE IN PRECISION

Enclose concepts in quotes. Subtract information foreign to our search. Use field searches to specify information qualities. These are all opportunities for precision. Here are a few examples set in the standard search engine punctuation used by Yahoo and Google.

### **"deep tissue" massage**

Search for webpages for the phrase "deep tissue" and the word massage somewhere on the page. Deep tissue is a concept, a certain style of massage, so we have good reason to use quotes.

### **diabetes -"childhood diabetes"**

Show us pages with the word diabetes but not the phrase ‘childhood diabetes’, which I understand is a different disease.

**intitle:cadbury**

Show us webpages with Cadbury in the title. We can expect this will include the corporate website for the makers of Cadbury Chocolates.

**greenpeace inurl:.au**

Show us webpages with the word Greenpeace but only those found on a webpage with .au in its web address. Thus, show us Australian webpages mentioning Greenpeace. As expected, Greenpeace’s Australian website leads this list.

**university sydney inurl:.edu**

List webpages including the words University and Sydney, with .edu in their web address. This list starts with links to several universities in Sydney.

**inurl:www.ccm.net/~jrsmith/**

Reveal all the webpages the search engine has found in this directory.

**link:stamps.com**

List and reveal the number of webpages linking to **stamps.com**.

**link:patents.ustpo.gov link:patent.gov.uk**

List webpages that link to both the US and UK patent databases.

We will have many more examples as we read further. Just remember, search engine punctuation allows us to ask specific questions. Search engines respond with far more focus. Precision is the second method of finding information with a global search engine. The first, of course, involves throwing a few words at a global search engine, then browsing the first few leads returned; a process commonly known as surfing.

## SURFING IS NOT ENOUGH

The internet is like a seventeenth-century Dutch painting. A small bitten apple in the corner of a picture, upon reflection, suggests the biblical story of Adam and Eve; the idea of sin. A dented pot suggests carelessness. A

half-eaten fig: sensuality. The more we look, the more we reveal, the more we understand.

Internet searching initially appears as a simple topic dominated by the simplest of questions: “What words shall I throw at a search engine today?” Now that we have sketched out a way to be precise with what we ask a search engine, thanks to quotes, minus, OR as well as title, URL and link, we can again confront the simplest of questions: “What words and punctuation shall we throw at a search engine today?”

Sadly, in so summarizing internet searching, we have lost almost everything that is wonderful and beautiful and delicious about the internet. Like a talented chef introducing a novice to their spice shelf – think of the disappointment. By all means cook with pepper. Cook with chili too. We certainly find pepper and chili in some of the finest dishes. However, please recognize that more than spice is needed to turn an egg into a soufflé. Spice is just one element of a grand feast.

Many a simple question can be answered without skill thanks to the internet. Many more can be answered with search engine punctuation. We can get some kind of answer to most questions. Should it take a little longer ... who cares? Should we get a mediocre answer, an untrustworthy answer ... hey, it's the internet! We should expect this.

But wait! Such an attitude is the complete opposite to that of talented internet searching. We are trying to accomplish something grand. A talented searcher draws far more complete answers, far higher quality answers and answers to far more challenging questions in far less time. And we should expect this of a skill like internet searching. Would a novice naturally make the right choices without experience? Has the internet somehow changed the value of experience? Can answers really be found by just whispering a few words to a plastic box?

Only if an exquisite meal is a matter of sprinkling a little spice.

There is something deceptively simple in the image of the internet as a realm we either search like any old database or browse like the shelves of a library. The unspoken image for such an internet is a mass of webpages dumped in a big pile yet searchable all the same. Perhaps the internet is so vast, all we can do is search. We search with luck and time but not skill.

That internet is a mirage – a horrible distortion of the truth. Internet searching is indeed a skill. In addition to search engine punctuation, this skill includes a great deal of library science that at first seems either self-

evident or completely off topic. Later we will start anticipating information, incorporating even more library science as well as sociology. Furthermore, the field of internet searching continues to develop. New techniques and concepts continue to emerge.

Our tools develop too. If we look at this historically, we have been rushing at a maddening pace through so many approaches to finding information.

With this in mind, let us revisit the word ‘surfing’ – that familiar sensation of moving from one site to another hunting for something that interests us. It is a close cousin to reading the newspaper and browsing the library bookshelf. In essence, we seek something of interest without a clear idea of what we seek or where we think we will find it. We search blind. It is one of life’s more rewarding experiences, this grazing on interesting information. Serendipity leads us to many beautiful gemstones. My personal love includes grazing on historical maps and Hubble photographs. Unfortunately, such grazing is not a good way to answer questions. When we have a particular question in mind:

- surfing wastes time
- surfing never tells us when to stop
- surfing rarely leads us to the best information

This is not to say that the key to searching is to know and accurately describe what we seek in advance. Sometimes this approach works. Sometimes this approach is maddeningly frustrating. Let us just recognize that surfing is not the solution. Allow me to explain.

## COUNTRY PROFILES

Suppose we are interested in Afghanistan. We type **afghanistan** into our favourite search engine. If we favour Google, we receive a list of 172 million matches, the top twenty listed for us to browse. Our search engine thoughtfully generates what it calculates as a helpful list but with just our interest in Afghanistan to go on, the search engine must make some very unfortunate assumptions. For example, the search engine must assume we know little about Afghanistan so it generates a list of several general and popular websites. In another setting, in another time, we would reach for a large encyclopedia.

Perhaps we are interested in something specific about Afghanistan. Say we want Afghanistan's vital statistics so to speak: its birth and death rate, its gross national product (GNP) and ethnic mix. We are looking for something called a 'country profile', a kind of standard document that describes a country briefly with statistics and precise descriptions. Country profiles may be familiar to you as books like the World Year Book. They read like the country descriptions found in encyclopedias. Perhaps you have seen the economic synopsis of a country by The Economist magazine.

It turns out country profiles are far more numerous than we probably expect. Many of the largest, most highly respected international organizations constantly update their country profiles and make them publicly available through the internet.

A search of Google for "**country profiles**" **afghanistan** lists some of the most popular of these. Their list includes such standards as the country profiles by the Library of Congress and the US Department of State as well as an extensive list of .com sites publishing something of news or the economy. The list also includes websites that link to country profiles like **corporate-information.com**. If we search a different global search engine, we get a slightly different list, though the very popular CIA World Factbook usually appears near the top of any list.

A directory like the Yahoo Directory could also be a fine place to hunt. Directories are still very useful and respectable research tools. The Yahoo Directory lists some country profiles, though the list is pretty bare. It includes the CIA World Factbook and country profiles by the Library of Congress, US Department of State and several .com sites.

Select any basic search tool and retrieve a similar list of resources that summarize Afghanistan. In this way, we can easily answer a question that could be answered by a World Year Book or a large encyclopedia.

What if we have a more challenging question? Or a question that demands greater depth? About six years ago, I researched country profiles in detail. I sought all the country profiles that existed at the time, listed them, compared them, tossed out the poor ones then crafted the results as an article. I included internet, library and commercial resources as well as other avenues to explore. This article sits at **SpireProject.com/country.htm**.

This article also vividly illuminates the perils of searching. At the time I wrote the article, I found free country profiles from more than forty of the most highly respected organizations in the world including:

#### General Country Profiles:

CIA World Factbook

Country Indicators for Foreign Policy (CIFP)

Organisation for Economic Co-operation and Development (OECD)

UN InfoNation

UN Statistical Division

UNICEF

US Census Department

US Department of State

US Library of Congress

World Bank

#### Travel Advisories from:

Australian Department of Foreign Affairs and Trade

Canadian Department of Foreign Affairs and International Trade

UK Foreign Consular Office

US Department of State

#### Country Health Reports from:

Health Canada

Pan American Health Organization (PAHO)

US Center for Disease Control (CDC)

World Health Organization (WHO)

#### Country Reports on War and Justice from:

Amnesty International

Canadian Department of Foreign Affairs and International Trade

Canadian Forces College

Care Country Profiles

Eldis – Gateway to Development Studies

Human Rights Watch

International Committee of the Red Cross

Initiative on Conflict Resolution and Ethnicity (INCORE)

UN Development Programme (UNDP)

UN High Commissioner for Refugees (UNHCR)

US Committee for Refugees

US Department of State

#### Economic Country Profiles from:

Australian Department of Foreign Affairs and Trade

Commission of the European Union  
Food & Agriculture Organization (FAO)  
International Monetary Fund (IMF)  
New Zealand Trade Development Board  
Organisation for Economic Co-operation and Development (OECD)  
UN Industrial Development Organization (UNIDO)  
US Department of State  
US Embassies  
US Energy Information Administration (EIA)  
US Trade Representative  
World Bank  
World Trade Organization (WTO)

Let your eyes skim over this list. Consider it. This is truly an impressive list. Many of the most significant organizations in our world are included.

Where does surfing lie in this picture? Throw the words **“county profile”** at a search engine and we retrieve a list that includes maybe five of the more famous country profiles listed above. The others are buried too deeply in the internet to surf to easily. Many country profiles are not popular, promoted or otherwise likely to rise to the top of a search engine’s results page. Instead surfing excavates a great many resources that serve as proxies for a good encyclopedia. Many will be .com sites lacking both the depth and authority that all great information possesses. Surfing would surely miss the huge published tomes from the OECD (Organisation for Economic Co-operation and Development) each running to over eighty pages of quality economic forecasting. Instead we reach out to a summary by a news organization that appears to miss economic commentary entirely. Oh, the specific documents we miss have changed but the fact we miss them remains. Great resources fill the internet. Surfing leads us to just a few.

Let us leave Afghanistan for a moment and think of something very specific. Sometimes information is invisible to all but a couple of search tools. Sometimes information is simply not online.

Say we seek the corporate website to a company registered in the United Kingdom. Surfing would suggest we have merely to keep looking and we will find their website. If not this results page, then the next. If not this search engine, then the next. If this website does not answer our question, try another. No step says, “Stop! Give up! Time to leave.” Surfing, unlike

searching, just does not address this possibility. We quit when we get bored or frustrated.

When searching well, we build a different relationship with information. Rather than just browse what is offered, we also work with notions of what else is out there. We anticipate our destination. If we have little chance of finding answers, abundant clues will tell us so. While surfing, we do not notice these clues.

Back to Afghanistan. Suppose we now have a specific question in mind. We are writing an essay on the evils of the Taliban and we are concerned that so much of our experience comes by way of the US news media. To correct any potential bias we need some additional proof that the Taliban really were bad people. We want convincing proof hopefully from a non-news-media source.

It seems a simple enough task until we get into it. If we roam the internet moving from one page to another, hunting for something that sounds reliable and trustworthy, we will probably find it. We will stumble upon something we can build a case for being unbiased and supportive of our conclusion. This is not proof.

We find proof in a paper published by Human Rights Watch that documents a civilian massacre perpetrated by Taliban forces.

**[www.hrw.org/reports/2001/afghanistan/afghan101-03.htm](http://www.hrw.org/reports/2001/afghanistan/afghan101-03.htm)**

The publisher of this document, Human Rights Watch, is widely respected and experienced in documenting human rights violations. This document arises from first-hand interviews with those affected. A witness list is attached. It is perhaps the highest quality information of this kind short of being there. And it is completely separate from the US news media, the potential bias we wish to counter.

Unfortunately, when I first looked, this particular Human Rights Watch report was not indexed by AlltheWeb. Nor by AltaVista. Google did index the page but we would never find it on Google unless we searched for the word 'Yakaolang', the site of the massacre. And why would we search for Yakaolang? Must we already know a document exists to find it?

Records of the Yakaolang massacre have grown more prominent with time. If it is easier to find today, years after the event, though not because indexes have grown more comprehensive. Though larger, search engine

indexes may not have grown more comprehensive. We may need to expect that information not yet prominent will simply not be indexed until later.

Surfing has us stumble upon this page using search tools that initially ignore it! If they do not ignore it, they at least do not recognize its importance. In the example of country profiles, we surf through a list of prominent sites looking for sites without prominence – a rather stupid endeavor if we think about it. Surfing tells us we will find our proof, our not yet widely respected or recognized proof, thanks to providence, serendipity and accident – three techniques we simply cannot trust to deliver quality answers. This is why surfing rarely leads us to the best information.

We know this. Anyone wandering the internet today knows something is amiss. We know because we feel frustration when we search. We waste time. We do not know when to stop. We only occasionally get the best information. Of course we are frustrated.

Let this frustration drive us to move beyond surfing. As soon as we feel frustrated, as when we ask complicated and challenging questions, we should reach for a different arsenal. We should search in a different manner. Frustration is in fact one of the clues to listen for. It is our friend. When we become frustrated, stop and search another way.

\* \* \*

## A DIVERSION

*The mid-morning mist still held its grip on the valley below. The cold stones had not yet lost their moisture. A small boy of twelve sat quietly in the window alcove on the second floor of the castle tower as he looked south to the hills speckled with grey-white sheep. As the morning chill tried gently to crawl into the blanket wrapped tightly around him, the cold stones he sat upon chilled him with a more brutal directness.*

*With a long shiver and a sigh, Albert stood, then moved back from the cold world beyond the window. He quietly retreated to an adjoining room warmed with the help of aging tapestries and a fire just of embers from the night before.*

*On a cold January morn, in the year 1195, a young French boy named Albert, second son of the regional magistrate for Toulouse, quietly decided his life's work would be as a knight. Knighting as a career path was well regarded in the Moyen Âge; the Middle Ages. His soft downy hair, small*

*hands and skinny frame betrayed his youth but he had connections and the support of a father keen to promote justice in the realm. It was a fine arrangement. Albert would settle into the task of learning to be a knight.*

*He had surprisingly much to learn too. Certainly Albert needed a great deal of technical skill in the use of weapons but the City of Toulouse also expected its knights to be religiously pure and relatively educated in the fields of the day. A knight was not only expected to stand for justice and equality. He was expected to recognize the just and righteous path.*

*Albert had a great deal to learn.*

\* \* \*

We too have a journey ahead of us: a journey filled with complexity and confusion. To ease this journey we shall follow this fictional Albert through a time in the Middle Ages when his own humble and simple journey became as complex and confusing as our own. Perhaps Albert's story will help us periodically lift our attention to the grander picture; to the art and insight that infuses the best of search.

Internet searching is not so very difficult. Most likely, you can already find non-internet information easily enough. You can find a book in a library and ask directions from a stranger. We just need to extend these skills to cover internet information as well. The struggle ahead is not in grasping a vast and unfamiliar field of expertise. We struggle instead to understand how skills and techniques we already know and use elsewhere apply on the internet as well. We need only clarity.

What should become apparent quickly in this quest, so I will alert you now, is that searching the web is about being aware of many aspects of information we frequently overlook. For example, say we ask a passer-by for directions. Did he just fling his hand in a seemingly random direction? Did he look confused and lost himself? Was that a bottle of cheap wine in his left hand? We look for such clues. Such clues have a bearing on the value of his advice.

With talented internet searching, we use the same tools and ask similar questions as less experienced searchers but we ask in a way that reveals more about the information involved. Every aspect of information – the web address, publisher, author, context, format, pages that link to the information, the intended purpose of the information, how the publisher

justifies their efforts – everything comes to have much more meaning than we usually attribute.

We are helped in this journey by the insights of no less than three disciplines: computer science, library science and sociology. We can explore explanations and move freely among all three. Thus, we will craft historical explanations. We will explore patents, newspapers and books. We will delve into how global search engines rank their results. We will explore a variety of publishing models and consider the future of the internet in view of the tension between capitalism and utopianism. In short, we will wander all over the place as we aim for effective use of internet information.

Our young French boy, Albert, was a simple soul. In an era bleak by today's measure, he chose to be a knight – a noble profession with generous opportunities to do good in a world of much hostility and fear.

Searching is not a profession. These days, searching is an element of so many professions. However, librarians have perhaps the closest ties to searching. Certainly, librarians consider the social importance of their work, worry about issues of access and are employed to help patrons find their way through the confusing and unfamiliar world of information. This always sounded noble to me.

For several years the librarian profession drifted, uncertain of its role in an internet empowered society. It seemed to some that libraries and library science had become passé thanks to the internet. Perhaps we do not need libraries and librarians as much as before the internet arrived. I will take this opportunity to dispel more of this uncertainty. Library buildings stocked with aging books may lose some of their luster but one of the pillars of this book is that library science is vital to the effective use of the internet. Many existing advances first emerged in libraries decades ago. Many future advances born in library science are already in the pipeline.

Library science is not the whole picture though. We must also learn some fairly arcane computer technology. Learn about the bookmarklet and the domain name. Juggle windows. Use shortcut keys to speed us on our way. Our quest for pattern and structure also takes us to investigate capitalism and academic recognition. We will reveal a more holistic picture of the internet's role in the flow of information. Why do people publish? Where would certain kinds of information be published? Who publishes that kind of information most successfully?

If the internet is a galaxy, this galaxy of ours has a history and a future evolving from this history. How the internet has evolved fascinates me. It is surprisingly understandable too.

Library science, computer technology and sociology: so much ground lies before us. So much insight to consider. So much to help us make better use of internet information. Before we wander too far, however, I wish to introduce a searcher's most trusted ally: the elevated vista.

\* \* \*

## ENGAGING THE WORLD OF INFORMATION

*Anyone can hold a sword. Anyone can stride into battle with a weapon in hand and try to strike the enemy. Connecting is entirely a different matter.*

*Albert started his lessons not with the sword but with the pike – a long solid stick with a sharp blade at one end. Albert was to hold the pike firmly in his hands, stand in formation with 15 other soldiers, four to a line, four deep, then run at the enemy. If the pikemen worked effectively as a team, the enemy soldier would meet four sets of sharp blades before they could begin to slash at the first pikeman.*

*Of course, the best defense against pikemen is more pikemen ... with longer pikes. The ancient Greeks under Alexander the Great used pikes as long as twenty feet. They decimated the troops of the great Persian King Darius in this way, literally running through the enemy lines.*

*There are different pikes too. Some have sharp hooks on the end for unseating a knight from a horse. Some have blades for slicing. The pikes Albert worked with were heavy, laborious weapons but they could be very murderous. Albert studied hard.*

\* \* \*

War is a little more complicated than grabbing a pike and racing at an enemy. A search is too. Grabbing the first available weapon, a global search engine, then thrusting words at it is just one of many approaches to searching the internet. If we did little else, we would often feel frustrated.

Let us extend our reach. Let us look beyond the recommendations our chosen search engine offers us and consider the view. Let us interact with the world of information.

Whenever we do a search from now on, the first item I want you to notice is the number of matches or hits reported by the search engine. Whether this number is five or five million, this number answers several important questions:

1\_ Did we do something wrong?

A very small or very large number indicates a spelling mistake or a problem with how we punctuate a search.

2\_ Can we refine our search further?

A large number of matches invites us to ask a more specific question.

3\_ How much information is there on this topic?

The number of matches indicates the size of the reservoir of information we have to draw from.

Lift our view to the horizon. Look at one page of results and see the world of information.

Suppose we work for a government agency looking after the interests of seniors. Our task today: uncover the issues involved in seniors using the internet. We decide our keyword is ‘aging’ and a simple internet search for **aging** returns a large number of matches – 212 million matches as of mid 2006 on Google. However, as we restrict our interest to just Australia (by typing **aging inurl:.au**) the number drops to 804 thousand. From over 200 million to less than a million. Seems strange? Australia generally accounts for around 4% of all Internet content, not half a percent as suggested here. This may be our only hint that in Australia, the word ‘aging’ is spelled ‘ageing’. A search for **aging OR ageing inurl:.au** returns 4.5 million matches, adding 3.7 million pages to our list.

Let us try again. Say we wonder if it is possible to search the internet as a career – perhaps as a commercial researcher. We search Google for **commercial database OR commercial research** and receive 216 million matches. Far too many, I should think? Perhaps something is wrong with our search query?

Do you see it? We have used **OR** incorrectly. We have asked for the word **commercial**, then the word **database OR commercial**, then the word **research**. That is not a specific search at all. I think we meant to type: **“commercial database” OR “commercial research”** remembering to add the quotes. Look at the horizon. Notice it is not where it should be.

Say we visit the website of our state library as we hunt for a book on research techniques. A title search for **research** returns a list of over four thousand books. Shall we craft a more specific request? We could add more words or specify a particular subject we are interested in. For instance, market research does not interest us today so perhaps we can search in a way that reflects this. The number of books on research tells us we can refine our search further. It works the same way on the internet.

This reminds me of a fine technique used in commercial article searches. When searching a commercial-quality database, keep limiting a search until we build a list that returns only as many records as we are willing to consider – usually about fifty. Now browse this list. Read the titles. Notice the publications. Consider the length of each article. From this list, select three to five articles to read or five to ten articles if we must find them in a nearby library since some will be unavailable. This tactic works exceptionally well with commercial-quality article databases like those in university libraries and those available through database retailers like LexisNexis and Dialog.

Let us now apply this approach on the internet. Craft a specific search. Refine the search so it generates perhaps fifty matches. Now browse this list. Select several likely candidates worth perusing. The criteria we use to select peruse-worthy information will be discussed later in this book but briefly, it involves matching clues from the web address with where we anticipate our answers will reside. Approaching the internet in this way is the perfect foil to search engines that offer answers that seem far too general and prominent.

When we want a specific search, we focus. At first glance, this can mean we add words to our search query until we have something very very specific. Better to add punctuation. Ask that words appear together as a specific concept. Change **artificial intelligence** to **“artificial intelligence”**. Should a word be in the title? Can we discard information on market research? Can we limit our search to a particular type of resource: perhaps a certain country? This kind of thinking leads to a much more rewarding search than just adding more words.

We also build a specific search as a process. As we build, we watch the number of matches. It tells us how much further we can refine our search. We will usually search several times before we stop and read the list of results. A good search gradually takes shape.

We type **shakespeare**  
then **shakespeare unconditional love**  
then **shakespeare unconditional love romeo**  
then **shakespeare “unconditional love” romeo**

Remember, the number of matches tell us something of the quantity of internet information available to us. Suppose our special friend is coming to dinner next week and we want to cook a favourite childhood recipe. We search for **“brazil nut cake”**, her favourite, and find over a hundred recipes indexed by Google. Just five of these recipes do not include the ingredient ‘flour’ to which our friend is allergic.

These numbers have meaning. These numbers suggest our search is a challenging search – a search that ranking technologies cannot assist. The recipe we seek may not be published in an easy-to-reach location. We may need to move beyond the global search engines. My thoughts turn to various recipe databases and cooking discussion group archives. My thoughts turn to other places where recipes pool.

Say our hacker friend talks about smurfing – a denial of service attack that can take down a website and land us behind bars. Shall we find the software that does this? A search for **smurfing software** returns just forty-seven matches, many of them glossaries.

Once again, these numbers have meaning. This will be a challenging search. Ranking technologies will not help us. We may need to look elsewhere, in more private locations.

Match numbers also tell us something of the awareness of information on a topic. Sometimes this alone is important. In a search for **“David Novak” “spire project”** we are given a number of matches that directly reflects the public awareness of my work on the internet. A similar popularity number emerges from a link search as in **link:spireproject.com**. Websites with more links have been promoted more effectively, have been on the internet longer and have demonstrated an ability to attract interest. Such sites often have better information, an assumption we will explore further in Chapter Two.

When we ask a specific question, the number of matches we encounter tells us something. It tells us if we are on the right track. It tells if we made a mistake. It tells us if we have found the right words – words that someone in the industry would use. A search for **staff loyalty**, for example, leads to many resources in business but very few in nursing. Why? Because nursing

literature uses a different term. The literature does not describe it as ‘staff loyalty’. I think to look more closely only because we found so few matches.

When we discuss feedback research later in this book, the elevated vista tells us even more but we will never hear what is being said if we don’t listen! Glimpse the elevated vista in the number of matches returned. Savor this momentary view.

## MY CHOICE OF SEARCH ENGINE

*Slice, Parry, Thrust, Lunge. While the pike relies on strength, a sword depends on skill. At his father’s insistence, a mentor started to teach Albert footwork.*

*Swordplay is a dance: forward, back, side to side. We constantly vary our momentum and balance. Albert thought he understood footwork. He tried to move more quickly; to improve his balance. It was frustrating though, for try as he would, his sword skills scarcely improved.*

*Albert had missed something. More than keeping his own balance, Albert had to judge the footwork of his opponent too. Attack when the opponent has least control over their movement. Lunge when the opponent steps forward. Step to the side as the opponent thrusts. Slice as their side becomes vulnerable. Swordplay is a deadly dance for two. Footwork establishes balance. Footwork creates opportunities to attack.*

\* \* \*

The two global search engines I use are Google and Yahoo’s AlltheWeb, though I may shortly change to Google and Yahoo. My choice rests on what I need to build a fine specific search: good field searching and database size.

I declare my preference not to suggest others are not important or to ask you to change your preference. I wish only to explain why I use these search engines and not others. Perhaps this will help you choose what is right for you. I cannot advise you further because specifics change too quickly for a book to address and comparing search engines never fully captured my interest.

Google originally attained fame for introducing a ranking technology built on link behavior. This approach to ranking has since been enhanced and implemented in all global search engines. Google now deserves our attention and praise because of its size and the flexibility of its field searches.

Size is a fuzzy issue now. Up from eight billion records in late 2005, Google is now much larger but of an undisclosed size. A similar story covers the other search engines. I find the views of Danny Sullivan of SearchEngineWatch persuasive when he describes how we cannot easily compare size across search engines anymore and how counts do not measure comprehensiveness.<sup>5</sup> However, relative size remains a reason why I look to Google. When we search in a specific manner, size matters, at least in theory. We want to reach for the largest search engine near at hand. Unfortunately, it can be hard to decide which is largest.

Here is a simple demonstration based on typing “**spire project**” search on April 4<sup>th</sup>, 2006:

Google:	18,800 records mentioned, 742 displayed
AlltheWeb:	4,710 records mentioned, 1100 displayed
MSN search	2,925 records mentioned, 450 displayed
Yahoo	5,620 links recorded, 1000 displayed

Repeating this search on January 15<sup>th</sup>, 2007 sees these numbers fall some but still shows a similar margin and a similar gap between mentioned links and those available for display.

Google:	12,300 records mentioned, 704 displayed
Yahoo	2,590 records mentioned, 1000 displayed

Yet size accounts for just half my reasoning. How flexible is the search technology? Unfortunately, Google is clumsy with some of its search techniques. All three top global search engines are clumsy with plurals but by using OR, we get around this. Google also does not display many results in a link search, so I use an alternative – my old favourite AlltheWeb.

A link search for **spireproject.com** on January 15<sup>th</sup>, 2007, retrieves:

Google:	64 links recorded, 76 displayed
AlltheWeb:	655 & 180 recorded, 304 displayed
MSN search	1,030 links recorded, 450 displayed
Yahoo	474 & 263 links recorded, 440 displayed

The second numbers emerge when we include **www.spireproject.com** in our search. To complicate matters further, Yahoo has **linkdomain:** (a specialty link field search) and numbers like those just listed change quickly over time. It is enough to drive one crazy. Once we get our minds around the fact that match numbers are estimates that can change mid-search, that a few hundred more matches can be found in a pinch and that some recorded links can never be seen while other links were never visited when indexed, we get a taste of the wonderful clarity enjoyed by global search engine observers.

With sanity we can say Google is not strong on providing links at this moment – so I use another search engine for that purpose. Google has other weaknesses too. At this time we cannot use the link field search to triangulate related information. Google has a field for the date of indexing but it is based on the number of days since noon, January 1<sup>st</sup>, 4713 BC. Don't ask. Don't even think to ask. A rough index-date search appears on Google's advanced search page and Tara Calishain and Rael Dornfest describe several script-based solutions in their book, *Google Hacks*.<sup>6</sup>

Google is responsible for maintaining a lovely database of newsgroup discussion now called Google Groups. Google's image search is very large. Google's news search is promising too. I love their support for significant internet resources but I consider these side databases as completely different and distinct from the Google search engine. I do not let such side databases influence my choice of search engine, for reasons that will become evident in Chapter Five.

In summary, I start with Google unless I have reason to start elsewhere. I start with Google because I am familiar and satisfied with their search engine punctuation. I occasionally wonder if it is time to change; time to favour another search engine.

Whether Google deserves your attention or not, do take the pressure off the constant quest to compare search engines. Consider:

- 1\_ We need a large search engine.
- 2\_ We need a decent URL field search.
- 3\_ We should move freely from our favourite search engine to other search tools for search tasks they do better.

Make sure we have the required tools nearby, get familiar with them, then get on with learning how to make searches more revealing and

rewarding. Frankly, we do not need that many global search engines anyway. If you love another, fine, as long as it has good field searching and is big.

In terms of the all-important rivalry between the global search engines, I am particularly mindful of Yahoo's experience and Microsoft's efforts. I see no reason to believe either firm cannot produce a superior search engine. I see many reasons why we would not realize they had developed a better search engine already. In purchasing AltaVista and AlltheWeb, Yahoo acquired most of the internet's best search interfaces. AltaVista allows for NEAR and was the first big search engine to offer brackets and true truncation. However, I suspect the search interface is not such a significant obstacle to making a great search engine. Remember, much of this technology was worked out in the commercial information world and implemented in commercial databases decades ago. The future rivalry between leading global search engines will be monumentally important to them, I am sure, yet of limited significance to us.

Before we proceed, let me confess one point of far more significance: the popular misconception that search engines index everything on the internet. This is misleading and very wrong. Throughout internet history, all the leading search tools have made similar claims. Now that we no longer have even rough estimates of the size of our search engines, we will surely fall into this trap again.

How much of the internet is indexed by our favourite search engine? It is very very hard to say. Perhaps ten percent. Perhaps twenty. Certainly not fifty percent or eighty.

Just how much is missing largely depends on what we mean by being 'on' the internet. Older estimates of the internet's size range from ten to three hundred billion records, growing at who-knows-what rate. Google has grown from a claimed two billion records in June 2002 to eight billion records in November 2004 to a suggested twenty billion in September 2005. Given that the sheer size of the internet, its rate of growth is probably slowing (growing but doubling less quickly). Given that the latest round of search engine size wars have indexes growing faster than before, perhaps we are closing the gap. Perhaps.

Against this conclusion we must weigh several discordant notes. Several studies call into question the claimed size of these databases. Database numbers have in the past included unindexed, merely referenced material.

Wild claims like Google's statement in November 2005 stating it was three times the size of any competitor seems implausible.<sup>7</sup> Quoted index sizes are not what I would consider good information. However, we have equally poor information about the size of the internet.

One approach to this confusion is to focus on the information world from which internet information is drawn. Do not underestimate the size of the world of information that surrounds us. It is vastly larger than the internet and if the internet is not far beyond a hundred billion records by now, this is only because information publishers have not found ways to justify publishing more, more swiftly. We will discuss this further in Chapter Nine. This means that even if search engine databases do incorporate much of the internet, they cover little of the information world around us.

Our question of coverage remains unanswered – an unhelpful conclusion but one I cannot avoid.

Will search engines continue to grow more swiftly than the internet? The costs of computer memory and computing power are falling and publishing rewards are falling as well. We can hope. However, if I am right and coverage hangs around ten to twenty percent for the next five years, then do ask yourself, "How could I possibly find information not indexed by a global search engine?" We have problems enough making the search engines cough up the information they do contain. How do we reach beyond them?

Until we can answer this question, we have not truly touched the heart of internet searching. We are bound to our search engines, encumbered by every bias they display. Eventually we will reach beyond them and in the process achieve a far more realistic and rewarding relationship with search engines and our world of information.

Let us first just recognize that we can be very specific with global search engines. Punctuation is the key. This is a first step to a better search. The next step is prominence.

## Chapter Two

---

## PROMINENCE

*Displaying a particularly fine mix of daring and caution during a group training battle, Albert got badly clubbed. Too quickly, his inexperience showed and his head hurt terribly for it. Afterwards, his ever-watchful Captain approached and offered words of encouragement. This greatly relieved Albert and his spirits finally began to lift.*

*Fame rested easily on the shoulders of his Captain. Citizens of Toulouse looked up to him and respected his wishes. He had only to ask and doors would open, gifts would be offered, peace would be imposed. Albert had none of this. In comparison he felt so ineffective.*

*Two days after the training battle and back in Toulouse, the Captain sent Albert on a simple errand. Seated grandly, enjoying a mid-morning drink, the Captain's peace and tranquility was disturbed when a loud argument broke out nearby. Albert was told to calm the disagreement. Restore peace and quiet. A young boy of barely fifteen whom no one respected, Albert was told to intervene.*

*Albert waited and thought. Timing would help. The argument rose once more in pitch, Albert walked straight to them, then boldly interrupted the two shouting gentlemen. He said four words, turned, pointed to his Captain, then ushered them to a nearby ale house where he bought them a drink. The ever-watchful Captain sat once more in peace, impressed.*

\* \* \*

Prominence is fame. Public awareness. Whether popular or notorious, we are discussing a central feature of public life. Some of us have a fine soapbox with which to express our views while most of us have little influence over events and public perceptions. Those who host TV shows and write newspaper columns are blessed by prominence. They are known. Their views are heard. They have an opportunity and perhaps the power to mold the thoughts and actions of others. While this power is different from the power to decide as given to elected officials and corporate boards,

prominent people are empowered simply because they have our attention. Their views have an audience.

Prominence invades the internet too. We can talk about information having prominence. Prominent information is known and read. It has traffic, recognition and influence. Since internet users rely so heavily on the global search engines to find information, internet prominence ties tightly to search engine ranking. Search engines offer the more prominent information first.

We can measure internet prominence in about five ways:

1\_ Count the number of webpages that link to a given page. More links usually means more popularity and presumably, more traffic, audience and influence.

2\_ Judge the significance of the organizations linking to or describing a website. When government agency websites, newspapers and peer experts mention a project, it suggests greater significance, audience and influence.

3\_ The Google toolbar has a small tab that displays PageRank. PageRank is a number from 0 to 10 that describes how prominent Google considers a webpage. Google uses this number as one of many factors in ranking webpages. [To install the Google toolbar, simply search for **google toolbar** since we want the most prominent one.]



4\_ Traffic numbers, not hits but visits, also give an indication of prominence. Hits only distantly relate to public awareness, as described in the glossary. Hits measure the activity of the computer serving a website. It is the number of pages or images requested of a computer, a number that varies with the number of images found on a webpage. Visits, on the other hand, correspond to actual individuals looking through a website. One visitor may look through many pages, request dozens of images and trigger over a hundred hits. When considering traffic, only consider visitor counts. More visitors suggest more attention and more prominence.

5\_ Lastly, as a crude measurement, notice a website's position on a search engine results page. First among ten thousand suggests greater prominence, traffic and more influence than websites listed lower on the list.

In many ways, prominence resembles business goodwill. It is revealed in public awareness and in the awareness and patronage of significant voices in our community: the wealthy, the informed and the popular. To be clear though, prominence is the notion of public awareness, not one of these measurements. We may measure prominence using link numbers and PageRank but prominence is not equivalent to PageRank or visit numbers. Many a marketing firms would do well to remember this distinction. Like CD sales to pop star fame, one indicates the other but they are not the same.

Prominence is also relative. No matter how famous we are, if another holds more fame we are relatively less known and less influential. This may be less important on the internet where near famous is often good enough but to clearly appreciate a website's prominence, compare it to the prominence of competing and comparable sites.

Once we understand this notion of prominence, we will begin to notice prominence everywhere on the internet. We arrive at a website by asking a simple question and clicking the first match given by a search engine. We rightly presume the site has prominence because of how we found it. We reach for the Yahoo Directory and know all the sites listed have prominence. With the Google Toolbar installed, we glance at the tab that indicates PageRank. Oh, this page has prominence. It has a PageRank of six. To look more closely at prominence, we retrieve a list of links to the page we are on, notice the number of links, then peruse these links for a feel of the types of organizations linking to the page that interests us. Oh, this page earned links from several government departments and many private law firms. It has prominence.

Later in this book, I will show you a bookmarklet (something similar to a bookmark) that lets us retrieve a list of inbound links at a single click. It is a little thing but helpful. I will also show you how to juggle windows so a good look at prominence will not interrupt the flow of our search. Even in detail, noticing prominence will take only a few seconds.

What influences prominence? How do we get prominence? Time is obviously a factor. In so far as prominence is reflected in the number of links pointing to a given webpage, the longer a webpage is on the internet,

the more people have the opportunity to find and link. Promotion also helps. While I understand paid web advertising usually does not count as links, any kind of promotion introduces a webpage to a larger audience and helps entice additional links. Original appreciated content helps too. We want an audience thrilled, or at least pleasantly surprised, by our content. We want a memorable web address and a colourful, memorable visitor experience.

We also want more traditional promotion like a good newspaper article and a well-known customer bragging about our excellent service. We want name recognition, choice affiliations and the appearance of significance. In short, we want all the benefits that traditional public relations and promotion offers the non-internet world. If this sounds like a book being judged as much by the quality of its cover as its contents, then you have the right idea. Prominence has its imperfections.

We use this concept of prominence in two ways. Firstly, prominence is an asset belonging to the web address that attracts our attention. As an asset, it has a monetary value. This view of prominence directs how we promote and market information. Most internet users spend most of their time in the prominent portion of the internet so projects driven by a need for attention must generate or acquire prominence.

Secondly, prominence describes a feature of internet information. Prominent information has unique characteristics we may desire and appreciate. Perhaps we seek only prominent information to answer our question. Perhaps we want to hear the views of those with the loudest voices. This time, prominence belongs not to the address but to the information that earns the attention.

## PROMINENCE AS AN ASSET

Anyone marketing on the internet today quickly learns that prominence is a primary ingredient to achieving anything on the internet. It is an asset. We need this asset if we wish to influence the internet world. If we do not have this asset, we must buy it or borrow it. Albert's solution to calming an argument was simple: he represented himself as doing the bidding of his Captain. He borrowed his Captain's prominence.

This was not always necessary. In earlier times (and still in certain sectors of the internet), prominence would flow easily to the deserving. Prominence depended only on content value. Write an important FAQ and people would find, read and tell others without any further intervention. Write excellent software, then simply place it in a popular, free software archive. This was enough to introduce it to the world and spawn the attention it deserved. As the internet matures, however, the need for awareness grows. Little can be accomplished today without it.

Vocalist Bernadette Robinson, whose daughter attends school with mine, lamented one day how her newly developed website appeared so far down the search engines results page. She feared clients wishing to hire her vocal talents would find their way first to the website of a speakers bureau and not notice her own website. This has a financial sting since a speakers bureau would simply call her, arrange an event, then take a sizable commission.

Of course, Bernadette's new website had no prominence. No one linked to her page. According to ranking algorithms, it belongs near the bottom of a list of sites describing Bernadette Robinson. After all, it is not a popular page and no popular page mentions it. Informing the search engines of the website's existence and asking them to index the website does not change the fact it has no prominence. Yes, this completely overlooks the fact that this is her 'official page'; that this page leads directly to her as an individual. This fact simply does not enter into the ranking equation.

As a solution, I add a single link from the bottom of **SpireProject.com** pointing to **BernadetteRobinson.com**. The webpage at **SpireProject.com** has prominence – it has a PageRank of six and numerous links from university and library websites. As I write, the prominence I lend her by linking is enough to place her website third on a Google search for her name.

Other factors are at work in search engine ranking but Bernadette's difficulties stem from not having sufficient internet prominence to be heard. Even with her name in the title, even as her official website, she needs prominence to reach the people seeking her.

As a second example, following a lecture I delivered last year to a class studying public relations, I spoke with a student considering a job with a search engine optimization firm. Now, I don't appreciate search engine optimization much. Too many operators are ill informed and too slick for

my liking. However, there is a need for these services and a certain future for the industry. With this in mind, I asked the student, “Is the firm prominent?” A decent track record and a healthy internet prominence would indicate to me a greater likelihood of succeeding in this industry and therefore more opportunity for a fresh public relations graduate. I advised her against working for a new, unproven business. Prominence would tell us if the firm was a recent or established player in this industry.

In this case, relative prominence speaks of corporate strength. Any fly-by-night operator can make a flashy website but few can create a meaningful nexus of links, recognition and perceived importance.

As an aside, if you ever use prominence in business, make sure you use relative prominence and always glance at the list of references for the appropriateness of their endorsements. I occasionally notice seminar speaker websites with a good number of links suggesting respect. Look closer, however, and I see few links come from appropriate sources. Too many links are simply self-made garbage.

Prominence deserves a book of its own. It has diverse applications from credit management to web promotion to web design. Internet marketing focuses on some aspects of prominence but often overlooks or diminishes the need to develop a footpath beyond the search engines. Fix ailing links. Compare, contrast and mine the footpaths of comparable sites. Now reach beyond links to other types of endorsements. This topic is not the purpose of this book so I will leave it for another day. Some guidance is present in a white paper at [SpireProject.com/white.htm](http://SpireProject.com/white.htm) but suffice to say a researcher’s perspective exists and it differs from the perspective usually associated with internet marketing.

The need for prominence in business will return to us when we discuss how it affects the publication process. The commercial model is only one of three ways to publish information; a very significant model that depends on achieving sufficient prominence to be heard, then capitalizing on this attention. Authors and organizations publishing in this way but unable to achieve sufficient relative prominence fail and often fail miserably. This dilemma means that while the internet is often portrayed as a free or near-free medium to publish in, those who need or seek attention must generate or purchase this asset called prominence to be heard. The internet is not a free medium for them at all.

## PROMINENCE AS A TRAIT

Prominent information has something that non-prominent information lacks – primarily a loud voice and the presumption of significance. As we wander the internet, we may prefer to dwell on prominent resources. We may seek prominent information. Perhaps we wish to hear only from influential and prominent voices. Perhaps we want to download only the most famous Google toolbar or visit only the most prominent astronomy picture archives. Let us now discuss prominence as a trait.

When I seek the experience of comparable speakers who discuss the internet, I want to hear most from speakers who are acknowledged experts. While in the past ‘acknowledged’ may have meant ‘published’ and specifically ‘published with a famous book’, on internet topics such a restriction is too brutal. Many internet experts do not bother to publish journal articles. I publish only the occasional article myself.

Prominence is the answer. I look for speakers with prominent websites. If a suggested colleague publishes a website with a PageRank of six, I will listen with more attention than I would if the colleague has a PageRank of two. Similarly, say a colleague publishes a page that has earned links from the Yahoo Directory, the Open Directory Project (ODP) and several university websites. This colleague has earned my attention. I may quickly abandon their website once I discover it is aimed at primary school students but the suggested significance is sufficient to earn my initial attention.

This brings to mind one of the least enjoyable aspects of publishing on the internet: plagiarism. The second time I encountered gross copying of my website involved a graduate of library science who lives in India. For several years I was unable to locate a valid email address with which to demand the pages be removed. I eventually did reach the person involved and he apologized, saying he did not know the material had become publicly indexed.

That was the trouble of course. Yes, he doctored a copy of my text, then replaced my name with his as author. But so what if it remained relatively unknown? Unfortunately, his website earned a listing in the Open Directory Project under Computers: Internet: Searching: Help and Tutorials. Yes, the same directory page listing my Spire Project and Information Research FAQ once listed an almost exact copy of my work supposedly published by a library studies graduate in India.

The doctored website attained a level of prominence that leant it significance and some authority. I fear it also tossed my own authorship of this material in doubt. A reader who visits my website second could well conclude that I copied the material from its Indian author. And why not believe so? A library studies graduate listed in a prominent directory sounds reputable.

Herein lies my solution. I first published an article describing the infringement in detail.<sup>8</sup> I next used the article to have the infringing pages stripped of its Open Directory Project listing. Essentially, I tore the prominence from the doctored webpage. As it returned to relative anonymity, the copy no longer warranted my concern. I do not fear plagiarism. It is a compliment of sorts. I greatly fear plagiarism married to prominence.

As an aside, my article about the infringement has enough prominence to be noticed, as indeed are these words here. I leave a persistent embarrassment that the event occurred – perhaps more than my Indian fan deserves. In the future, I fear savvy business strategists will use similar tactics to intentionally tarnish the internet reputations of competitors. Internet reputation is not often discussed or even recognized as something of value. The key response to attacks of this nature involves publishing a rebuttal on a page with prominence that directly links to and mirrors the title and text of the page in question.

I saw this executed beautifully by the British Wind Energy Association in their response to a rather biased publication by the Country Guardian titled: The Case Against Wind ‘farms’.<sup>9</sup> The very similarly titled: BWEA Corrects Some Misconceptions In The Case Against Wind Farms,<sup>10</sup> further mirrored much of the text of the Country Guardian article. Gifted with prominence, their rebuttal appears close to the Country Guardian article in many searches. Of course, this avenue is unavailable to those without prominence to spare. Those without a voice are more defenseless to misrepresentation and plagiarism. Oh, the horror. The horror.

Back to the topic of prominence as a trait. Not long ago I received an invitation to speak in southern England next time I travel there. The invitation came from a gentleman working for the local council but something in his letter suggested he had talent of his own so I searched for some background on him. His email address led to the local council’s website but I was unsettled to find just two pages mention his name, both

only in passing. Someone with talent would have more exposure. They would have more prominence. Did this invitation come from a novice?

The name was too common to search directly so to find my answer I added a geographical marker – the name of the city where he works. Indeed, he was until recently an independent business advisor with expertise in this field. I surmised he only recently stepped into the government post. Finding the prominence I had suspected reassures me that the invitation is heartfelt and valid from someone who understands what I try to say. Without this evidence of prominence, of links and discussion and advice mentioning his name, I may well conclude the offer was made from someone without experience in the field and given without much thought.

Do you see how prominence entwines with the notions of trust and apparent significance? Prominence obviously has a role in quality assessment – the topic for the next chapter. However, let us first look at prominence as the search engines consider it. Search engines used in a blunt manner use prominence as a proxy for importance.

By ‘blunt’ I mean a simple search, the tossing a few words at a search engine. A blunt search leads to ten thousand matches or more. In a blunt search, we look at only a few of the many qualifying matches so what we see is heavily dependent on prominence.

Ask ourselves this question: “Are we seeking a prominent resource?” If the answer is yes, then we want the assistance of tools that lead us to prominent resources. We want to search a global search engine in a general blunt manner. We want to visit global directories.

We want to use these tools because they depend on prominence to filter information. If our answer is no, if the information we seek is unlikely to be prominent, then we will regret staying with tools that direct us towards prominence. We want to move beyond the prominent portion of the internet.

To better understand this idea, let us contrast prominence with the closely related notion of importance.

## IMPORTANCE

Something important is something we value. For an internet search this primarily means valuable content. We set the criteria by which information is judged important. Perhaps information must be recent. Perhaps

comprehensive. Perhaps definitive or influential or popular. Our criteria changes with the questions we ask. What is important, what is significant, depends on what we need to answer our question.

Importance (a measure of information value) differs from prominence (a measure of public awareness) in that prominence does not vary with our question. These two concepts obviously entwine. Many prominent sites are important. Many important sites earn a justified prominence. Nevertheless, differences between importance and prominence lie at the centre of our frustration with the internet and define the most significant division in search technique.

As I explained earlier, I believe we should select a global search engine based on size, good field searching and familiarity. With these criteria, I choose Google and AlltheWeb, two important and significant global search engines I am very familiar with. If I judge search engines by different criteria, like the value of the first ten responses to basic questions, then perhaps some of the younger search engines with novel approaches in database mining would be more important. Tools with smaller databases or fewer fields like Ask.com may lead such a list.

Importance depends on my criteria. Prominence is an independent measure quite unrelated to my needs and criteria. Google and Yahoo are the two most prominent global search engines as I write this line – not because they do what I want but because these two names lead any list of famous global search engines.

As a second example, the Library of Congress (LOC) is a most important and prominent book resource. It is important because their freely search-able catalogue lists over twenty-nine million books and offers a very refined search with over thirty fields to choose from. It is prominent because many people know its name, use the catalogue and mention it online. This prominence is evident in how Yahoo tells us 5.8 million webpages mention **www.loc.gov**. And since prominence is best understood in a relative manner, the British Library has 0.68 million, or an eighth as many references.<sup>11</sup>

As a book resource, my local library is much more important to me than The Library of Congress. My local library is important and significant because it lends books, has friendly staff and can be found just down the road. It probably has no importance to you. It has little prominence too. Few beyond my suburb would know and recommend it.

If I am seeking my local library website, I will not find it by typing **library** into a global search engine simply because I am not searching for a prominent site – not as I just phrased my question. If I insist on searching bluntly, then I must phrase my question so that my local library is the most prominent answer to my question. On this occasion, I need type only **library toorak** since Toorak library is the most prominent library in the suburb of Toorak.

In general, blunt searches succeed not because we add another word – that just reshuffles the deck so to speak. No, they succeed because we add the very word that serves to rephrase our question so that the information we seek becomes the most prominent answer to our question.

## RECOMMENDATION ENGINES

Look closely at the differences between importance and prominence. Some of our searches will benefit best from prominent resources. Indeed, if we can phrase a question as a request for a prominent resource, then blunt use of a global search engine is our strongest ally. Ask search engines and directories first since they will undoubtedly recommend the most prominent resources. Their algorithms judge prominence in such a refined way, with such precision. They know prominence.

However, ask a question that requires the assistance of a page we do not think will be prominent, and search engines cannot so easily help us. Not the blunt use of a search engine. Consumed by the assumption that prominent information is important, a global search engine will recommend prominent resources in the hopes such sites will satisfy us.

In Chapter One we saw how many of the world's most significant international organizations publish country profiles on the internet. Publications like the CIA World Factbook, first published to the web in 1992, have enormous fame. I remember it as one of the very first US government documents to achieve celebrity status. There was always something so satisfying about reading something by the secretive CIA.

However, many important and significant country profiles like those by the Pan American Health Organization (PAHO) and the obscure CIPF project by the Canadian Department of Foreign Affairs did not have sufficient prominence to reach our attention easily. When I first encoun-

tered Country Indicators for Foreign Policy (CIFP) it was barely known beyond those directly involved. The country profiles by the OECD, while famous and well-loved in print, were not widely known to be online despite many of these economic profiles being over 70 pages long and filled with world-class expert commentary. Importance as I judge it – primarily authoritative quality content – simply does not equate to prominence. We simply will not find such documents by searching for **country profiles**. An important but near-anonymous profile could easily rank ten thousandth and never reach our attention.

This situation is not ideal. We would prefer search engines recommend important resources – that search engines would list of resources that match our criteria – whatever criteria we have for today’s question. Indeed, this is one of the aims in generating specific search queries. We try to convey to the search engine just what is important to us.

However, search engines cannot judge websites by criteria we don’t supply! Failing to know we want the library down the road, we type **library** and get links to the Internet Public Library, the Library of Congress and the British Library. These are, after all, the three most prominent libraries in the internet world. Prominence is used to fill in the gaps between what we want and what we tell the search engine we want.

Let me explain this another way. Next time we approach a search engine and undertake a search that is not specific – that leads to a list of ten thousand matches or more – then we essentially precede our search query with the words, “Please suggest some prominent resources on ...”.

A Google search for **Jupiter** is actually asking: **Please suggest some prominent resources on Jupiter**. A search for **internet search skills** is asking: **Please suggest some prominent resources with the words: internet search skills**.

Quietly adding this preamble to our search query makes for a clearer distinction between occasions when we want the most prominent resource and when we don’t. **Please suggest some prominent resources on Jane Austen** is not going to help us search for a doctoral dissertation on Jane Austen’s role in advancing nineteenth-century feminism. The most prominent resource on **feminism “Jane Austen”** is no better since we seek a special, unique resource that will never attract much attention and would never become prominent.

As our searches become more challenging, we will find this bias towards prominence often gets in our way. Any comprehensive, definitive or detailed search is by definition not a search for prominence. Any search for quality is only indirectly tied to prominence, as we will see in Chapter Three.

Here is the essence of this argument. Search engines recommend. They RECOMMEND prominent resources. Yes, the epiphany for some readers is this: SEARCH ENGINES DON'T SEARCH! Not when they return ten thousand matches or more. They merely recommend. Used in a blunt manner, search engines are better called 'recommendation engines'.

Let me justify this label carefully for if misunderstood, it is an insult. Firstly, when we search a global search engine, retrieve a list of ten thousand records, then stay within the first fifty, what have we done? We ignore 99.5% of the answer, right?  $50/10,000 = 0.5\%$ . We never look at answers fifty-one through ten thousand.

Select matches randomly and we could suggest we have a sample but we now know of this bias towards prominence. Best to call them recommendations and avoid the suggestion we search anything.

Say we look at the first fifty matches. How is this different from looking at a list of fifty recommendations? How is it different from looking at fifty recommendations from the Yahoo Directory or the Open Directory Project? The only real difference deals with how specific we ask our question. Indeed, a search for **library** or **motorcycle** on Yahoo's search engine provides much the same answers as the same search on the Yahoo Directory. How could it be otherwise? Both use similar criteria.

We do not search the internet – not when we toss a word or two at a search engine. Instead, we ask for a recommendation. "Search engine," we say. "I am interested in a library. Please recommend a few of the most prominent." In response we get addresses for the Internet Public Library, the Library of Congress and the British Library.

Now that we know the bias of the global search engines, and prominence is a bias common to many search tools not just global search engines, we have defined the circumstances where we want their help and when we don't.

I am doing a background check on the activities of a colleague Dean Gates who just started a conversation with me on serendipity. I like to know

something of the people I communicate with. I will peruse anything he has written.

However, a blunt search for “**Dean Gates**” will not help me. A search for “**Dean Gates**” translates as: **Please suggest some prominent resources with the phrase “Dean Gates”** and this just strikes me as a really bad way to search for past statements by one specific Dean Gates. In this case, I search for his email address as well as his nom de guerre, “T. Dean Gates”. Both searches are specific and lead to fewer than two hundred results.

Seeking information unlikely to be prominent, we must either rephrase our question to ask for something prominent or discard a blunt approach in favour of another approach – perhaps a precise search.

Rephrasing our question is often easiest. For instance, a global search engine will gleefully supply us with the most prominent local directory of meeting rooms but would have difficulty coughing up the addresses to small meeting rooms individually. We just need to ask in a way that positions the answer we seek as the most prominent answer to our question.

If we cannot phrase a question to highlight prominence, then use another technique like feedback or precision or triangulation or the page-next-door as discussed in the next few chapters. Much of the success of these other search techniques rests in how they help us rephrase our question into something anchored to prominence.

Prominence/importance is the most significant division in internet search technique. Where as once we discussed the difference between browsing and searching, between directories and search engines, thanks to prominence ranking both browsing and searching lead to similar information. Today, it is far more significant to distinguish a search as either leading to specific information or prominent information. What kind of information do we seek?

'Tis true. There's magic in the web ...  
A sibyl ... in her prophetic fury  
Sewed the work. – William Shakespeare

Yes, William Shakespeare wrote about the web.<sup>1</sup> To confirm this, just look for a really big database of Shakespearean quotations. Do we want the most prominent database? Of course we do. Don't lead us to someone's list of ten favourite quotes. We do not want an obscure quotation either. Nothing Shakespeare wrote will ever be obscure. We want a really big

searchable database of the complete works of Shakespeare. The most famous one will do very nicely, thank you. A blunt search of a search engine or a quick perusal of a large directory will surely assist us in this quest.

However, if we search instead for a quote by some famous historical figure about the web – not thinking of Shakespeare in particular – don't toss a word or two at a search engine. Don't approach a large directory. It won't help. Our question is not phrased in such a way as to benefit from prominence ranking. What would we search for? **“Historical figure” quotations internet OR web**? We don't want the most prominent historical figure. We don't want the most prominent quotation on the web. Yes, in a sense, we have a bad question. That aside, when we want something obscure, specific, comprehensive or quietly unique, we will probably not find our answer in a list of prominent resources.

Just on this example, consider the subtle difference between searching a global search engine for **“William Shakespeare”**, **“William Shakespeare” quotations** and **Shakespeare quotations database**. All three searches are blunt searches. All three return more than ten thousand matches. All three include prominent databases of Shakespearean quotations. Only the third query positions our hoped for database as the most prominent answer to our question.

In summary, we want to use our search tools in a way that brings out their best qualities and acknowledges their worst. Prominence is the specialty of the global search engines. So is precision but never at the same time. Which applies depends on the number of matches found. If we have ten thousand matches or more, we use the search engine to point out prominent resources. If we have two hundred matches or fewer, we have precision. We search. A specific and precise search leads to very different information than a blunt request for prominent recommendations.

## IMPROVEMENTS ON PROMINENCE

Prominence is not the only influence on search engine ranking. Search engines rank more subtly and demonstrate more finesse. On occasions, search engines assume we prefer recent resources or national resources. Pages rapidly gaining prominence probably rank higher than pages with falling prominence. If we type two words, like **jupiter pictures**, search

engines will presume we prefer these words appear together, appear in the title, appear in the linking text, the subheadings and sometimes the meta-tags. If one of several words is relatively rare, search engines will place extra weight on the position and frequency of that word. Furthermore, search engines continually improve and refine their ranking algorithms. The bias towards prominence is not as severe as it was a couple years ago.

Notice I used the words ‘bias’ and ‘preference’. This is another way of thinking about the effects of prominence. Global search engines prefer prominent resources. Search engine bias drives us towards the prominent sector of the internet where we usually, but not always, wish to be.

Prominence ranking is a vast improvement on earlier ranking systems like a reliance on word frequency. Besides, what would we have a search engine offer us? We ask for **jupiter pictures**. We want and get some of the most popular and respected of the 6.5 million matches. This is not a fault. It is a problem only if we don’t want such resources. It is a problem only tossed up by our willingness to look at fifty matches out of many million and call it a search.

Global search engines deliver recommendations splendidly. They do not deliver so well on tasks we should not ask of them but ask anyway for want of another search tool. Comprehensive or complete searches require precision and something else we will cover in time. Unique but unpopular or unrecognized resources require luck and time or some kind of advance knowledge of where to look.

When we ask more than search engines are designed to deliver, we may still find they deliver admirably answering perhaps 90% of our questions with ease. However, this only underlines how much we shy away from asking the more challenging questions!

The next significant improvement to search engines seems certain to be Yahoo’s efforts in social searching. Recommendations now tied to prominence can be replaced with peer prominence, perhaps better called peer respect. Some social tools already exist. They help us find music we like (CDNOW, Rate Your Music), people with common interests (LinkedIn Network) and the blogs we should be reading. The same approach works with internet resources. Ask.com invites us to browse a search tool biased by the preferences of acknowledged experts.

In a sense, this is the next step along a path of interpreting more and more from a given link. At first, meta-search engines counted links. Next,

Google measured the popularity of links. Now, Ask.com measures the presumed knowledge behind a link.

I like this idea, not least because it mimics one of the techniques we will delve into in Chapter Four: that of the link companion. However, changing bias does not remove bias. Social searches will bias their results another way – towards peer recognized resources and away from quiet, non-institutional achievers.

My dream tool would allow me to scale the degree of dependence on peer input, prominence and reliance on word frequency according to my needs. I suspect we will gradually see this emerge in the form of a collection of different global search engines, each biased in a slightly different manner.

Our problem remains, of course. Just what do we want to notice, and what are we willing to overlook, when our search returns a million matches?

To conclude this chapter, let me state this simply. The global search engine is a simple tool that works in one of two simple ways. Either it recommends prominent resources or it allows us to search in a precise, specific manner. If we use it in a blunt manner, recognize search engine bias. Use it to our advantage. At least do not use it to our disadvantage.

I have more to say about search engine bias and recommendations. I have more to say about precision. However, we must first learn about the nature of quality since often we do not seek a prominent or specific answer. We seek a quality answer – and this draws us in a very different direction.

----- End Notes -----

<sup>1</sup> William Shakespeare, Othello, Act 3 Scene 4: “’Tis true. There’s magic in the web of it. A sibyl that had numbered in the world the sun to course two hundred compasses in her prophetic fury sewed the work.” Yes, we start a book on internet search skills by twisting the words of Shakespeare.

<sup>1</sup> Subscribe to her newsletter via ResearchBuzz.com. Tara Calishain is prominent as co-author of Google Hacks (O’Reilly Media 2005)

<sup>1</sup> Information Today Inc Periodicals [[www.infotoday.com/periodicals.shtml](http://www.infotoday.com/periodicals.shtml)]

<sup>1</sup> Online Currents [[www.onlinecurrents.com.au](http://www.onlinecurrents.com.au)]

- <sup>1</sup> End Of Size Wars? Google Says Most Comprehensive But Drops Home Page Count (SearchEngineWatch 27 Sept 2005) [[searchenginewatch.com/searchday/article.php/3551586](http://searchenginewatch.com/searchday/article.php/3551586)] Retrieved Sept 2006
- <sup>1</sup> Tara Calishain and Rael Dornfest, *Google Hacks*, 2nd Ed (O'Reilly Media 2005)
- <sup>1</sup> John Battelle, Google Announces New Index Size, Shifts Focus from Counting, *John Battelle's Searchblog* 26 September 2005 [[battellemedia.com/archives/001889.php](http://battellemedia.com/archives/001889.php)] Retrieved April 2007
- <sup>1</sup> David Novak, Plagiarism Therapy [[SpireProject.com/art23.htm](http://SpireProject.com/art23.htm)]
- <sup>1</sup> Country Guardian, The Case Against Wind'farms' May 2000 [[www.countryguardian.net/case.htm](http://www.countryguardian.net/case.htm)] Retrieved June 2005
- <sup>1</sup> British Wind Energy Association, BWEA corrects some misconceptions in The Case against Wind Farms [[www.britishwindenergy.co.uk/you/cgcase.html](http://www.britishwindenergy.co.uk/you/cgcase.html)] Retrieved June 2005
- <sup>1</sup> This number compares mention of [www.loc.gov](http://www.loc.gov) to [www.bl.uk](http://www.bl.uk). 5.8 million matches from a search for “[www.loc.gov](http://www.loc.gov)”. 0.68 million matches for a search for “[www.bl.uk](http://www.bl.uk)”. Retrieved using the Yahoo! search engine May 2006

---

## YOUR NEXT STEP

**T**his book continues for another eight chapters; another 250 pages of cutting edge guidance on internet search skills delivered with similar flair and ease, covering topics with greater significance and originality. Q4 quality assessment, link companions, deep URL interpretation, publishing paradigms, the elevated vista and more lies ahead.

Purchase a print copy for AUD\$59.95 +p/h or ...

Download a PDF version – or Listen to an Mp3 version NOW. Yes, the full book is available to read online, published under a creative commons license.

Spireproject.com has a copy.

David Novak

Author – Internet Informed.

### Chapters:

Precision  
Prominence  
Quality  
Identity  
Haste  
Structure  
Attention  
Utopia  
Pursuit  
Choreography

For permission requests, see **[SpireProject.com/copyright.htm](http://SpireProject.com/copyright.htm)**.

For advance notice of further publications, leave your contact details at the base [SpireProject.com](http://SpireProject.com)

Feedback most welcome. Again, use the email form at the base of [SpireProject.com](http://SpireProject.com).

---

## ABOUT THE AUTHOR

**T**here is always an element of “Ah ha”, and “Oh, I know this” to my work. We ‘know’ these ideas I share because we use them every day when away from the internet, walking the street, reading a magazine, living life.

Yet on the internet, we set such ideas aside. We focus on the computer - on the technology - on the beautiful yet limiting view that we are looking at the internet.

This is our difficulty. We look but do not see how internet information only masquerades as something unique. We see a different realm, with different standards and a very different feel instead of seeing how similar it is to searching in everyday life.

Since publishing the Information Research FAQ in '97 (likely the first internet document on serious internet search skills), I have continuously drawn the field of internet search skills closer to library science and practical information research, always pointing out how familiar searching should be.

Yet don't let this overshadow the very real advances I have brought to light. Tools and techniques like the InUrl field search, the context bookmarklet, Q4 quality assessment, the notion of an internet history built on publishing models – these original thoughts deserves your attention. It makes your life more productive, more Internet Informed.

My work includes some of the most significant resources on this topic including:

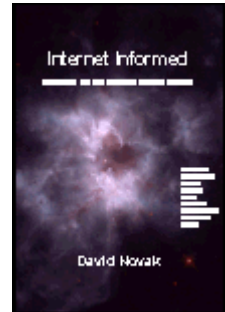




## The Information Research FAQ



And now:



I lived in Melbourne when I created this work. For the next few years I will be in South East Asia.



For a more personal touch, here is my voice from a section off my AudioCD: [SpireProject.com/beyond.mp3](http://SpireProject.com/beyond.mp3)

Internet skill was once about vision and computer experience. The future take us much further afield. I look forward to once again travelling the world speaking and meeting with colleagues so do make contact if this excites you.

Thank you,

David Novak

(Email via a form at the bottom of [SpireProject.com](http://SpireProject.com))

The Spire Project - better ways to find information

[SpireProject.com](http://SpireProject.com)