INTERNET INFORMED

GUIDANCE FOR THE DEDICATED SEARCHER

Precision
Prominence
Quality
Identity
Haste
Structure
Attention
Utopia
Pursuit
Choreography

DAVID NOVAK

Internet Informed

Guidance for the Dedicated Searcher

by David Novak

an element of:
The Spire Project

FOR MY FAMILY

AND FOR THOSE WHO SHARE THE UTOPIAN DREAM.

Microsoft Windows and Microsoft Explorer are registered trademarks of the Microsoft Corporation in the United States and other countries. Google Brand Features are trademarks or distinctive brand features of Google Inc. Many further identifying names used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where these terms appear in this book, they usually start with a capital letter.

Lastly, this book is not professional advice; rather a knowledgeable author writing candidly of his experience. 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.

'Tis true. There's magic in the web ...
A sibyl ... in her prophetic fury
Sewed the work.[1]

– William Shakespeare

La Terre est couverte de gens
ne méritent pas qu'on leur parle.

The earth is covered with people
not worth talking to.

                                    – Voltaire

PROLOGUE

The midnight sky stretches overhead. Thousands ... millions ... billions of stars twinkle and shimmer. Each star shines through the velvet blackness of the night. Red, blue, bright, dim, near, far, new and old, each star reveals so much more than just the twinkle. Each star also testifies to the long legend leading to its existence; to the age-old transit to our eyes; to the vast emptiness between us. A cascade of stories and histories fill our view. So vast. So beautiful. So inspiring.

The internet reaches across a vast mental space stretching beyond all horizons. Thousands ... millions ... billions of bits of information sparkle and chime. Each piece of information shines its light on those seeking guidance. Detailed, superficial, incisive, trite, new and old, biased this way and that, each statement speaks of so much more than the facts contained. Each attests to the insight, perspective and effort bringing it to light; to the challenge of capturing our attention. Once again, a cascade of stories and histories fill our view. So vast, beautiful and inspiring.

Let this be our metaphor. Let this image of an internet galaxy guide us. It will help dispel the trendy but misguided notion of the internet as a web – a web mostly of lost objects from near anonymous sources. Internet information certainly consists of more than facts. Indeed, one of the central themes to using the internet well is to reconnect information with its history; to reconnect with the people, purpose and perspectives that give rise to internet information.

For example, encountering an internet statement of uncertain origin, I click a button on my web browser that reveals context. In essence, I retrieve a list of further publications by the same author and publisher

found in the same directory as the page I am reading. This simple act – this single click – reveals more about the author and publisher than I would ever find within the page itself. Knowing an author and publisher in this way makes their information more vital, more valuable.

Context helps us enormously. Context is also just one of about forty techniques that lead us to a more valuable internet experience.

Let us return to this image of an internet galaxy. At first glance it may seem messy but this is no cloud of information; no ocean of facts. It is more than a web of interlinked pages. Our internet is a galaxy – a vast collection of information where each item of information has a location. These locations and the links between them are adequately described by computer science. Each item of information also has a history partly defining the nature of that information. This history arises from the information's context, format and source; a history closely following the insights of library science. Furthermore, all this information, far from being objects, are indeed messages; presented with purpose; competing for attention. This offers us a third approach to understanding the internet embedded in the social act of publishing as seen through the eyes of sociology.

Location, history and message – concepts like these lend the internet structure, order and organization. And just like our galaxy spread across the night sky – this structure, order and organization is not immediately obvious.

Step back. We must view this creation from a distance to see clearly. Stars and the internet both display a holistic complexity ... and beauty; a beauty simply lost when we focus solely on the objects themselves.

See this beauty and so much becomes clear. Sweep away the appearance of chaos for just beneath our feet rests a solid platform to support us. We may see only chaos at first. We may only suspect a degree of order. But there it rests – a firm foundation that bears our weight and more.

In this book, we shall initially gather a collection of search techniques that extend our ability to collect and appreciate internet information. This will include the use of field searches, an understanding of prominence and endorsements as well the influences of context, format and source. I will also show you a very effective way to reveal quality on the internet.

We will then adopt several ways to move more swiftly. We want to liberate ourselves from some of the drudgery as well as notice the many clues that already flash before our eyes.

Following this, we will address the development of internet information from a sociological and historical perspective. The more we understand the internet, the less confusion will bar our way and the more we will feel at home. This also reveals one further elusive structure to the internet.

Finally, we will discuss choreography, as we decide what to search for and how to frame our questions. A comprehensive or controversial search can be particularly tricky.

Finding answers on the internet is not simple. If it were simple, we would just throw a few words at a search engine. A search would be straightforward; a three point plan; a task we undertake in our most sleepy hour.

Such searching works only for the dullest of questions and search engines used in this manner only point to the most prominent resources; the most brilliant stars. While such searching offers occasional hints of brilliance and generally satisfies us, the best the internet has to offer so often hides from view. The comprehensive, illusive and challenging requires more from us than the tossing of a few keywords at a search engine. As we learn to demand more of the internet, what we find in this way will not impress us.

There is talent to using internet information. This talent saves time; saves frustration. It leads us to better information. It also changes us. We become connoisseurs of information instead of consumers. We relate differently to information. We hold, hoard and value information differently. This talent empowers us - empowers us to frame troubling challenges in terms of questions we can answer. As we answer these questions, we overcome our challenges.

All journeys start somewhere. I shall assume you know how to toss words at a search engine so we will start our journey with making our searches precise – a step that requires a technical understanding of internet fields and field searching.

There is much worth achieving. Much more than you will at first suspect.

Let us begin.

CONTENTS

_____ Part Two: Intimacy _____

_____ Part Three: Finesse _____

_____ Support _____

Part One:

FOUNDATIONS

———————————————

PRECISION

I f a search is war, then the global search engine is our sword. Grab this favourite weapon of ours, 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 already. 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 in Chapters Two and Three, 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 believes 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 AND with the plus symbol (+) and 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 does 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 words we seek. 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, should 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, set it aside.

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 [by which I mean Google™ Search but henceforth will just refer to as 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 Microsoft's Live 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/or-both. Just one word is required. If both appear, that's 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 litera-ture. 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) created the International Patent Classification (IPC), a complex system of patent classification that it updates and improves every three years. The IPC, now on version eight, 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 start 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.
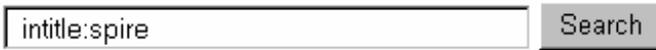
```
intitle:spire                          Search
```

If you prefer Yahoo (by which I mean the Yahoo! Search), 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:**. Microsoft's Live 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-other. 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. Microsoft's Live 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 the similar search of **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 ** ... **.

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 address.
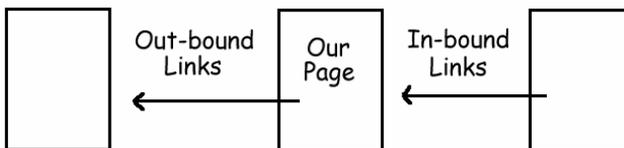
In a superficial way, the more in-bound links a webpage has, the more popular and more recognized the webpage. This is why search engines use the number and presumed significance of in-bound links in their ranking technologies. References that appear at the top of a search engine results page usually 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 using 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 search has 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 reward-ing to search for .ppt files (for Microsoft's PowerPoint) 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 accomplish some unique and rare task but we will not need them often.

The real complexity comes when we step beyond our favourite global search engine and closely follow 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. Microsoft's Live 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, 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,[2] 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[3] as well as Online Currents[4] here in Australia. We can scan **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 searching. 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
Show us webpages with 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 has a different cause.

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, 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.

Stop! Such an attitude is the complete opposite to that of a talented internet searcher. 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 an 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
- and 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 such an approach works. Sometimes such an 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, with 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 an economic synopsis of a country published in 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 fairly 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 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 six 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! Its 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

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 when the bullets fly. 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 substantially larger, search engine indexes have probably grown less 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 feel 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.

<p style="text-align:center">*   *   *</p>

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 to grasp 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 the 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 the inner workings of 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 often 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. 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. We will get to know her much more intimately. She is the elevated vista.

<p align="center">*　　　*　　　*</p>

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 defence 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.

<center>*   *   *</center>

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 punctuated our 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. Is something perhaps 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 in much 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. I usually search several times before I stop and read a 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. That 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, this 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 tutor 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 strived to move more quickly; to improve his balance. It was frustrating though, for try as he would, his skill with a sword 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. In this way, swordplay is a deadly dance for two. Footwork establishes balance. Footwork creates opportunities to attack.

<center>*   *   *</center>

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 and 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 Sulivan of SearchEngineWatch persuasive when he describes how we cannot easily compare size across search engines anymore and how counts do not measure comprehensiveness.[5] 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 4th 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 July 5th 2007 sees these numbers fall some but still shows a similar gap between mentioned links and those available for display.

Google:	12,700 records mentioned, 1000 displayed

Yahoo	3,470 records mentioned, 1000 displayed
Live Search	2,717 records mentioned, 1000 displayed

Size accounts for only 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 that. 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 15th 2007, retrieves:

Google:	64 links recorded, 76 displayed
AlltheWeb:	655 & 180 recorded, 304 displayed
Live 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 1st, 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.[6]

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; if it is 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.

1_ We need a large search engine,
2_ we need a decent URL field search
3_ and we should move freely from our favourite search engine to other search tools for 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 search interfaces are not so significant an obstacle in 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. I think it will be less significant 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 billion 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 unvisited, merely referenced material. Wild claims like Google's statement in November 2005 stating it was three times the size of any competitor seems implausible.[7] Quoted index sizes are not what I would consider good information. Regrettably, 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 could incorporate much of the internet, and they do not, 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.

_____

PROMINENCE

D*isplaying 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 he felt a little less frail.*

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 both a drink. The ever-watchful Captain sat once more in peace, impressed.

<center>*   *   *</center>

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. This number from 0 to 10 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 matured, however, the need for awareness grew. 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; that this page is by Bernadette Robinson. 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 speakers on 'wealth creation' have websites with a good number of links suggesting respect. Look closer, however, and I see few links come from appropriate sources. Too many of these 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** 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, it 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, a prominent website indeed, 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.[8] 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 webpages. 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 (BWEA) in their response to what I considered a rather biased publication by the Country Guardian titled: The Case Against Wind 'farms'.[9] The very similarly titled: BWEA Corrects Some Misconceptions In The Case Against Wind Farms,[10] further mirrored much of the text of the Country Guardian article. Gifted with prominence, their rebuttal is referenced close to the Country Guardian article in many searches. Of course, this avenue is unavailable to those without prominence to spare. Those without a loud voice are more defenceless 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 but you and I set the criteria by which information is judged as 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 searchable 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.[11]

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 sending **library** to 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 CIFP project by the Canadian Department of Foreign Affairs did not have sufficient prominence to reach our attention easily. When I first encountered 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 find a doctoral dissertation on Jane Austen's role in advancing nineteenth-century feminism. The most promi-

nent 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 either rephrase our question, to ask in a way the information is prominent, or we discard our 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 leading to either specific information or prominent information. What kind of information do we seek today?

> 'Tis true. There's magic in the web ...
> A sibyl ... in her prophetic fury
> Sewed the work.[1]
> $\qquad$ – William Shakespeare

Yes, William Shakespeare wrote about the web. 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 far more than ten thousand matches. All three include prominent databases of Shakespearean quotations. Only the third query positions the database we hope to find 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. We use our search engines to show us the brightest stars. 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 these days. 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 similar 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 the 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! Why are we avoiding those questions left unanswered by the loudest among us?

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 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 of a search that 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.

Chapter Three

_____

QUALITY

Albert accompanied Friar Carlo, an aging Catholic missionary, on a journey into the Pyrenees mountains. Church officials wished to dissuade the local pastors from teaching doctrine contrary to the prescribed texts. Certain rural priests had begun preaching concepts of humanism and communalism. To Friar Carlo, such ideas were abhorrent heresy; an ultimate sin imperiling the very souls of these country peasants. Albert believed differently. He knew the locals as peaceful frontier folk. He genuinely liked them and easily tolerated their occasional oddities.

The white-haired missionary from Rome sought to convert these heathens back to a life with Jesus, which was a little amusing since the locals already considered themselves Christians. Only foreigners like Friar Carlo considered them heathens or Cathars. Of course the priest needed no protection. No one would do violence against a travelling missionary. Albert's presence only served to ease political tension back in Toulouse.

On this trip, Albert felt certain the peasants would listen, a few would convert again to Catholic doctrine and local practices would continue as they had for several centuries, building, slowly building, into a distinct southern French culture.

However, this friar's preaching included a serious threat. Vatican leaders were becoming frustrated. A persistent alienation threatened the foundation of Catholic unity. Like a slow poison working its evil magic first in the limbs, a cure was most urgently required. If this trend towards religious diversity was not reversed, it would imperil everything the Catholic church had worked to achieve.

As Albert trudged along the rocky path leading into the Pyrenees, he pondered an unspoken question: How would the church draw the line between Catholicism and heresy?

<p style="text-align:center">*   *   *</p>

The information revolution washes over us. It picks us up and pushes us forward like so much driftwood. From now on, our lives will forever be awash with information. We eat it. Breathe it. Live in it. Drown in it. Some of us will even learn to live for it.

Yet we do not crave the consumption of more information. Overeating just makes us fat. We crave instead answers to questions – even to questions we cannot quite articulate.

As we hunt for that perfect item of information that will satisfy all our concerns and solve our challenges, remember, such information may well exist. It often exists. Yet it stands alongside much comparable but less satisfying information. Perfect information hides in plain sight.

We deserve information we can trust. We deserve quality. We deserve definitive and comprehensive answers too but we are half a book too early to tackle that challenge. Let us first deal with this issue of quality, for astonishingly we can 'deal' with this concern completely. Once we know how, quality is obvious. The internet is the most quality transparent medium we have.

I find an insightful article to accompany my steaming cup of hibiscus flower tea as I sit reading in a teahouse I do not often visit. The article is lively. Well written. It describes with clarity the peculiar nuances to the bloody conflict between India and Pakistan over mountainous Kashmir. It makes several intriguing suggestions. I put down the well-thumbed magazine, sip my hot tea and ponder an unspoken question:

<p style="text-align:center">How much truth resides in this article?</p>

Quality is always an issue. I am not so naïve as to believe everything is true all the time. People lie. People persuade. Even when facts are rendered accurately and faithfully, facts can be selectively presented. Choose a few undisputed facts that favour a chosen position. Marshal them into a logical argument. Now set aside the facts that would refute or weaken this position. Voila! A persuasive argument that leads in a natural progression all the way to our chosen conclusion. Such arguing, such persuasion, will probably lead others to agree to something they would not agree to if they knew all the facts.

Because of this, all of us have a serious difficulty. We simply cannot judge information solely by its reasonableness. Sound arguments are reassuring, most certainly, but a declaration of truth, quality and trust must stand on more than the facts as they are presented. As a topic, quality has depth and spirit.

Who wrote the article on Kashmiri politics? Unfortunately, I do not recognize the author's name. The author is not Dominic Dunn who writes for the magazine Vanity Fair and whose slant and degree of honesty I think I have tied down. The author is not another of perhaps fifty authors I recognize by name. No, this article comes from just another writer I have never encountered before.

The magazine is also new to me. I subscribe to few magazines but know many by reputation and enjoy several with zeal. I trust The Atlantic for a reputable, detailed read. I know Foreign Policy as a respected, incisive policy journal. I consider Time as a superficial, biased yet timely magazine. I do not know the magazine in my hand.

Two strikes. I know neither author nor publisher. The information reaches me from an anonymous source as it were. I am none the wiser as to the quality of this intriguing article.

There is more to consider. For instance, good articles keep good company. I take another sip of my tea, then reach again for the magazine. Glancing first at the Table of Contents, I rapidly page through the magazine looking for evidence of quality in the articles that appear nearby. I see name-brand advertisements. I see an article on the future of NATO. I continue browsing.

If I see one word about reading tea leaves or a fundamentalist slogan, my suspicion of value will shatter!

Good information keeps good company. This is not just a reflection of the work of a publisher. It reflects any selection process. Good magazines keep good company too. I found this magazine on the corner table of a teahouse so most likely it is a good mainstream magazine. Teahouse owners select magazines to interest their patrons. When I find a magazine in a library, it may be less mainstream but will be respectable or at least popular. Librarians purchase the important and prominent magazines for their patrons. When a friend passes me a magazine to read, my friend selects that magazine for my attention.

In a sense the teahouse owner, acquisition librarian and friend all vouch for these magazines. I do not mean they put their reputation on the line – that they believe deeply in the value of these magazines. No, they probably never read the article in question. This influence is more subtle. They vouch for the importance of these magazines, or at least their popularity. They select these magazines instead of others. I in turn trust them not to lead me to the trashiest of magazines.

Such vouching can be vital. In the field of international politics, some reputable sounding policy journals are little more than marketing pieces for special interest groups. They become vehicles to broadcast a perspective – and some perspectives in our world are seriously unbalanced. However, such magazines are unlikely to find their way to the side-table of a teahouse, onto the shelves of a public library or into our hands courtesy of a friend.

Drawing our attention back to our article on Kashmiri politics, we learned very little just now. An interesting article by an unknown author writing in a mainstream publication that appears to publish good articles. That about sums up our assessment. Our avenues explored, I file in my mind the premise of this intriguing article. I attach a questioned uncertainty as to its quality. Perhaps when I read of Kashmir again, I will drag this article from the depths of my mind to compare and decide again if the article's solution has merit.

A half-cup of tea, an intriguing article and a little detective work makes for a delightfully civilized moment. As I consume my morning policy prediction, I weigh, measure and judge information. I undertake what is called a 'quality assessment' that for the most part I leave unfinished.

Can we investigate further? Of course we can. I forgot to look for the author's credentials in the byline. Sometimes the first few pages of a magazine will introduce contributors with a paragraph-long biography. What was the date on the magazine? An old issue would not be so valuable. When was the article first written? Too much may have happened since it was prepared. Can we confirm this magazine is popular? Magazines usually include their circulation numbers in the January issue. If I see this magazine selling in a Borders bookstore, that too would indicate a fair circulation. Lastly, I forgot to nudge the tea drinker beside me and ask, "Do you know and trust this magazine?" Perhaps this stranger is an acclaimed scholar of Kashmiri politics.

Quality assessment saves us from believing in unsupported unworthy conclusions. This can be very important when decisions must be made. Quality assessment also saves us time. We discard doubtful information when better information is thought nearby. We spend more time with the information we trust most. Yet quality assessment works best when quick. This must not be a trial by minutiae or a strenuous test of our patience. We must somehow consider and swiftly decide if information is valuable to us.

The inspiration I want you to see is that each of the steps and concerns we shared with our Kashmiri article apply equally on the internet. Permit me to describe a simple framework I call Q4 Quality Assessment. It is fast, straightforward and deliciously revealing. I have only gradually pieced it together over the last three years but if you let it, it will dramatically change the way you look at internet information from now on.

Q4 Quality assessment has four dimensions:
Q1 • Internal Clues
Q2 • Author and Publisher Identity
Q3 • Context
Q4 • Endorsements

For each dimension, look in a specific direction, ask a specific question, then make a judgement. When assessing internal clues, observe the spelling, grammar, date and the internal logic of the information. Is the information delivered in a sane, professional manner? When assessing the identity of the author and publisher, consider the credentials, occupation and source of funding. Does this author and publisher have the experience to deliver this information? If so, where would their bias lie? Context and endorsements direct our attention elsewhere. When we finish, much of the purpose, bias and value to an item of information will become apparent. The whole process may take as little as two minutes.

Let us start with internal clues, the most understood aspect of quality.

Q1: INTERNAL CLUES TO QUALITY

Good information shouts clarity, sanity and professionalism. We see evidence of this within the spelling, grammar and writing style. We see further evidence in the presence of a copyright notice, well-crafted

images, a link to biographical information, a date and more. These are all internal clues, meaning simply they are found within the information under consideration. Just read and judge how sane and professional the information seems to be.

Evidence of professionalism suggests the manner the author will deal with information when it counts. Make correct use of spelling and grammar and we presume the author makes correct use of facts as well. Marshal and express an argument succinctly and we presume the author considers and researches their arguments.

Unfortunately, this link between professionalism and value is fairly tenuous. Essentially we trust that when an author or publisher remembers to check their spelling, then they probably won't misquote Shakespeare. When the author protects a document with a copyright notice, perhaps the document is worth protecting. The internal logic to an author's argument seems to make sense when we understand it, so perhaps the internal logic holds true when we don't.

There is a professional approach to working with information. The hallmark is an attention to detail and placing the reader's interests above those of the author. Address our reading concerns about bias, currency and supportive detail. Take steps to help readers appreciate factual nuances and confirm details. Include references. Provide sufficient information to verify pivotal facts. Link to biographical details. These steps and steps like these help us as readers to move beyond persuasion. These steps help us believe and trust a conclusion.

This professional approach is very distinct from the inexperienced work of high school students and the empty writing of marketing staff. It is a style built on revealing relevant facts instead of authors pushing their conclusions. When the author – and publisher – work on our behalf, we more easily trust their content and conclusions. We trust, for instance, that we can verify important details with external sources. We probably will not take the time to verify anything but the professional approach suggests this trust is not misplaced. In general, we stake an author's reputation on the truthfulness of their arguments.

Except that we should verify details. As this picture unfolds, we are implicitly told that should an author try to deceive us, we are helpless to see though this deceit unless committed by a preschooler. We can only hope the culprit slips up, as it were, and some culturally inexperienced

political pundit spells a word incorrectly. Perhaps a publisher, intent on convincing us to agree with some key point they passionately believe, absent-mindedly declares their affiliation with the Ku Klux Klan. For our part, we stay alert for such mistakes and wonder if the author presents facts or opinions.

This is not a very satisfying state of affairs. The eminent Japanese scientist may have excellent reasons for his atrocious spelling. Similarly, an author forgets to include a copyright notice or considers a link to a useful book as clutter instead of scholarship. Perhaps we should consider nothing more of such unforgivable gaffes. Yet this is what we do. Tear a manuscript apart as we look for blunders. Tally them up, then make a judgement.

Professional handling of information includes supporting conclusions. Shortly I will introduce you to an article that tells of international aid agencies kidnapping Muslim children for sale in western countries – an accusation courtesy of an Islamic fundamentalist much in the news this last decade. Does the author support this accusation effectively? Such a loaded statement surely requires supporting evidence. Not supporting such a loaded statement allows us to label the work as unprofessional and less than sane.

Quality information shouts sanity and professionalism. Part of this equation is that serious authors understand that persuasion is delightful but not in the interest of the reader. To work to the interest of the reader, authors must disclose the uncomfortable facts, the alternative interpretations, the opposing views. They must share something of themselves – especially their credentials. Readers have the right to know.

Providing biographical information is simplicity itself on the internet. Just link each article to a separate biography. Away from the internet authors still must mention credentials and experience, though usually all too briefly.

The sane and professional author also reveals any serious bias they may have. An author holds a political office? Let the reader know. An author has a relationship with an organization significant to the issue? Let the reader know. Having bias should certainly not stop anyone from publishing but bias should be admitted, early. A wise reader can easily counter bias by reading information biased an alternative way.

This is not too much to ask. When authors do not share their bias and biography, we can level one of two accusations. Firstly, they fail to consider it important. I call this neglect. Perhaps it truly is unimportant. Perhaps including such information is unimportant to the author – the intended audience of peers already knows the author personally or knows the supporting evidence to the conclusion. The author does not care if their message is conveyed poorly to a distant, unfamiliar audience. We simply are not their audience. Newsgroup and specialist discussion often falls into this category. It was created for a specific group in mind, a group informed by a continuing discussion extending over many messages. Each post is not structured to stand-alone or present a conclusion to outsiders effectively.

The alternative to neglect is deceit (or more softly, untrustworthiness). The author intentionally decides not to include biographical details and not to mention their bias. Disclosing such details would probably hurt the persuasive power of their information in some way.

Many a business or industry will spin off an association or institute to support its primary aims. While these associations and institutes are nominally independent, their perspectives and bias remain firmly with those of their founding and supporting organizations. Ah, but do authors and publishers reveal these affiliations? When they do not, or when they disguise a relationship, from our perspective as readers, the author and publisher place their interests at persuading us above our interest in gathering clear information. However common and normal this behavior may seem, unrevealed relationships damage trust.

Yes, trust. As authors we can lose our readers trust by being clumsy or by unsuccessfully distancing ourselves from our bias. How can we earn trust?

Firstly, mention or reference the work of others who support our conclusions. Quoting the New York Times or the Louvre helps. References show that other credible authors and publishers support some fact or position too. The NASA Technology label on face cream (La Mer), an energy drink (Tang!) and many a water purifier tells us the technology conforms to NASA quality standards.[12] Oh, such a relief. If it works on the moon, it will probably work in my humble kitchen.

Secondly, reference the work of people who disagree with our position. In scholarly papers, referring to alternative arguments assures us the

author is aware of such arguments even if considered, then rejected. Such references also show the author tries to present the whole picture; tries for a balanced perspective.

Thirdly, provide the information necessary to confirm stated facts. Instead of indicating a survey says 9 out of 10 dentists like toothpaste, give us the name of the organization that undertook the survey. Instead of saying scientists discovered a new planet circling a distant star, tell us who discovered it and who verified its existence. Go one step further and link to the confirming statement by a space agency. With internet material, single-click verification to pivotal facts is a great advantage to the reader and so very simple for an author to include.

Fourthly, provide details of the research behind a conclusion. Provide as much as possible so interested readers can confirm for themselves how we reached a conclusion. Consider the technical research report: it includes specific research methodologies. Any reader unsatisfied with a conclusion can read over the method and consider the evidence directly. An author summarizes the results of a survey. Where are the survey questions? Perhaps question four was phrased in a misleading manner.

Fifthly, consider the concept of margin of error as used in statistical analysis. When the Australian Bureau of Statistics compiles and publishes census and survey results, they also calculate how mathematically certain their results are. With small sample sizes, the margin of error grows. Sharing this uncertainty with the reader shows professionalism. Away from statistics, a careful author may simply draw attention to a drought of relevant studies on a topic. The author may describe a discovery as provisional, awaiting confirmation.

Trust is more complicated than this. Indeed trust, like quality, ultimately depends on other factors like context and what others say of a document. We are not done discussing trust.

Surveys deserve our most careful attention. Abbreviated statistics are particularly prone to abuse. John Mueller, in a 2005 article in Foreign Affairs, demonstrates this beautifully with the following:

> "It is close to impossible to judge how many people want to get out or stay the course [in Iraq] at any given time because so much depends on how the question is worded. For example, there is far more support for 'gradual withdrawal'

or 'beginning to withdraw' than for 'withdrawing' or 'immediate withdrawal'."[13]

John goes on to recount how in August 2005, the Washington Post found a 54% - 44% split when questioned: "Do you think the United States should keep its military forces in Iraq until civil order is restored there, even if that means continued US military casualties or do you think the United States should withdraw its military forces from Iraq in order to avoid further US military casualties, even if that means civil order is not restored there?"

In the same month, a Harris poll reported a 36% - 61% split when questioned: "Do you favor keeping a large number of US troops in Iraq until there is a stable government there or bringing most of our troops home in the next year?"

A 54% - 44% split or a 36% - 61% split based on little more than how we phrase our question? How important then to hand over the question and not just summarize a survey as "a majority supports ..." or "barely a third favour ..."?

John Mueller's example reminds us that reasonableness does not always equate well with quality. We could easily make a very reasonable argument with either of those two surveys just mentioned. Our very reasonable and persuasive argument, delivered in isolation, seems honest and true – backed by survey integrity even. That each seemingly honest and reasonable argument concludes differently merely proves we must not judge quality by reasonableness alone.

SUPPORT

Presenting unsupported information is a serious misdemeanor. Without a source, date, margin of error and a way to confirm for ourselves the truthfulness of information, an author essentially says, "Trust me". And however much we may wish to trust the author, however much we may be led to expect this kind of behaviour, an author displays bad manners in asking for this trust. Here was a perfect opportunity to earn the reader's trust and the author messed up. This is like a journalist quoting unnamed sources or a news personality saying, "rumour has it ...". Why state something without mentioning where it comes from? As reader we must either

trust the author's word or toss our hands in the air and shout, "What?" to an author who cannot reply.

It reminds me of a criticism I long held with a certain set of business benchmarks. These benchmarks provided remarkable detail on how comparable businesses budget their resources. Unfortunately, for all their uniqueness – they even provided a quintile breakdown of the more important business ratios – they did not reveal the sample size. They went to so much effort to convey such excellent statistics but left off the one piece of information I sought most. Without a sample size, how could I know if the statistics reflected the experience of five or five hundred comparable businesses? We can, of course, trust the publisher knows what they are doing; that they would not sell business benchmarks compiled from just five businesses. I suppose we have to trust them, right? Should we?

I trust the World Health Organization. I trust the British Museum. I do not naturally trust a commercial business. I have watched too many television shows where the evil bad guys are business managers trying to keep a secret. That said, to ask for such trust is just inconsiderate and irresponsible.

Noam Chomsky touched on this idea in reference to documentaries:

> "Personally, I don't even like to watch documentaries where I agree completely with the thrust of it because there is something about it that strikes me as sort of false. Namely, it's presenting the facts in a way in which you can only evaluate them from the point of view of the person who put it together. Now that's also true of the printed page but less so."[14]

Yes, marshal an argument from a few undisputed facts favouring a chosen conclusion. Set aside those facts that weaken our conclusion. Voila! We have a persuasive argument not easily refuted without these absent facts and unmentioned interpretations.

The personal and visual nature of documentaries, and I would add especially network news, makes it difficult for the incompletely informed to disbelieve a persuasive argument. We too easily trust distilled images. We have the images, after all. Pictures don't lie unless they are fabricated, right? We seem to overlook video clips can be real but sliced, diced and biased. Seeing the excited Iraqi crowd toppling a statue of Saddam Hussein

speaks to us loudly in a way we cannot easily disbelieve. A year later I see the documentary, WMD: Weapons of Mass Deception[15] and see how the event was manufactured for the media with a crowd intentionally brought in for the purpose of filming. I am again persuaded, this time that the event was manufactured. It seems I cannot hold a degree of disbelief in front of visual material. My eyes do not let me evaluate images as anything other than truth or fabrication.

Remember the Rodney King episode in 1991 that led to riots in Los Angeles, California? Edited footage gave a very vivid appearance of police brutality – an obvious bashing of Rodney King by no fewer than four police officers. We quickly offered our trust and outrage at a time when perhaps we should have remained more uncertain. Yes, real pictures. No, biased and isolated. Perhaps unworthy of our trust. Instead, we accuse and judge the Los Angeles Police Department solely on the basis of a sliced, diced video segment.

That may well have been brutality. The Christopher Commission found the Los Angeles Police Department did have trouble with brutality. The law courts concluded that instance was not brutality, though a later civil case found two of the police officers guilty. Then Los Angeles Police Chief Daryl Gates recounts:

> "No one knew what Rodney King had done beforehand to be stopped. No one realized that he was a parolee and that he was violating his parole. No one knew any of those things. All they saw was this grainy film and police officers hitting him over the head …"[16]

A variety of opinions surfaced but we saw a video and attained clarity in ten seconds. Can we not keep a modicum of uncertainty? No, as the police officers were acquitted of all charges but one, citizens rampaged.

I believe it probably was police brutality but in the absence of something stronger than a slice of confronting video footage, perhaps the most sensible response was qualified outrage: "I am appalled at the brutality if what I see bears true under closer investigation."

The phrase "If what I see bears true" and perhaps also "We are simply too far from an event to see clearly" may help us frame our difficulties. We are not without information but we are starved of the clues we need to believe a conclusion.

We are often starved of information. Just who is right in the Kashmiri conflict? The answer should not depend on which side set off the latest bomb. Do we really want to make a five second judgement and blame the suggested perpetrator at a time when we have little credible, unbiased information?

Have you noticed how some scientific research papers try too hard to present discoveries and conclusions in all possible lights, interpreted according to all possible perspectives? The author/researcher sometimes pretends not even to have an opinion of what the results mean except in a concluding line or two. In good science, there is no persuasion. The scientific paper focuses solely on communicating facts.

Academic circles expect such behaviour. In political discourse, such raw unspun facts would be so unusual as to border on unprofessional. How could a political pundit fail to take an opportunity to spin and bias and persuade us of something? The same can be said of marketing. If nine out of ten dentists agree, just what was the question? We shall never know.

In some arenas, we expect subterfuge; we expect statements not to have supporting evidence. Spin-doctors spin. Marketers market. Yet, such expectations should not blind us to the fact that this is less than sane, professional, informative writing. In terms of our quality assessment, the author loudly proclaims they are not working with our interests in mind. The author prefers to persuade. While we may not declare this as less than sane, certainly the author and publisher does not deserve our complete trust. They have decided not to earn it.

Yes, to a degree, all discussion tries to persuade. What kind of political documentary would engage our minds yet not express a perspective? We may believe in ever-present bias. We may believe in something akin to nepotism and kleptocracy. Perhaps everyone with a public voice has the right, even privilege, to try to bend our views ruthlessly to match theirs. Realistically, so very many authors will try hard to persuade us – will reach beyond facts and objective conclusions with persuasion. So many authors we encounter will select facts to suit an agenda, hide uncomfortable details and generally deliver information in an unbalanced, untrustworthy manner. I am sure I have in this book, though I am not certain where. Perhaps it is the author's right to persuade us with one-sided arguments, questionable statistics and undisclosed alternatives.

If we believe this, then we have merely shifted the burden of trust to ourselves. Waive the author's responsibility to deliver trustworthy and balanced information and we must assume the role of ensuring truth and balance ourselves. We absolutely must gather alternative information and synthesize a decent conclusion to mitigate the effects of persuasion. If we do not assume the mantle and neither does the author, then perhaps we deserve to be misguided and wrong.

CURRENCY

One final issue to cover: date. The internet loves to confound us here for any date mentioned may mean when the information was prepared, published, compiled or assembled. As I write this book I am mindful that some of my guidance dates back to the mid 1990s. The foundation of this information comes from my experience creating the Spire Project between 1997 and 2000. Some elements, like context-based quality assessment are but a year old and some ideas come together as I write. The book you hold prints "Copyright 2007" at the start of the book but since the book was quickly sent to print, most of this book was typed in mid 2006/mid 2007. Somehow "Copyright 2007" fails to express the range of dates involved.

I shall release two chapters of this book on the internet. As a webpage, it may well have a date reflecting when the chapter was published to the internet or perhaps even the day you download the chapters depending on how the webpage is structured. How very confusing!

Statistics suffer this same confusion. Statistics in a 2004 publication by the Australian Bureau of Statistics may refer only to information surveyed a year earlier. Industry statistics are often severely backdated. Perhaps trade to the financial year July 2005 to June 2006 is compiled and published in February 2007. Compiling census material in particular takes time – often a year or more.

Articles can also have confusing dates. The publishing process may take days or months. Some magazines reprint articles published else-where, earlier. An article may be a book excerpt, the material prepared years before. Date can easily confuse us.

Here are some possible solutions:

1_ If the page is prominent, consider consulting the Internet Archive at **archives.org** for a historical look at a webpage. The Internet Archive will often present prominent webpages as they appeared three years ago, two years ago and eighteen months ago.

2_ Sometimes the URL itself indicates a date. **xyz.com/news/0601/** was probably published in June 2001 (the sixth month of '01). Messages sent to newsgroups or mailing lists often have a specific date etched into its complex web address. When were these documents published?

>
> **recipes.auraskitchen.com/2005/04/Almond-Date-Cookies.html**
>
> **the.honoluluadvertiser.com/ article/2005/May/04/il/il16p.html**
>
> **currents.ucsc.edu/03-04/04-19/news.html**

3_ Your internet browser may be able to detect when a page was last uploaded to the web. With Microsoft's Internet Explorer, under the File pull-down menu, select Properties. Occasionally this will include a created and modified date.

4_ How many links are no longer valid? Links age over time and break. When many links are broken, we should worry about the age of the resource. In my own work, I found about one link in thirty dies in a three month period, though this rate is tied to the topic and degree of deep linking so is a very crude estimate.

5_ For internet and print media alike, look for dated events within the document. Perhaps the author mentions a news event. I just quoted from John Mueller's article in the Nov/Dec 2005 issue of Foreign Affairs. Shortly, I will mention the case of Elian Gonzales washing ashore in Florida. Such statements date the writing.

6_ Statistics also date a document. If a document refers to statistics published in 2006 but not 2008, then the document was probably prepared between these two dates.

7_ Guess. Searching is not the exact science we may prefer it to be. We will guess more as this book proceeds. Perhaps the

document presents the internet a little too enthusiastically, a little too rosy, and this alone brings to mind an earlier era. Perhaps the author likes AltaVista and while still around, AltaVista was more famous a few years ago.

We have now covered the use of use of internal clues to judge quality. In Q1, we simply ask, "Do I find the author sane and professional?" Notice this is a different question to the favourite, "Do I agree with the author?" The advantage in asking both questions is that our conclusion is tempered from an implied, "I believe and trust you" to a more limited, "I find your writing sane and professional." We will not read too much into the appearance and reasonableness of an argument in this way. Is the author clumsy, reserved and suspicious? Is the author helping us to trust and verify their position? In a sense, "Do I agree with the author?" is a completely different issue and quite unrelated to what the author/publishers behaviour suggests. We may easily disagree with an author's perspective and conclusion but we cannot so easily ignore a perspective delivered in a rational, professional manner.

WHAT DO WE MEAN BY 'QUALITY'?

Before we continue on and address source, context and endorsements, let us review some of the traps and subtle nuances to this issue of quality. I have three points I think you will find worthy of a diversion.

1_ Internet quality is a fluid concept.
2_ Quality is plural.
3_ Notions of objective truth may obstruct our search for meaning.

The issue of internet quality has changed remarkably over the last few years. In the early days of the internet, all information was suspect. The graphic image of this was a dog sitting before a computer. The caption reads, "On the internet, you never know who you are talking to." Consider nothing on the internet as credible. Everything has to be confirmed with something in print.

In time, of course, we improved on abject disbelief. With an early flood of good government information, belated agreement surfaced that some

internet information could be trusted. Not everything on the internet held to the standard of dog chatter. Part of this advance rested in identifying government agencies and prominent corporations since they took great pains to usher only reputable information to the internet. The presence of **.gov** in the address was an easy clue, keeping in mind, of course, the often repeated reports of hacked websites.

The next improvement on presumed disbelief was buried within the information itself. If careful, we can spot questionable material. We can sort the work of high school students from professionals. This of course is the use of internal clues to quality.

Time progressed and a more serious approach to internet quality emerged, thanks primarily to members of the librarian community. Library science literature includes numerous checklists and mnemonic devices to help judge quality. Two I often see are:

> RAP: Reliable, Accurate and Plausible
>   and
> CARS: Credible, Accurate, Reasonable and Support

A questioning searcher starts a merry chase to determine the character of an item of information. For each, there are several hints and clues. Reliable means repeatable. Confirmable. Accurate deals with margins of error, survey size and precision. If you detect my minor annoyance, this is true. I am thankful for context and endorsements since they support the reader more firmly as you will shortly see.

Our understanding of quality continues to evolve. It matures with the internet. Perhaps in another two years, new ideas on internet quality will emerge. Quality is a fluid concept. I certainly expect this field will continue to develop.

For the second trap, I ask you to confront how the notion of good and bad quality distorts more than it reveals. To suggest the word 'quality' is singular is just grossly simplifying a complex situation. When we label information quality as good or bad, we make a value judgement based on who wrote it, when and why. We also base this on the reputation of the publisher, the credentials of the author, on what others say of the information, on supporting evidence and presentation and style and so much more. We cobble all this together into a summary: good or bad information.

Too simple. Firstly, we really want specifics. We want to get intimate with information. Like a series of stepping stones, quality assessment often takes us to the next task, that of synthesizing a conclusion from several pieces of information. Creative synthesis requires diversity, agreement and transparency. We will address this shortly but transparency is the big one. Think about what we do with information. We want to appreciate the many strengths and weaknesses of information, not just apply a label.

Secondly, information value depends on our question. Old information may be ideal to identify the history of a person but would be useless in locating their current whereabouts. Change our question and what we value changes too. If we do label information as good or bad, we must re-label with each new question.

The label good or bad tells us very little. Good information could be old or new, come from a respected publisher or an anonymous one. It could be from an experienced scientist, a journalist or a psychotic bystander. Saying 'good information' tells us only that we trust it in isolation. We strip strengths and weaknesses. We reduce bias to a linear measure; to acceptable or not.

The 'good quality' label also hints at a mistaken mathematics of quality. Do two good quality items add to a great conclusion just as two bad items equal one good one? This simply is not so. Two articles from the same old newspaper speak no stronger than one. Two fine articles may reflect the same perspective and refer obliquely to the same supporting evidence. Again they speak no stronger than one. Yet wildly dissimilar information when combined can be much stronger. Suddenly one plus one equals three. We will not notice this if we strip information of its specific strengths and weaknesses.

Time magazine occasionally delivers information in this way. Two articles by different writers describe some newsworthy event from different perspectives. These two apparently separate articles happen to appear in the same magazine edition side by side. Do these two articles assist us to gather a complete picture? It depends. In a sense, we are discussing just one article artificially split in two for stylistic reasons. Both parts to this split article reflect the publisher's bias. As we read and determine one argument as more persuasive, we cannot interpret this as strongly as we

would were the articles truly separate and published in unrelated publications. These two articles share a clear bias.

Here is how quality addition really works. Dixie the call girl reports a murder. Says 'Bruno the Cleaver' did it but Dixie is not a reliable quality witness. She could be lying for many reasons. Frankly, a call girl has little credibility in our society.

As gumshoe detectives we hit the road and find a 12-year-old boy who accurately describes the face of the killer. Golly, it sure looks like Bruno the Cleaver! Yet young boys also lack credibility. We consider children far too uncertain to tell the truth.

Taken together, Dixie's and the young boy's statement offers a strong case for a search warrant. Yet if the young boy were Dixie's son, this is not nearly strong enough. Clearly, details are important. To label both sources as poor quality simply loses the plot.

We need details. We need to know who wrote what. Not just a name but credentials, depth of knowledge, other articles they have prepared and an indication of their professionalism. We need to know the publisher, their reputation and other publications they have overseen. What do others say? We need to know this and many other aspects to the information. We need the many 'qualities' of information.

Remember the old detective shows on TV? The extremely British Inspector Morse listens to classical opera for hours as he ponders the subtle nuances of the information he collects. Inspector Morse does not tally with a scorecard. However, working with the many qualities of information need not be laborious and usually does not require opera music. As we will see with context, many clues positively leap at us in haste. Many clues to quality already flash before our eyes needing only our recognition to be meaningful. Besides, much of the time we only investigate when quality seems doubtful or suspicious. Let us simply see many qualities to information and strive to keep these many qualities distinct in our minds, not lump them together as a simple statement: good or bad.

Now for a third trap involving this topic of quality: notions of objective truth can encumber our quest for a meaningful conclusion.

Is there truth? Yes and no. To simple questions, there are absolutes. We have John's telephone number. It works. Fact. With complex questions, we are often better served avoiding the terms fact and truth. Presume ever-present bias instead. Presume facts are unavailable. In a very postmodern

perspective, perhaps there are alternative truths. Perhaps truth is simply not a practical possibility. Set aside our search for the most perfect information or pristinely honest information. Instead, let us synthesize something far more valuable from pieces of flawed information. This may seem strange so let me marshal this argument carefully.

Hammurabi, the ancient king of Mesopotamia, lived and ruled in the very early days of civilization in what is modern day Iraq. Hammurabi was responsible for the very first recorded legal code including such timeless classics as an eye for an eye. Our question is this: "When did Hammurabi live and rule?"

I reach for the internet and quickly find two conflicting dates. Surely he lived just once? We have made our mistake. The best answer may be a rough date of 1800BC, give or take a hundred years.

This conclusion may not suit us. Perhaps we will look closely at the two dates, learn that one comes from a US university professor discussing legal texts. The second date comes from the staff of the Louvre Museum in Paris. The Louvre displays Hammurabi's famous stone tablet. The Louvre is reputable. I am fine with accepting institutional reputation over a date suggested by a lesser-known university scholar.

Was my choice driven by a conviction that I had to present just one date for consideration? If we are not careful, our convictions in truth and reality can restrict solutions to our detriment. Hammurabi's life is incompletely known. We already have two dates. Continued searching may lead us to conclude the date of Hammurabi's rule can only be guessed. Truth, however enviable a destination, may well elude us.

Another example: doctors struggle to save a young lady's life. She has cancer so we hospitalize her. We pump her with chemotherapy. We strive to banish the evils of cancer. Oops. We forgot to ask the young lady if she agrees. We forgot to ask if she wants to spend perhaps her last remaining days of life vomiting into a hospital toilet. Perhaps she feels there is something more important she should be doing. Perhaps her culture, religion or personal philosophy does not subscribe to our conviction that cancer is evil.

Most likely our patient wants to live and willingly undergoes any treatment for the slightest chance. However, by assuming this, we restrict our conclusions. We close down our options once we stand by certain facts. We distance ourselves from understanding alternative perspectives.

Once we label a position as truth, we cannot so easily doubt it. Truth tends to be unquestionable and undeniable in this way.

In society, we use various proxies for truth. We refer to expert credentials or institutional reputation. The legal profession seeks not truth but sufficient proof to convince a jury of our peers (or an educated judge in countries that do not use the jury system). Politicians decide a course of action not based on truth but on best advice and experience at hand. More cynically, perhaps politicians decide based on what is doable rather than right.

Oh, we need our avatars. Let truth inspire us. But truth is often simply not part of our reality. As with our earlier discussion of good and bad information, truth is too simple a concept.

Let us revisit fundamentalism for a moment and question whether the truths of a bible-touting Christian missionary might not restrict their ability to understand a situation with alternative truths. Convinced of the need to protect souls, our young knight Albert will shortly bear witness to a crusade against the Cathars. History will treat this crusade as genocide.

If truth has limitations, so does disbelief. We can mistakenly enshrine disbelief. We can move beyond a reluctance to believe in certainty to where we hold all events as always uncertain, always indecisive, no matter how clear the evidence. This too is not helpful.

Consider the sociology of Holocaust Denial. There was a Holocaust – the death of a great many Jewish and other minorities during World War II. Let us not discuss whether it occurred. The evidence is supremely strong; the counter-evidence offered by holocaust deniers is frankly pathetic in comparison. Yes, history is constantly rewritten and reinterpreted based on sound argument and evidence. However, we do not rewrite history without evidence. At some point we draw a line and say we believe this, until strong evidence emerges to suggest otherwise.

That is the trap. Start by stating truth is relative. "Let's talk about it." Now use this principle to attack any inconvenient but soundly proven argument. Preach, "Let us discuss whether the holocaust occurred" but do not then engage in a discussion. Draw strength and awareness from, "Let us discuss" but not from any persuasive argument that should follow. Empowered in this way, Holocaust Denial has driven an unargued flight of fancy into the public arena with wide public awareness and wide academic disgust – a most amazing achievement.

We get another taste of this effect in a recent US federal court's decision to chastise a Pennsylvanian school board for making a curriculum change. This change referred to Intelligent Design (ID) as an alternative to Darwin's theory of evolution. Judge Jones declared:

> "To be sure, Darwin's theory of evolution is imperfect. However, the fact that a scientific theory cannot yet render an explanation on every point should not be used as a pretext to thrust an untestable alternative hypothesis grounded in religion into the science classroom or to misrepresent well-established scientific propositions."[17]

In comment on this event, Eugenie Scott, Executive Director of the US National Center for Science Education states:

> "It is already clear that the new slogan for the ID movement [Intelligent Design] is going to be 'Teach the Controversy!' – even though there is no scientific controversy over the validity of evolution in biology."[18]

ID has merit as a religious belief and we judge religious beliefs by very different criteria. However, ID claims to be science – a science belonging in a classroom where religion is not invited. This is just wrong. Unargued and untestable, ID is not science. Arguing Darwinism is incomplete does not make it science. It resembles Holocaust Denial in how both are unargued and unsupported claims standing against well supported, soundly proven positions. 'Let's talk about it' is used to lift both claims well beyond where they belong.

It is not that Darwinism is pure truth, nor that the Holocaust may not one day be reinterpreted in light of new evidence. It is only that you and I should never have heard of Holocaust denial or the 'science' of intelligent design. Why are we entertaining an unargued perspective? When do we discard unargued perspectives? We will not see an end soon to this style of intellectual slight of hand.

Truth and disbelief have their limitations. Sometimes no truth or reputable information can be found. Pretending truth exists in such an ill-informed environment just confuses the situation. Consider popular politics where we often see only strongly biased information. All is spin. Reality is very confusing half a world away as I watch staged events orchestrated by political action groups and reported through the filters of

biased news organizations. Whether we are discussing the politics of Chinese-Taiwanese relations or Iraqi redevelopment, we have little more than rumour and spin to work with. Draw conclusions if you will but do not trust them. Best admit we do not know and work from there.

We should have opinions but recognize when we are poorly informed. I simply cannot know with certainty if Iraq is improving as I write this line in early 2006. Perhaps I should wait a few years for the release of a documentary film since we all know documentaries never display bias.

As you can see, quality is a messy issue. In response, most professional researchers become slightly cynical. We look, decide but sustain a partial disbelief in everything. We recognize how limited our access may be to the truth and we know just how easily a reasonable argument can be neither reasonable nor truthful.

Now is a good time to return to the tenets of Q4 Quality Assessment. Information quality emerges from internal clues, source, context and endorsements. By looking at internal clues, we hope to judge information as sane and professional. Yes, we want information that shouts sanity and professionalism largely because information with these qualities tends to be, usually, by and by, on most occasions, more valuable. Information that appears sane and professional may not actually be sane or professional but most often is.

Better information also tends to emerge from credible authors and exacting publishers. So let us now turn our attention to the degree of trust we bestow on the author and publisher.

Q2: AUTHOR/PUBLISHER IDENTITY

An author and publisher with relevant experience reassures us. As our second task, Q2, we look at the identity of the author and publisher. We ask, "Does this author/publisher combination have the experience to present this kind of information and if so, with what bias?"

The internet is a splendid tool to investigate authors and publishers simply because so very much information gets published about authors and publishers. Anyone who creates a single internet document tends to make three or four. Anyone who writes three or four tends to include a short biography. We may seek information on obscure topics by reclusive writers but usually we just seek information. Since prolific authors say

more, and have more experience saying it, we usually encounter the prolific authors; authors with biographies. Our first task is simple: find this biography.

Q2 overlaps partly with Q3 context since one of the most useful ways to understand experience and bias is to read additional works by the same author and publisher. Additional publications will probably be found in the same directory as our current article. This will be our first working definition of context. However, for Q2 we will restrict ourselves to what the author and publisher tell us about themselves. We want to read their biography as they present it.

To find this biography:

   1_   Look on the page we are on for a link to personal details.

   2_   Seek this information nearby. Look especially for a link on the homepage. This may involved hacking the web address or asking a search engine to show all the pages it has found nearby; two techniques we will explore in Chapter Five.

   3_   Look beyond the website for this biography. Ask a search engine for an author's name and include a concept or email address unique to the author.

   4_   Email the author directly.

The third tactic bears a closer look. Since many people share the same name, when we search for a name, be sure to include an additional concept associated with that person. This makes for a specific search as we discussed in Chapter One. For example, there are many David Novaks on the internet. One prominent David Novak mentioned in the Wikipedia writes about Jewish history. That is not me. I am, however, the only David Novak of significance who writes about internet searching, the Spire Project or quality assessment. A search of one of these concepts (in quotes) like **"David Novak" "internet searching"** would reveal my work.

Also, consider using truncation to accommodate instances where the middle name or initial makes for difficulties. Many search engines do not allow for truncation. Google offers a truncated version of truncation with the wildcard * key so that a search for **"John * Smith"** will match John Maynard Smith the biology professor and John 'Hannibal' Smith the 1980's TV character that led the A-Team.

Hardcover books include biographical details on the cover slip. Past publications may be listed near the title page. Magazines may include a short biography at the start of their publication while journals usually introduce authors at the start of an article. The search for author biographies is not unique to the internet. What is unique is the depth of detail we usually encounter. Do not be surprised to uncover pictures of pets and children.

The other party to publication is of course the publisher. Who are they? Do they have a demonstrated past experience on this topic? Do they have notable commercial affiliations or biased funding sources that may tarnish their objectivity? This is the difference between "John said …" and "John, Senior Prosecutor for the World Court in Geneva, said …".

The more juicy details may come only from outsiders and emerge only when we look for endorsements but many traits will still be revealed by the publisher out of self-interest, in the interest of clarity or because their demonstrated experience forms part of their message. A publisher that is a government agency or non-governmental organization (NGO) or an involved company may want to tell us their background. The publisher of a magazine, a newspaper or a journal wants us to remember their role in bringing us this information. Our task is simply to consider what they want us to know.

Start with the publisher's homepage. If that is difficult to find, hack the web address or ask a search engine for a list of pages within the website. One page or several will describe the publisher's history and affiliations – at least from their perspective of what makes good marketing.

There can be plenty of surprises. In an example I often use, Radio Free Europe/Radio Liberty is a "private, international communications service to Eastern and Southeastern Europe, Russia, the Caucasus, Central Asia, the Middle East, and Southwest Asia."[19] Oh and it is chiefly funded by the United States Congress. Now how would the US government affect the reporting of news in Afghanistan?

A few pages ago I included a comment by Eugenie Scott, Executive Director of the US National Center for Science Education, in relation to Intelligent Design. How thoughtless of me not to add that the US National Center for Science Education is a private association, not government as I at first assumed. It presents itself as a grassroots association "Defending

the Teaching of Evolution in Public Schools." Ah, of course the Executive Director would state there is no scientific controversy to evolution.

Earlier in this chapter I also mentioned a documentary titled, "WMD: Weapons of Mass Deception". Visit the homepage at **wmdthefilm.com** and there on the left, fifth entry in a 'Main Menu' is the 'Filmmaker Bio'. In bold type it tells us filmmaker Danny Schechter authored four books, is the executive editor of an internet project and an award winner for Excellence in Documentary Journalism. A list of documentaries follow. I am greatly reassured this filmmaker has the experience to produce a fine documentary.

As a final example, let me draw your attention to the Astronomy Picture of the Day (APOD) site found at **antwrp.gsfc.nasa.gov/apod/astropix.html**. We will return to this project several times in this book. APOD features beautiful and intellectually inspiring pictures of planets, stars and galaxies. Each picture is accompanied by a detailed description of the science with further links to relevant research papers and official statements of discovery. At the bottom of the page is a link titled: "About APOD" that leads to a list of mirror sites and a brief paragraph that describes how:

> "[APOD] is originated, written, coordinated, and edited
> since 1995 by Robert Nemiroff and Jerry Bonnell."[20]

Robert works for Michigan Technological University while Jerry is a NASA scientist. A search for **Robert Nemiroff michigan technological university** uncovers a short bio that further tells us these two astronomers used to work together at NASA's Goddard Space Flight Centre.

Did you notice APOD is published on the NASA's Goddard Space Flight Center website? The web address starts with **antwrp.gsfc.nasa.gov**. At first I presumed it was the APOD project by NASA – though their brief biography suggests otherwise. Only by exchanging email with Robert Nemiroff did I learn that NASA involvement is limited.[21] "NASA controls no formal content nor exerts any explicit editorial control" explains Robert, though he deeply appreciates the NASA web assistance and a recent grant to continue their efforts. "Some people seem to think there is some sort of APOD team, but it is really only Jerry and me," adds Robert.

I also learned that Robert and Jerry are two professional astronomers with impeccable credentials. They have created perhaps the internet's first picture encyclopedia.

A NASA project implies editorial and management oversight. It implies a standard of quality and excellence and perhaps an avoidance of risk and other institutional habits. It might imply a preference for NASA images. It certainly implies a resemblance between APOD and other NASA picture archives like the smaller National Space Science Data Center (NSSDC) Photo Archive (**nssdc.gsfc.nasa.gov/photo_gallery/**), the Great Images in NASA archive (**grin.hq.nasa.gov**), NASA's Johnson Space Center (JSC) Digital Image Collection (**images.jsc.nasa.gov**) and the Hubble Heritage Image Gallery (**heritage.stsci.edu/gallery/galindex.html**) that includes the cover image for this book.

Does it matter? Perhaps it does. For reasons we will address shortly, the APOD project has many excellent qualities. It is, for instance, very famous. However, to judge APOD's quality based on its association with NASA is misinformed. NASA behaves more like their sponsor than publisher. We discover this only by taking the time to gather biographical information.

The APOD project also behaves more like a self-published project than a project with an institutional publisher. It routinely sources images from all over the world including the best amateur photos. As an astronomical picture archive, APOD is unique in this manner. APOD also focuses not on a tool (like Hubble Heritage and GRIN) or on a photo collection (like the JSC and NSSDC) but instead on science – recent science.

Remember, the question we ask at this stage of the quality assessment is simply, "Does this author/publisher combination have the experience to present this kind of information and if so, where would their bias lie?" We will reveal further detail when we look at context and endorsements so we need not gather a complete answer at this time. Besides, we do not want to rely solely on information provided by the author and publisher. We do, however, expect the author's and publisher's assistance in gathering some background at this stage. Remember, sane and professional writing includes the obligation of telling the reader of any relevant experience and bias.

This issue of the identity of the author and publisher – deciding if they have the relevant experience to produce good information – should not

overly tax our understanding of information. We draw conclusions from author and publisher identity all the time away from the internet. We expect careful analysis from government reports, very careful analysis from peer reviewed journals and oh so sloppy consideration from tabloid newspapers. Elvis may live in Jamaica but we won't know by reading a tabloid. We probably already wonder if reports of some new software could be exaggerated, if a politician is in the pocket of big business and if the latest sports star really likes their breakfast cereal, fruity drink and multi-coloured sports shoe they promote. This should all be very familiar to us.

I will demonstrate this by asking you to imagine you are the Managing Director of a fictional charity called The Smiling Refugee. In 2001, you are given three pieces of information about the starving and neglected refugees of Afghanistan. The first comes from a photojournalist just back from trekking through the war-torn hills of Afghanistan. Her website features a great many forlorn pictures of destitute refugees fleeing civil strife. The second piece of information comes from CARE, an international aid agency with many years experience working in troubled countries. Their report describes what they are doing to assist Afghan Refugees. It estimates refugee numbers and describes the challenges involved in helping them. The third piece of information comes from Bruce Pannier of Radio Free Europe/Radio Liberty, a journalist working for an organization we earlier uncovered as funded by the US Congress.

You have probably already selected the source you value most. You already recognize who has the stronger claim to experience and where the bias for each may rest. I trust you consider the photojournalist perhaps too coloured by the overwhelming experience of being among destitute refugees. I trust you noticed how the photojournalist cannot claim vast experience with refugees so cannot claim to have perspective. Personally, I suspect all war-torn refugees make for heart-rending photographs.

I expect you further recognize that while Bruce Pannier is a talented journalist, he too cannot claim considerable experience with other refugee situations. And I have not yet decided where the bias lies for his publisher, Radio Free Europe/Radio Liberty. As we strive to judge and weigh the risks and rewards of Afghan refugee assistance, Bruce Pannier offers excellent coverage of current events but we really want to listen to CARE and others organization already assisting refugees in Afghanistan.

Connecting author/publisher identity with quality assessment is both simple and obvious. Yet what if we avoid this whole step of identifying the author/publisher? If we anonymize the information we read? The internet already strips the author and publisher from their words. These details are kept separate from the information we are led to by a search engine. We all too easily read internet information and count it as just another author and publisher we've never encountered before.

I am continually surprised how often even experienced internet users fumble like this; how often we overlook source bias. With little interest in who is providing the information, our three documents on Afghan refugees devolve into simply three pieces of information from various sources, suggesting different actions. Do we really want such anonymous information of uncertain bias? Part of being a connoisseur of information is knowing what we consume.

Perhaps we really do not care if a pretty image of a pretty galaxy comes from a project published by NASA or only sponsored by NASA. We just like the picture. Yet author/publisher identity is more than a preference. The author/publisher is part of the message. It is like reading a wise man once wrote:

> "Always think of the universe as one living organism with a single substance and a single soul; and observe how all things ... play their part in the causation of every event that happens."

Nice words. Pretty sentiment. This passage was penned by Roman Emperor Marcus Aurelius in the second century AD.[22] I find knowing its origin adds something to these words. It adds nuances to the information. It makes me think different thoughts. Here is the difference between simply reading and being Internet Informed – a difference that reappears each time we reintroduce a new dimension to information like context, endorsements, prominence, format and publishing model.

We have one other reason to invest time learning more about the information we consume, a reason based on what we do with information after we read it.

CREATIVE SYNTHESIS

Information synthesis is the key to creating the strongest conclusions. We merge several pieces of information into something new. Take two or more bits of information, compare and contrast them, then generate a conclusion stronger than either fact would justify when considered separately. This process is synthesis.

The strongest synthesis emerges when we have three qualities:

- • agreement,
- • diversity
- • and transparency.

We seek agreement from a diverse collection of transparent information.

By agreement, I mean a consistent message coming from the evidence. Agreement need not be absolute but disagreement muddies the water so to speak. Disagreement may simply be within the margin of error or perhaps we can identify a clumsy survey question or too small a sample size. Perhaps Bruno's girlfriend lies when she says Bruno was with her at the time of the murder. If we catch this lie, then we discard her advice and the information is once again in agreement. We hope for agreement. We hope for corroborating evidence.

Diversity also fuels strong synthesis. Marry two items of information with diverse backgrounds and we generate a stronger result. It tells us that two independent assessments draw similar conclusions. In reverse, two statements drawn from the same primary source speak no stronger than one.

Transparency refers to our depth of knowledge about the information. We want to know the information we gather. Sample size. Credentials. Bias. Funding. When gaps are left in our understanding of information, we weaken synthesized conclusions. Only when we know Dixie and the young boy witness are not related can we with confidence ask for the search warrant.

After we gather information, we synthesize a conclusion. Keep this end in mind and perhaps while we gather information, we should look for the best information for this purpose. Seek transparent information. Get intimate with information. Avoid anonymous and unattributed sources.

Gather diverse information. Notice bias, then counter it. We must reach beyond pretty pictures and facts to synthesize a strong conclusion.

As an example, consider the value of wind farms. As a fairly distant observer to this minor conflict, let me recount not the truth of the conflicting opinions but rather my personal impressions as I investigate.

With a quick survey, I decide environmental groups, governments and the wind industry all firmly endorse wind farms. Wind farms are the solution to our desire for a cleaner world. A second, slightly less audible view hates them intensely. They question their value, suggest they kill birds and point out that they occasionally shed giant blocks of ice.

As I browse the internet, all the prominent internet resources on wind farms are very supportive. Look specifically for prominent resources against wind farms, perhaps by approaching a search engine with **against "wind farms" OR windfarms**, and I find what I consider as fairly irrational arguments primarily suggesting that wind farms are ugly, noisy and do little to diminish our reliance on oil and coal. In short, wind farms do not deliver what wind power promises.

Lost between these two positions lies a third perspective. I will describe it as not against wind farms but against how wind farms are being authorized. This position encompasses more than the motto, "Just not in my backyard." It seems to suggest that while wind power is wonderful, the way wind farms are established supremely irritates neighboring land owners in ways that perhaps country residents fear most. Establishing a giant noisy eyesore that hurts property values resembles someone laying train tracks by our front door. Yet governments support almost all applications and pay subsidies to encourage even more wind farms.

I am not well informed about wind farms, which is indeed the purpose of this example since I must synthesize a conclusion from an incomplete grasp of this topic. However, I have a problem. This third perspective, while critical to understanding the value of wind farms, is drowned out by those ranting and raving against the evils of wind farms or those blandly declaring the wondrous value of all wind farms. In short, this is a seriously unbalanced discussion.

For instance, the British Wind Energy Association, supported by prominence, claims wind farms do not hurt neighboring property values – a statement I find hard to believe. Yet with little research yet complete,

those living near proposed and established wind farms have neither the proof nor the prominence to counter such a statement.

I first noticed this third perspective when I found one of the primary websites supporting wind farms had discrete ties to the wind industry. I began to worry about bias so I searched for information against wind farms to counter this bias. Then I began to see errors in logic. Someone quotes a report by the German Electricity Industry but completely misuses the conclusion. A sensible claim of survey bias stands against one of the few surveys suggesting wind farms don't hurt property values. This confusion leads me to hunt for other angles and for the perspectives of those more informed than I.

Greenpeace supports wind power. Government policy seems to authorize all wind farms except under the most serious conditions. Only now do I notice and recognize a few discussion pieces on the sensible issue of property value. A third perspective began to emerge. Only by considering all three perspectives do I reach my final conclusion that I am pro-wind power but troubled by how wind farms are authorized.

To form the best conclusions, to synthesize the best information, somehow we must see through this mist of unbalanced discussion tilted towards the loudest voices. Dip lightly into the internet and we encounter only praise. Search for dissenting voices and we encounter what I describe as the irrational argument against. Only by listening carefully and noticing the identity of the author/publisher do we hear all three perspectives.

Whenever we wish for a comprehensive or definitive answer we must grapple with this possible situation. We must not let our search tools draw us away from valid but quiet positions. We must not synthesize solutions based on resources supporting just two sides of a three-sided argument. We will revisit unbalanced discussions again in Chapter Nine but synthesis remains a reason always to consider author/publisher experience and bias.

There is an abbreviated version to synthesis and it works like this. Grab a high quality reliable format of information – we will introduce format in Chapter Four. Next, simply accept their claim of truthfulness. Stay alert for anything unexpected or suspicious but essentially trust the author. If something unexpected or disturbing appears, we investigate. This shortened approach works well if we do not need great confidence in our

conclusion. If we need such confidence, then better to start with distrust. Interrogate everything to learn of its origins.

Say we read an academic research report from a peer-reviewed journal. We can probably trust the author and publisher. Peer reviewed journals tend to have high quality. The author's peers should notice significant errors before publication. Intent on saving time, we jump directly to the conclusion and read that section carefully. Yet as we read, something strikes us as odd. Perhaps the report mentions doing a survey twice or suggests the need for a larger survey size. We investigate. Not knowing the survey size disturbs me so I will interrogate the information until I am satisfied the sample size was sufficient. Of course in a research report, I have only to read the methodology section. If sample size is not described, I may well choose to disregard the report entirely or contact the author directly for clarification.

In this simple approach we trust information but only when it conforms to our expectations. We stay vigilant for anything unexpected. This stance works only when we do not expect a significant bias. When we work with less trustworthy resources, with more contentious issues and issues with serious repercussions, do not be so trusting.

Q3: CONTEXT–BASED QUALITY ASSESSMENT

We will now address a very simple technique that once understood, will take all of five seconds. This really is simple. There is even a bookmarklet we will discuss later to make this simpler still. It will become a single click and a glance. A mere gesture.

Information keeps good company. We know a publisher by what they publish just as we know a book author by their previous books. In regards to the article on Kashmiri politics mentioned earlier in this chapter, I looked at the Table of Contents, then quickly paged through the magazine. Nearby articles will reflect the quality of the article I am interested in. This is context-based quality assessment. Our question for Q3: What company does this article keep?

A magazine publisher strives to deliver a particular quality of information. All articles will reach that level or remain unpublished. Thus, a list of articles from the magazine, past or present, will reveal something of the publisher's preference and bias.

We know book publishers by the books they publish. The publisher's quality and preferred topics show in their past publications. A newspaper broadcasts its quality and bias through its news articles. London's Times, Guardian, and The Sun all differ significantly and these differences have little to do with geography and distribution. Their differences spring from the quality and selection of their articles, from the sexy girl on page three. Consequently, we know something of the quality of a newspaper by their articles past and present.

In all three of these examples, publishers select information to match their standards of quality and bias. Consequently, nothing would be more natural than for us to open a magazine, a publisher's catalogue or a newspaper and skim for an impression of quality, perspective and bias. After all, their reputation is revealed in what they publish. Yes, we are discussing an aspect of a publisher's reputation seen in the information held nearby. We are discussing context, local context to be precise.

Now let us reach for the internet and watch as this fine structure of context collapses into a pile of rubble. Internet information, we are told, has no context. Every page sits beside every other page in this vast cloud of information we call cyberspace. Every item of information has its own qualities and perspectives that exist unrelated to the next page we happen to visit.

By such a definition, cyberspace has no neighbors. As we search, we constantly meet new publishers of unknown quality and bias. However, the internet is not a cloud. It is not an ocean. The internet shares all the structure, order and organization of a galaxy – a finely structured galaxy. Every webpage has a specific location as defined by the web address. Publishers place their information together in one place. Now notice how our vision of context-based quality assessment snaps back into place. Internet information exists in a specific directory. With very few exceptions, all the information within a single directory comes from a single publisher. In most situations, information within a single directory is written by just one author. Context-based quality assessment merely asks that we glance at the other information found within the same directory. Like articles in a magazine, webpages within a directory share context.

The URL field search is the key to unlocking local context. With a particular directory in mind, ask our favourite global search engine for a list of all the other webpages that appear in the same directory. Google or

Yahoo! will happily provide us with such a list when we type **inurl:web_address**. Click the search button and we retrieve a list of perhaps ten or three hundred articles by the same publisher.

And if I see one word about reading tea leaves or a fundamentalist slogan, I will have my answer.

If a webpage that interests me is published right next to a webpage about royal jelly lipstick, I will assume:

1_ The publisher is the same for both webpages.

2_ Both articles are held to the same standards of quality.

3_ Both articles are stupid.

Yes we can dispute each of these assumptions. In a critical way, we should. Firstly, perhaps a new publisher has taken over and is destroying the credibility of a magazine or web project. Quality standards have begun to slip. Secondly, perhaps a brilliant article has found its way to a low quality location. Sometimes great works of art really are sold at garage sales and sometimes great authors publish great works in mediocre magazines. Thirdly, perhaps I am wrong about the value of royal jelly lipstick. Perhaps royal jelly lipstick really works and deserves its place on the cover of respected magazines and newspapers everywhere. Context is not a wildly accurate measure of quality.

Yet context is central to the way information is sold and consumed on and away from the internet. I would certainly not believe a change in our understanding of gravity when the tabloid newspaper, The National Enquirer, tells me so. Place the same news in Scientific American or The New York Times and I am a believer. Context matters. Contextual clues surround information like smells surround our evening dinner. They are ever-present. Where there is information, there is context. And just as the smell of dinner reveals what is for dinner, context reveals the quality of information.

A reputation is a complex creation arising from past work and the attention and comment it receives. Local context is but one part of reputation. If Q3 instead asked for the reputation, we would have much more difficulty in gathering together the many pieces – the past articles, comments and fame. It is the simplicity of our question that empowers us. We merely wish to consider the information positioned nearby. So framed, context is often so very revealing.

As our example, let us turn our attention to The New Powder Keg in The Middle East, an article by Mujahid Usama Bin Ladin published at **www.islam.org.au/articles/15/LADIN.HTM**[23]

The name Usama Bin Ladin positively drips with meaning as I write this line but let us consider this particular article in context. What else has the same publisher produced? What other articles appear in this same directory?

This article appears at:
> **www.islam.org.au/articles/15/LADIN.HTM**

So we will ask for articles found in:
> **www.islam.org.au/articles/**

And we will ask Google by typing:
> **inurl:www.islam.org.au/articles/**

This last line roughly translates as: "show me everything you have found with **www.islam.org.au/articles/** in its web address." We covered how to do this in Chapter One. We receive a list much like this:

> Interview With Mujahid Usamah Bin Ladin ...
> The Islamic Taliban Movement And The Dangers Of Regional ...
> THE MORO JIHAD: A Continuous Struggle for Islamic Independence.
> The Islamic Legitimacy of The "Martyrdom Operations ...
> Christian Missionaries in the Muslim World - Manufacturing Kufr ...
> The Liberation of Constantinople ...
>   and so on.

If we briefly scan one of these articles, perhaps the one of Christian Missionaries fifth down the list and prepared by a different author, we would read the following paragraph:

> "Perhaps the most insidious method used by [Christian] missionaries is to kidnap Muslim children from war-torn countries and sell them to non-Muslims to raise as disbe-lievers..."[24]

As I read, I begin to make my quality assessment. Oh please don't make an assumption of all Islamic communities from the actions of one apparently fundamentalist publisher. Such injustice. Also, I am not propagating a jihad against Islam by mentioning fundamentalism. That said, we should certainly make assumptions about this particular publisher from this

publisher's past actions. At our fingertips rests a large selection of past articles with which to judge them.

As a second example, say we search for **staff retention** using Google. The first search engine recommendation draws us to a page titled IDS HR Studies found at **www.incomesdata.co.uk/studies/impstaffretention.htm**. This page introduces a study on Staff Retention, a study that requires a three-month trial subscription priced at £50. Is it worth this price? Part of our assessment should include Q3 context – a quick look at information found nearby.

> This article appears at:
> **www.incomesdata.co.uk/studies/impstaffretention.htm**
> So we will ask for articles found in:
> **www.incomesdata.co.uk/studies/**
> And we will ask Google by typing:
> **inurl:www.incomesdata.co.uk/studies/**

Browsing the resulting list of nearby pages, we learn our page about a Staff Retention study sits beside 106 other IDS studies on other topics important to human resources. I interpret this as serious proof the IDS has worked in this field a long time and has the expertise I seek. Context supports their claim to quality.

Context-based quality assessment shines a strong revealing light on internet information. This simple technique exposes much about a publisher in a quick and concise manner. Furthermore, publishers are ill-equipped to disguise their past actions and associations unless they wish to paint their past as a void. This last option is foolish since publishers can all too easily sink into anonymity.

A publisher wishing to disguise a bias may project an image of respectability, accountability and authority. They may project an image of omniscience. They cannot project an alternative perspective without actually convincing readers of this alternative conclusion or appearing undecided. This means that when a website strives to inform us, a publisher cannot step around their own bias. A newspaper intent on presenting itself as the serious international paper better be filled with serious news and not gossip. If we look nearby and see gossip, it is a gossipy tabloid no matter how much the publisher claims otherwise.

Oh, a political party can still pretend to care about an issue, to make policy statements in support of an issue but do little of significance. Context will not reveal this since their purpose is not to inform us but rather to gather votes. However, publishers intent on informing us cannot hide easily from their context. And with so few internet users reaching for context, few publishers try. I have seen some incredible statements online in my time, statements that are so obviously biased when observed in context.

Wait a moment. Good information keeps good company. Can we really judge an argument by its neighbors? Surely arguments stand on their own. Authors only contribute to their message. Besides, we should read important information irrespective of the author, right?

We do not have the time to read everything so we use context to judge importance. The unproven expert does not really help us appreciate their experience or help us recognize the importance of their work. Without publishing a few articles, essays, a book or website, the self-published author essentially says, "I know I have not written much of interest before but this page is different." I find such claims insipid and unconvincing. Almost all great discoveries come gradually from people who work at it and have a track record of working at a topic. If the author has not published anything significant before, do they deserve our attention?

Furthermore, the author could not find a publisher willing to recognize the value of this information. "I found no publisher willing to vouch for its significance. I could only put it here." Of course I will judge such work harshly.

There are two issues to discuss that involve the mechanics of context-based quality assessment. Firstly, we may not want to look only at the immediate directory. In the example just mentioned, involving **www.islam .org.au/articles/15/LADIN.HTM**, we requested a list of all articles in **.../articles/**, not just those in **.../articles/15/**. Fifteen probably refers to the issue number anyway. We may need to step up another directory to learn something useful. Of course, a URL search for **.../articles/** will include anything found in **.../articles/15/** as well.

Secondly, the URL field search is not the only method to gather context. Two further ways will reveal information positioned nearby, though I hesitate to even mention these alternatives since the URL field is so simple. In Chapter Five, we will introduce a bookmarklet to add to our

web browser that makes using the URL field search as simple as a single click with our mouse. I will also introduce a search tool that automatically converts a pasted web address into a URL field search.

The second method involves hacking at the web address and guessing the address to information positioned nearby. Web addresses are chosen to mean something to the publisher so with a little practice and a few nearby examples we can usually guess a page or three. Hacking web addresses can be tricky but most every directory, for instance, has a home or index page that will end in either .htm or .html. We discuss this technique in Chapter Five.

A third method involves simply surfing to nearby information. This can be fairly inefficient but will lead us to recently published information – to information perhaps not yet indexed by a search engine.

Context reaches further than we have described here. In Chapter Four we will investigate the link companion, the electronic footpath and the information venue. However, as we search for quality, with speed in mind, we want to look only at information positioned immediately beside the page we read. In practice, as I reach Q3, I click a button on my web browser, I browse a list of other documents held nearby and I decide what they say of the depth, quality and bias of the information that interests me. At most, I might read a couple of documents for flavour. Am I impressed at the volume of information published nearby? On to Q4.

As an aside, when we find a lovely article on the internet, we may eagerly wish to read any similar information positioned nearby. Where there is one exciting article, perhaps there are two. The URL field search is once again the key. Simply ask a large search engine for pages they have indexed nearby. Indeed, in this way the URL field search introduces a second way to traverse a website. We can surf through a website as the publisher intends or we can pick and choose from a search engine's list of pages instead. For badly constructed websites, this approach works very much to our advantage. Just acknowledge that search engines may not index unpopular pages – not a problem, of course, for websites with prominence. Search engines may also miss the newly published page. I move 'horizontally' through a website in this way perhaps once every hour I search.

I love the flexibility of the URL field search. I also enjoy the role I have had in first bringing Google's inurl search term to public attention and

now in revealing context-based quality assessment. These are two of my grandest discoveries. Knowing my work includes these two significant discoveries should somehow enrich my other words and actions. It becomes part of my reputation – a reputation we can tap into by searching local context.

Q4: ENDORSEMENTS

Here is another simple technique that once understood, takes very little time. This too will become a gesture; a simple and very revealing aside.

Nobel Peace Prize Winner Aung San Suu Kyi remains under house arrest for yet another year in Myanmar (Burma) where hundreds to thousands are routinely detained for wanting freedom from tyranny. In 1991 she was awarded the Nobel Peace Prize and has never been anonymous since. Oh, she is not just one struggling opposition speaker among many. As the daughter of a famous general, she was never unknown but in a sense the Nobel Prize changed her into a symbol of Burmese democracy; her continued detention proof of its denial.

The Nobel Peace Prize is another milestone in her growing popularity with empowerment arising from this popularity. It is a strong statement of the nobility of her struggle. It is also the only reason I know her name.

One of the first pieces of advice I gathered from the National Speakers Association of Australia (NSAA) was that after each speech I deliver, always ask for a testimonial: a statement of support in writing. Post these testimonials to prospective clients. Put them on a website. Create a book of testimonials much like the portfolio of a graphic artist or the published articles folder carried by a journalist. Ask for glowing testimonials. Even suggest words. Guide the process if the client does not object but let it be their words, their sentiment.

Used properly, testimonials allow a speaker to establish credibility without sounding egotistical.

I love Japan. I love the land, the culture, the people. With high school memories still fresh in my mind, I toy with the idea of returning to teach English with one of Japan's big English training firms that always seems to advertise for foreigners in English newspapers. "Come and teach," they declare. "No experience needed. We will arrange everything."

Intent on making an informed choice, I use the internet to investigate the experiences of teachers who have taught in Japan already – and run straight into the most extensive hate mail campaign I have ever witnessed.

Literally hundreds of past teachers join together to share their horror stories. Stories like being followed then reprimanded for befriending locals. Stories like being asked to teach in a way that barely deserves the name. The near universal message: skip the big three and arrange directly with a smaller English school – one that demands teaching experience and a relevant degree. I read in fascination at the depth of outrage and anger.

Awards, testimonials and hate mail; we are discussing various statements made by people and organizations not involved in what we are investigating but commenting on it. In one sense, these statements recommend something to our attention. Statements like a Nobel Peace Prize and testimonials speak of significance and support. Hate mail speaks loudly of significance too. There is anger, certainly, but also significance. This recognition is very valuable on the internet where attention is so fundamentally important to reaching an audience.

In another sense, awards, testimonials and hate mail all convey a worded message. The noble effort of Aung San Suu Kyi. The proven value of a public speaker. The collective anger of a hate mail campaign raging through a discussion list. Each of these statements speak of much more than importance.

I shall use the term 'endorsements' to describe these statements, even though very negative endorsements will occasionally surface. English does not supply that perfect word suggesting appreciation but occasionally meaning slander. Perhaps we can focus on how even negative statements lend a kind of support. Like the Hollywood starlet making scandalous headlines; famous or notorious, just as long as people notice.

We cannot call such links as mere 'references' since this term only acknowledges they refer us somewhere. A reference directs our attention. Labeled endorsements, we stress their supportive role in lifting a page from a mere quiet existence towards public attention and significance. We also stress they describe a person, page or project; they are opinions often by people or organizations much more familiar with our object of interest.

Away from the internet, endorsements may imply advertising. Beyonce Knowles endorses Pepsi. Tiger Woods endorses Nike. This happens on the

internet too. Indeed, some endorsements we encounter will be purchased and we generally will wish to ignore these. We can also use the term endorsement in a political sense. Greenpeace endorses a candidate. After investigating the candidate's background, carefully considering their past voting record and future promises, perhaps sharing their concerns in person, Greenpeace as an organization suggests their supporters vote a certain way.

This second sense, this political sense, best describes the humble link. After considering the alternatives, a web publisher decides to include this link and not others; asks us to visit this page and not another. We happily listen, thinking their decision rests on a superior understanding of the subject.

Drawing once more back to our article on Kashmiri politics, an endorsement is seen when the teahouse owner, the librarian or our friend vouch for a magazine. They purchase this magazine for us to read, not another. They select this article for our attention, not another. They nudge information our direction in a manner that speaks of importance, significance and appreciation. These are all endorsements.

On the internet, this delicate web of vouched for information, this web of endorsements, persists. Some endorsements may speak very loudly, like the Nobel Peace Prize. Most will speak very quietly indeed. Our task is simply to gather some of these endorsements together, peruse the most interesting, then consider what they say about a resource.

Thankfully, gathering endorsements is actually very simple. We already met some of them as links.

Endorsements on the internet come in three varieties:

1_ links,
2_ mention of an address (but not as a link)
3_ and referring to a project by name.

Yes, we are interested in more than just those webpages linking to our page. Indeed, we are interested in several aspects of each endorsement:

1_ who endorses,
2_ what they say
3_ and how many endorse.

Retrieve a list of endorsements. Peruse them. Look at the number, the source and perhaps some of their comments. We now simply add this to our quality assessment.

Our question:

Q4: What do others say about this information?

Such a revealing question. For the first time, we move beyond the direct control of the author and publisher to the words of others. Like testimonials, some endorsements will be carefully cultivated. They may be purchased, seeded or swapped.

At other times, like hate mail from past employees, endorsements may reveal information authors and publishers desperately prefer to keep quiet. We can confirm bias or reveal conflicts of interest thanks to the comments of outsiders. Yes, endorsements speak of more than value and quality. They may also speak of perspective, position and stance in a marketplace.

An absence of endorsements tells us something too. It suggests a lack of fame and prominence that in turn suggests a new website, a site without promotion or a site without value. Do be careful interpreting a quiet and unnoticed site as low quality. It may well be brilliant, just unrecognized. However, sometimes quality information may be insignificant because it does not capture the attention needed to influence us. Sometimes, the unrecognized website is insignificant indeed.

This notion of seeking endorsements is neither new nor foreign to us. We ask these questions already in life. We are, however, uniquely able to retrieve endorsements from the internet so very quickly. With little more than a flick of the wrist, a gesture of interest, we can uncover a list of other people commenting on a website. In comparison, locating dissenting voices to a magazine article takes time and probably money. Sometimes a magazine will publish in their next issue a rebuttal or a counter argument to a confronting perspective. Sometimes letters to the editor respond with corrections but we must wait and hope. If not, we must reach for a commercial-quality database and start searching for an article that carries an argument further.

On the internet, alternative perspectives are built into the system. We merely interrogate the information. We merely ask.

Endorsements come in three varieties:

1_ Webpages that link.
Search for links by using the link field search.
link:web_address

2_ Webpages that mention an address.
Search for pages that mention an address.
"web_address"

3_ Webpages that mention a web project by name.
Search for a web project by name in manner specific enough that other projects are not included. This usually means we add a choice keyword certain to accompany the project name.

To gather endorsements for The Spire Project at SpireProject.com:

- search for **link:spireproject.com**
- search for **"spireproject.com"**
- search for **"spire project" "internet search" OR "internet research" OR "David Novak"**

Naturally, we can combine these searches. A bookmarklet in Chapter Five to assist with this. As a general rule, about as many webpages will include an address as will link to the address. As I write, search engines like Yahoo and AlltheWeb will show more linking sites than Google. Also, different search engines may lead us to different endorsements. Lastly, Yahoo has the linkdomain field that allows us to retrieve pages that link to any webpage within a domain. **linkdomain:spireproject.com** reveals links to **SpireProject.com, SpireProject.com/country.htm** and **SpireProject.com/ cn/past.htm**. We can almost always find a few more endorsements if we wish.

Few endorsements will link to or discuss specific pages buried deep within a website so we may have to chase where the endorsements point. In regards to a contentious report, will someone link to the address of the report, only mention the address or mention only the organization's website? If a meaningless flash page sits on the top of the website, will people point there anyway or to the more meaningful second page? Getting a list of endorsements can be finicky in this way. We may need to search several ways to get what we want.

Where is this discussion taking us? As I mentioned, the internet is uniquely transparent. We can easily put our ear to the ground and listen

to the murmur of conversations discussing a website, business or author. Furthermore, publishers are little able and generally unpracticed in hiding the diverse opinions we can dig out in this manner. Because of this, internet endorsements represent a significant reversal of power. I cannot meaningfully accuse and insult a business in the real world but I can cripple one online.

As a case in point, several newspapers trumpeted the near-inevitable lawsuit against the Wikipedia for a transient statement made by a volunteer editor. The Wikipedia is an encyclopedia where anyone can add entries but where bad entries are gradually corrected in time by peers and editors. The Wikipedia is a collective project on a collective space. It is also in a legally grey area since the law does not yet clearly accept the existence of a media commons – public media held in common. Indeed, public land held in common does not have special legal protection to my knowledge.

The accusation is serious. The threatened lawsuit appeared serious. Then I read an article on the Wikipedia that totally disemboweled the individual threatening the lawsuit. A chronicle of past failed internet projects; revealed relationships with deeply biased publications.[25] All the skeletons in the closet are on parade and I am more inclined to laugh than take the lawsuit seriously.

I mention this not to suggest publishers do not have a good recourse against negative endorsements. Indeed, rebuttal can be very effective. I mention this because publishers cannot simply ignore the words of negative endorsements when they lodge in a place where people see them. At least for a transient time, accusations of misconduct travel well on the internet. Perhaps in time, as more and more surface, they will become as routine and ignorable as spam. However, they do travel now. Furthermore, as internet users learn to reach more frequently for endorsements, the visibility of negative endorsements will grow.

Most endorsements are neither negative nor particularly positive. They reference and suggest significance or represent support, perhaps just financial support. Indeed, this is one task we must set ourselves as we view endorsements. Why were these endorsement made?

We are interested in trends so thankfully, our answer is usually very clear. We retrieve basic information from the web address of each endorsement so reading endorsements will take little time. We will cover

how in Chapter Seven. A link from some professor's university course notes means significance to university students. A link from a newspaper means it was in the news. A link from a personal page may mean many things but a quick look at where the link sits on the page should be enough to decide. Collectively, a link on thirty directories definitely means significance – and we can return to our Chapter Two discussion on prominence to explore this further.

Here is the first of two examples:

Astronomy Pictures of the Day (APOD)
www.antwrp.gsfc.nasa.gov/apod/

Google tells me it has 600,000 references for **APOD astronomy picture of the day. link:antwrp.gsfc.nasa.gov/apod/** on Google reveals 3700 links and 56,000 links on AlltheWeb.[26] These links come from a diverse range of pages, some personal, some university department websites, many from international websites. I also see a collection of mirror sites and a Yahoo listing without looking deeply.

With numbers as large as these, our quality assessment is overwhelmed with presumed significance and longevity. I am not seeing evidence of peer respect but I did not look in a way that would reveal it clearly. Certainly the university links and NASA's involvement suggests peer respect. It appears APOD finds its way to many a page of assorted links and many a page pointing to astronomical pictures. This is partly a result of being on the internet since 1995. However, an AlltheWeb search for **link:antwrp.gsfc.nasa.gov/apod/ url:gov OR url:edu** reveals 2,990 links from government and educational sources. Many of these links would have been hard to earn without peer respect.

As a small aside, I could not get Google's help with this last search so I quickly jumped to AlltheWeb, a search engine that has less difficulty with complex mixed field searches.

The Washington Institute for Far East Studies
www.washingtoninstitute.org

I first encountered the Washington Institute several years ago through the work of Noam Chomsky in his book, "The Prosperous Few and the Restless Many". In this 1993 book he wrote:

"Martin Indyk, [whom President Clinton appointed to the Middle East desk of the US National Security Council,] headed a fraudulent research institute, the Washington Institute for Near East Studies. It's mainly there so that journalists who want to publish Israeli propaganda, but want to do it 'objectively,' can quote somebody who'll express what they want said."[27]

For those readers not familiar with his writing, this is vintage Noam Chomsky. He likes to throw out a confronting statement – the plain truth – then discuss it further. I think of him as the academic forerunner of Michael Moore.

The internet offers us a great opportunity to confirm or dispel his accusation against the Washington Institute. In early 2005, I search Google for:

link:www.washingtoninstitute.org OR "the washington institute for near east" OR "washingtoninstitute.org"

This returns more than 50 thousand matches. All these endorsements link, mention or refer to this organization. Obviously, this is a significant institute. People talk about it. Buried somewhere in these fifty thousand endorsements may be some confirmation that they are, as Noam Chomsky claims, a "fraudulent research institute".

The first few references make statements like the following:

"Founded in 1985, The Washington Institute for Near East Policy is a public educational foundation dedicated to scholarly research and informed debate on U.S. interests in the Middle East."[28]

This does not sound like a 'fraudulent institute' to me. I will not find confirmation easily by just browsing such endorsements. So, I search. I add **neocon OR neoconservative** to my earlier search. I reason that if someone considers the institute as fraudulent, it is probably in relation to the neoconservative movement, a US interventionist or hawkish political movement.[29] My list of fifty thousand statements shrinks to just 938.

I could search with different terms. In the wind farm example, I simply used **wind farm against.** I just need to find the right term.

With **neoconservative** added, the 938 remaining endorsements discuss the Washington Institute in less glowing terms, some in terms very similar to Noam Chomsky's statement above. Now for the critical step: some of these statements are from significant and trustworthy sources – from academics, international newspapers and the activist scene. Some are definitely not trusted sources. In this circumstance, I suspect statements found on **aljazeerah.info** need not be considered with great care.

I appreciate the work of Joel Beinin, professor of history at Stanford University, in his article, "US: the pro-Sharon thinktank" as it appeared in Le Monde diplomatique.

> "The Washington Institute for Near East Policy influences the thinking of the United States government and has a near monopoly on the supply of 'expert' witnesses to the media."[30]

Keep in mind, the number of endorsements that mention neocon or neoconservative does not signify anything. That was only a means to reach for significant voices making similar arguments as those of Noam Chomsky. Yes, they exist. We have not proven the Washington Institute is a "fraudulent institute" but we are exploring this perspective and it looks less unlikely.

Endorsements move us beyond the control of the author and publisher to the comments of interested bystanders, peers and involved institutions. We learn of prominence, significance and bias. Endorsements also reveal related and comparable information through the link companion, a topic for Chapter Four. We will see endorsements and local context solve many further challenges as this book proceeds.

Time to bring this together. Here is a complete Q4 quality assessment from start to finish.

PRACTICE IN QUALITY ASSESSMENT

There was an online book titled, "Information Research on Internet: Techniques Strategy and Resources" once located at **in.geocities.com/samdarshipali/library/.** Here is a complete Q4 Quality Assessment:

Q1: Yes, delivered in a sane and professional manner.

The writing is good. Sane. It includes a link to biographical details and also a bibliography. Furthermore, the conclusions are well argued and succinct; the language precise. The copyright date indicates 2002 and other indications suggest the information is a few years old but otherwise nothing of concern.

Q2: Yes, the author has the experience and expertise to present this information. It is self-published.

The information is found on the Indian Geocities webfarm, perhaps a concern but the author graduated in Library & Information Science and holds an MA in Economics with some computing expertise. The full resume is actually very impressive and reassuring - quite appropriate as the author of a book on internet research.

Q3: OK, this information keeps good company but nothing to shine any further light on his search experience.

Context does not reveal much on this occasion. Beyond this lengthy book, Samdarshipali also placed some information about MySQL and web databases. MySQL is a fairly technical topic, a good sign, though the website contains nothing else on internet research.

Q4: Hmm ...

We reach endorsements. Early in 2004, a search for **link:in.geocities .com/samdarshipali/library** revealed a listing in the prominent DMOZ directory. This was a significant and prominent link. Beneath this link, we would have found a reference to an article I wrote on how this online book is plagiarized. Yes, Samdarshipali extensively plagiarized my work on the Spire Project. Proof sits at **SpireProject.com/art16.htm.**

Plagiarism is sad and a little irritating. I am actually flattered someone would hold my work so dear as to steal it, then spend hours improving it. This episode was more of a challenge because "Information Research on Internet" gained prominence. This prominence lent Samdarshipali the appearance of integrity, and by extension, tossed my integrity into doubt.

Does it matter? This is not so subtle an influence. When I hold an author in disrespect, I do not believe their conclusions and generally discard their opinions. As this book proceeds, I hope to convince you that

identity matters; that it forms part of our conversation. We will see further examples where endorsements supply some critical piece of information as this book proceeds.

Please note how nothing in our initial quality assessment suggested plagiarism until we looked at endorsements. This is the only dimension of quality assessment truly beyond the control of the author and publisher. Until I published a counter article, the document had convincing quality. Generally, too much of the appearance of quality comes solely from the trust we have in the author and publisher as well as various internal clues. Too much of our quality assessment can be manipulated because of this.

CONCLUSION

I intended this chapter to sketch for you a clear and concise way to tap into the quality dimension of the internet. Q4 Quality Assessment is my current advice. Q4 proves so helpful because it draws us to ask four distinct questions very suited to the internet:

> Q1 • Is the information delivered in a sane, professional manner?
> (Internal Clues)

> Q2 • Does this author and publisher have the experience to deliver this information? If so, with what bias?
> (Author/Publisher Identity)

> Q3 • Does this information keep good company?
> (Local Context)

> Q4 • What do others say about this information?
> (Endorsements)

We could easily ask other questions. We could ask more questions. I like these questions because they reveal most of what I want to know about information quality and reveal it quickly. Easily. I am personally never far from my context bookmarklet, never hesitant to ask the author's identity and always ready to reach for the opinions of others. It is simply something I do when I wander the internet. As a connoisseur of information, I want to know what I consume.

There are other approaches to quality to consider. Internet literature on this topic seems to pool on the Open Directory Project's listing for Web Site Evaluation.[31] Of course I like Q4 most but it helps to be flexible.

For instance, we may require more or less certainty depending on what we do with information. A hunch is enough to place a bet on a horse. Much more is required to declare war on a small country. How much time and effort do we have for quality assessment? If quality means very little to us, we may simply judge quality on format alone, a topic for Chapter Four. Alternatively, if we need something stronger, we delve deeper. We may confirm elements in the author's claimed identity, retrieve previous published material by those involved and discern a history and origin of the ideas present. If we wish to assassinate someone's character, we will need more ammunition than Q4 provides. In an extreme case, particularly for scientific research, we may replicate the original work. Let us align our telescope and calculate for ourselves if that asteroid really will hit New York City. Q4 Quality Assessment is only a tool, though a most versatile and revealing tool.

I must mention one last issue on quality: that of the state of information in our world today. We will encounter a vast range of quality on the internet as well as a vast range of bias. Do not underestimate how much the appearance of truth can be twisted without our awareness.

Years ago I spoke with a man who described how the national GDP figures of Singapore were artificially inflated after their land-value bubble burst. Certainly, we know the Budget Deficit figures for Pakistan were completely rewritten after the 1999 coup from 2 percent to 7.5 percent![32] I am not suggesting the Singaporean accusation is true, rather that such national statistics, widely considered as truth incarnate, are not intrinsically trustworthy.

Another acquaintance previously worked in the management team of an international dental supply company. He described how he routinely approached the loudest of the researchers speaking out against his company's products. He would offer them considerable sums of money to do research into the effectiveness of his products. Well over $100,000 at times would go to support research by scientists publishing reports critical of the company's products. Such substantial sums usually silenced the negative publicity. Do not underestimate the influences brought to bear upon information.

At the other extreme, I am continually surprised by the depth of understanding that a simple expert has of situations that appear complex from a distance. A whole class of experts, driven to understand their field through years of effort and experience, go on to publish vast extensive websites or tell-all books that describe in great detail, information only visible to insiders. Often such individual experts can be identified from the depth of their thinking and the extensive nature of their websites.

I am also continually surprised by talented government staff. I jokingly refer to this class of workers as a collection of highly motivated, talented and impassioned individuals tightly bound up in red tape. They may be unable to publish directly, often for political reasons, yet approach in person or through a personal email and many become fountains of highly prized information. They are, after all, deeply experienced and paid to understand a critical situation – even if they find themselves powerless to improve it.

Let me offer this simple warning. However skilled our assessment of quality and bias, sometimes we will be wrong. When we investigate quality, we will be wrong less often.

Part Two:

INTIMACY

Chapter Four

_____

IDENTITY

A lbert leapt at the brigand's mistake. Stepping on the head of the spear as it dug into the ground, Albert swung hard with the flat of his sword striking his opponent's hand. 'Yield' he shouted, so loud he scared even himself. This would become an important personal victory.

Back in Toulouse, Albert's father was all congratulations. As judge, he presided over the short trial and one less thief disturbed the very important pilgrimage route. The town council was pleased too. Albert confessed to his father how the thief seemed to have tripped but it mattered not. A tale was told of his strength and valor. The young Albert was named a hero.

It was then that Albert was introduced to Captain Robert D'Matan. This aging and accomplished officer, known as El Capitan, looked kindly upon Albert. At his insistence, a tutor was engaged to help Albert with his reading and writing. Once a week, Albert joined El Capitan in the map room to discuss military strategy. Knowing the land well, Albert found he had much to contribute.

The City of Carcassonne, two days ride to the south, was growing. This city petitioned for further defensive works but El Capitan felt the strategic value of Carcassonne castle was minimal and planned to direct more effort strengthening smaller castles further to the south. To understand why, Albert had to learn about the nature of war. While cities like Toulouse and Carcassonne were the economic life-blood to the region, no large city could hold out long when outnumbered. It simply required too many soldiers to defend. This was not true for small castles. Perched high on a hill, a small band of soldiers could hold a small defensive structure for weeks, even months. They could protect citizens in times of raids or disturbances and delay even the most determined adversary long enough to raise a larger army from nearby towns and cities to drive an enemy away.

Together, El Capitan and Albert discussed where to strengthen defences.

If we are not careful, the word 'information' takes on a meaning of some kind of nebulous, fundamental particle. Loose 'data', unstructured and unaffected by anything else. Like driftwood. Like so much driftwood washed up on the shore, we know not and care not where it comes from. It is here. We collect it, step over it or fling it back into the ocean as we please.

Such an image of information blinds us. Every item of information has a history; a history that imparts a wide range of values and qualities upon the information. This history affects the information. We must notice this history if we are to find better information and better appreciate the information we find.

Described another way, information is coloured by its history. Bias, validity and other traits are colours painted onto information during its creation. The raw fact is often only half the message. We certainly cannot trust nor discern its bias without seeing colour.

Albert learned there is more to securing peace than a stout heart and sharp blade. Sometimes it is the show of force. Sometimes it is the public trial, the commendation and the planning. So it is with information. There is more to information than the facts contained. Sometimes significance lies in the context, the format or perhaps some facet of the author and publisher; the source. This notion of context/format/source is central to library science and provides an elegant entrée to information literacy.

I demonstrate this visually in my seminars. I stand up, typically in a crowded library, grab a book off the shelf and rip a page from it. Librarians go nuts at this but it has great startle value. I then thrust the ripped page into the air and declare, "This is what the internet looks like. We've ripped information from all its foundations, from all facets of context, format and source. We present it as an isolated factoid devoid of history."

This page. This ripped page thrust violently above my head. We do not know if it comes from a book, a magazine or a diary. We do not know if it comes from a specialist library, a coffee shop or a dustbin. We may not even pause to think who wrote it. We may never wonder of the publisher and their influence.

Internet technologies rip information from their foundations. It separates supporting details from the document and presents information

without it. Indeed, we can even use HTML to frame someone else's page and present a factoid ripped from the paragraph above and below.

However, the internet does not remove this supporting information far. It is all there, simply waiting for us to notice it; to reattach the history to the information we find.

Every other item of information we encounter in our lives, whether a book by an expert bought at a second hand bookstore, heartfelt advice by a dear friend delivered over coffee or the colourful feature article in a Sunday newspaper insert, all this is richly endowed with a halo of supportive detail. Only on the internet do we encounter information stripped of such detail. Only online do we encounter such raw meat, not just raw but of an unknown animal, of uncertain taste and questionable nutritional value. We encounter anonymous information in an anonymous form without any notion of origin.

Yet this ill-informed nature is of our own making. The halo of supportive detail is all there, nearby, each piece revealed at the slightest gesture or momentary consideration. Once we are adept at seeing colour and reconnecting history, it is easy. We can even reverse this picture. We can describe the colours most likely to answer the questions we pose. We can then seek the contexts, the formats, the sources most likely to have the information we want. We can take our first step towards anticipating information.

Learning about context/format/source – learning to see colour – is a journey into the nature of information; a journey as applicable to internet information as to library and commercial information. It applies to all information – even asking a neighbor's impression of the next footy game. This journey traverses very familiar terrain. Indeed, try to notice just how familiar it feels since yet again, we are not learning new techniques so much as transferring familiar techniques to the internet environment. There should be few "Ah ha!" moments and many more "Well, of course!" occasions as we run through this chapter.

CONTEXT

Just as we judge a person by their friends and judge a magazine by its articles, so we judge a webpage by the company it keeps. We find it

embedded in other information. We find a single work among many. Nearby information informs us.

In our discussion on quality assessment in Chapter Three, we interpreted context to mean webpages on the same website, on the same computer directory and so presumably by the same author and publisher. Such related, neighboring information demonstrates the bias and skill of the author/publisher. Critically reviewed, such information helps reveal information quality.

There is, however, another, less restrictive, less narrow interpretation of the term 'context' to consider.

Location on the internet can mean more than physical location. We can also describe location in a relative, logical manner. Cyberspace is a four or five dimensional space, stretched and twisted by the humble link. Links draw items of information close together. Once distant and unrelated, such information now sits only a click apart. While perhaps physically positioned on computers located in different countries, once I link to your website, your website is found from mine. We are neighbors, friends even. And this friendship has meaning.

Pages linked together are related but not in the same way as pages on the same website. Clearly they don't share the same author/publisher. Instead, links make a statement; a kind of endorsement that states more than just, "Nice site". Consecutive links also share topic.

I want to suggest a strange yet also obvious notion. How we find information reveals something of the information we find. Like a tour guide's commentary on a bus touring through town, as we travel, we listen to a subtext telling us of the information we are visiting. We have covered the best part of this subtext already – notions of prominence, Q3 Context, Q4 Endorsements. Other elements of this subtext are rather banal – pages found through cool.com are probably cool. Pages found from a primary school child's webpage are probably pitched towards primary school children.

This subtext is simply context – the book bought at a second hand bookstore, the advice by a dear friend delivered over coffee and the feature article in the insert of a prominent Sunday paper. All this subtext, this context, reveals something of the information we encounter.

Listen to this subtext. It enriches our journey. It offers us ways to refine a search in progress before we look at a page. This notion of context

delivers two further gems too; two ideas that will help us find better information. The first is the information venue, the internet's equivalent to a specialist library. The second is the link companion, a way to find comparable information. Context has a marketing application too: the idea of the internet footpath. All three of these concepts deal with pages away from the page we are on; the page next door, pages nearby or pages twice removed.

THE APPROACH TO OUR PAGE

Start with the journey. In an environment where we jump from site to site so very quickly, it is tempting to postpone the use of our brain until we reach a page that interests us. Idle until we reach our destination. Only then shift into gear and critically read a document.

We can do better. Reach a page by way of an internet directory or the recommendations of a global search engine used in a blunt manner and the page has prominence. It has fame, fortune and a presumed quality. With practice, we notice as we approach prominent pages. We can use this to check we are searching effectively since we want to avoid prominence to answer certain questions and desperately need it for others.

We can also discuss a feel or impression. Something we find on Wired.com is probably fairly topical and cool in a techy sort of way. Something suggested by a youth message board on graffiti probably is cool in a youthful sort of way. "CEOExpress.com. Connecting busy executives to information that matters." Hmm, probably directs us to serious projects. A link to a speaker on instant success. Strange how this page serves no purpose beyond directing visitors to their seminar. No review, no advertisements; not saying anything speaks very loudly of marketing. I am already meeting the speaker and the speaker has nothing to say.

As we move to click a link, the page we are on already reveals some aspect of the destination. This is the magic of context.

Step away from the internet and this kind of context is everywhere. A dusty pile of magazines in an op shop. The reject sale pile of a local library. The swish new digs of an over-puffed friend courting an article in Architectural Digest. Each image I paint has meaning independent of the words that follow. Each image forms part of the message.

Such context adds spice. It adds pizzazz. Why listen to a public speaker if we ignore how the speaker walks, dresses and inflects their voice? Why fall in love if not to see the morning light dance upon a delicate cheek of a lady in slumber. If it's night, light candles instead. Context is the essence of life. And to dramatize this point excessively, we risk stripping that life from internet information when we jump straight to the words on a page without noticing anything of our journey.

Learn to notice context. Every step of the way, like a breadcrumb trail, clues tell us what we are about to see. Stay attentive and listen. For example, while search engines speak loudly of prominence, they also speak of a muddy or clean nature – something we will address again, later in this book. Briefly, a clean search indicates little confusion between what is and is not on topic. We search and most of the search recommendations yield relevant, on-topic results. A muddy search is the opposite: a search where we find only the occasional suggestion helpful in answering our question. We search but the search engine recommend mostly inappropriate, irrelevant and off-topic results.

Clean and muddy are traits of both our search and the information environment. It speaks to us of our search words, of the information we will find and of the wider information environment we are working in.

While it is tempting to consider context as a discrete collection of statements – of endorsements, prominence, a clean or muddy nature – there is a holistic nature to it as well. We help ourselves by considering context as full of nuance and meaning rather than a checklist of possibilities.

INFORMATION VENUE

One discrete element of context involves topic. Context draws comparable information into orbit together. More precisely, people like you and I link to interesting resources on the internet and in the process, draw comparable information together. The best of these baskets of comparable information I call 'information venues'.

The information venues resembles the directory but our focus is on the best list rather than most extensive. Large global directories do not often make the best information venues. They are fine venues for general

questions but better, more focused lists can usually be found produced by someone with demonstrated expertise in a topic.

This is fairly obvious. No surprises yet. But watch carefully because one page from now, I will rip the carpet out from under you.

Stepping away from the internet for a moment, consider the specialist library of a government agency overseeing childcare. This library holds many books, periodicals and articles on the topic of childcare. This venue for information contains many fine resources from many fine and varied authors and publishers. Everything relates to the topic of childcare.

The Electronic Frontiers Foundation (EFF) is an association deeply concerned with building and protecting electronic freedoms. As a small part of this task, the EFF maintains a well read and appreciated archive of publications relating to internet freedoms and legislation. The part of this collection at **eff.org/Privacy/** discusses privacy and encryption. Since privacy is an internet topic, as expected, most of the material is present in a digital form. The articles, working papers and submissions to US Congress are all electronically shelved together.

Superficially, this EFF collection on privacy and encryption is little more than a list of internet resources and publications on a webpage. However, look closely and it shares all the same traits as the specialist government library devoted to childcare or the private library of the association of Wind Power Producers. In particular:

- All the publications share the same topic.
- All items are selected and vetted.
- We should expect a selection bias.

Do not overlook this last point. An association of wind power producers may stock their library largely with very pro-wind power publications. This is not unprofessional, so do not consider it a nasty surprise. Selection bias simply reflects their interests. They want quality resources, so shallow and poorly argued publications are placed on their shelves. They want it on topic, so everything is about wind power. They also want resources that assist them to express their opinion. In addition to many technical and scientific publications, we would expect to see many supportive publications. To counter this bias, merely visit the archive of an opposing view.

The most obvious strength of an information venue is that all the information shares the same topic. Next to a book on childcare, shelved in the childcare library, we find another book on ... childcare! And next to that book, yet another!

Now here is the critical step: find a book we like. If we look around and notice we are standing at the shelves of a specialist childcare library, not only have we found a book we like, we have also found a collection of other books on the same topic, many we will probably like as well.

Let me phrase this using our new vocabulary. When we find an internet document we like, we can use the context of this document to find comparable resources. Welcome to the link companion.

THE LINK COMPANION

Book in hand, we look up and notice we stand at a library shelf stocked with other books on the same topic. Like our earlier examples of local context and endorsements, here again is a simple, almost obvious, empowering observation.

According to context, an internet document sites close to a great many places – any place that references it. In the same way, a booklet of events in my city sits on the shelves of the tourist information centre, sits in the lobby of a backpackers hostel and lodges with other brochures in the state library.

Some of these locations are information venues – sites that specialize in bringing together fine resources on a particular topic.

If we can find an information venue that references a fine document we like, we have found a list of comparable resources. With one good resource, and the help of an information venue that references it, we can find other, similar documents. We can't do this easily in real life but on the internet, this is simplicity itself.

Three steps:

Step One:
Select a good document on a topic that interests us.

Step Two:

Retreat to a suitable information venue that lists this initial document.

Step Three:
Move forward to other comparable resources as listed by the information venue. Look first near where the reference appears to our initial document.
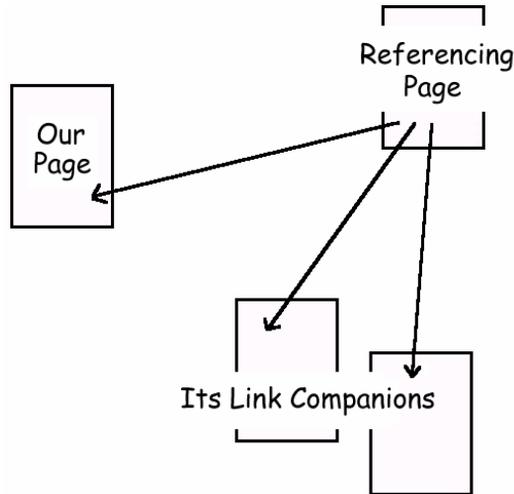
Once you do this a couple times, you will see this really is a simple process; not cumbersome, just counter-intuitive. We essentially move backwards, then forwards again. We zigzag through the internet.

As our example, suppose we find and devour the delicious Internet Privacy FAQ. We want further resources of similar calibre so let us use this FAQ to find another resource or two on the same topic.

Step One: Search for sites that link to or mention this FAQ.
We search for: **"Internet Privacy FAQ"**
or **"www.faqs.org/privacyfaq/"**
or **link:www.faqs.org/privacyfaq/**
or we simply click the endorsement bookmarklet.

Step Two: Choose a promising information venue.
Third down the list I see the About.com's list on privacy. It links to the Internet Privacy FAQ, which is why it appears on the list.

Step Three: Browse About.com's list for similar resources.

For our second example, the Electronic Frontiers' special library mentioned two pages ago will prove helpful. Say we find and read the work of the Privacy Foundation (**privacyfoundation.org**). It is good but incomplete for our purpose. We want more on this topic. We want other, comparable resources.

To find them, we look for places on the internet where the Privacy Foundation is mentioned. We look up from reading the Privacy Foundation and notice where we are. Crowded around us are hundreds of assorted webpages, several of them information venues, that mention or reference this website.

One place among these many sites is the internet privacy archive of the Electronic Frontiers Foundation (EFF). To use our metaphor, we look up from reading and see if any specialist libraries happen to have a copy of Privacy Foundation material. To our delight, we find the Privacy Foundation sits 'on the shelves' of the EFF privacy archive.
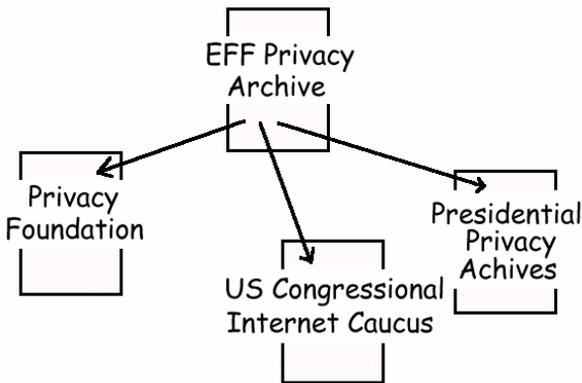
The EFF also happens to shelve two further documents immediately beside their copy (well, reference) of the Privacy Foundation: the US Congressional Internet Caucus and the Presidential Privacy Archives. These are comparable publications, according to the editor of the EFF's internet privacy archive. These are two similar resources, similar to the privacy foundation archive.

I use the term 'link companion' to describe this relationship. Two resources are not linked together – the Privacy Foundation does not link to the Presidential Privacy Archives – but links to both appear together on the referring page by the Electronic Frontiers Foundation. They are tied together. The EFF shelves them beside one another; mentions them in the

same breath. They are linked once removed as it were. They are second cousins.

The link companion emerges from a wider view of context. Information we find on the internet is positioned close to many other resources all over the internet – anywhere it is linked or mentioned. The Privacy Foundation is on the shelves of the EFF archive, the shelves of Yahoo's list of privacy resources, on Joe Nobody's personal page and many further places across the internet. That its physical location is at a specific web address on a specific computer is unimportant. Logically, the Privacy Foundation is next to other sites that in turn link to comparable resources.

I first uncovered the link companion three years ago while observing reference librarians of the Western Australian State Library as they provided online assistance to users of the Australian AskNow! service. The reference librarians are charged with leading internet users to at least two resources that should answer or at least advance their question. The link companion became a way to turn one good resource into a list of further likely resources to consider.

I have five pieces of advice to share about link companions. Firstly, using link companions is a genuinely alternative way to move through the internet. Instead of searching for information based on criteria we supply, we spot one site or resource we like, then gather comparable resources based on someone else's reasoned, personal assessment of what is similar. We reach very different material in this manner.

Secondly, this technique feels much like the constant surfing of further resource pages once so familiar to early internet users. Once upon a time, eight or nine years ago, surfing was a valid way to find information. Early web publisher took great pains to alert us to the best comparable sites. And who better to decide what was significant than those who publish on that topic? We cannot do this anymore since so few publishers put much effort into further resource pages. However, link companions serve the same purpose.

Thirdly, this is not a cumbersome technique. We essentially zigzag our way through the internet. Instead of stepping forward to the next page, we step backwards, then forwards again. Once we have our work desk set up properly – the topic of Chapter Five – this will involve only a click on a

bookmarklet, then a glance at a list of candidate venues; seconds not minutes.

Fourthly, this technique only works well when we have a great initial page to work from. We want a page people talk about. We want a page that almost tells us what we want to know. If we just grab link companions for the first page we come across, better to just toss words at a search engine.

Lastly, we also want to choose a fine information venue. The best link companions come from information venues created by publishers strong in their topic. This is not usually the larger global directories like the Yahoo! Directory or the ODP. A good page for link companions, very much like a fine specialist library, is often not particularly famous but has that ever so desirable quality of specialist experience.

With our last example, just how helpful the EFF's list of privacy resources will be depends on how much care goes into the preparation of their list of resources, on how the archive is organized and a few other conditions. The more effort the EFF puts into their list, the more likely it will become popular and prominent too but that is a different issue. Prominence depends more on the reputation of the EFF, the time it has been online and any past promotion of the archive, not just on intrinsic excellence.

The most promising venues will tend to stand out from a list of links once we learn URL Interpretation from Chapter Seven. They look like lists, not articles or discussion pieces. Better link companions come from sites produced under the utopian publishing model (more on this later), from sites that show a selective approach to linking and from sites that demonstrate their experience.

In practice, I gather link companions only when I have a specific need, not a general interest, and only when a specific search proves clumsy or muddy. If my interest is general, I will just search for a prominent site instead. If I want an alternative viewpoint, then I do not want a comparable site and link companions would not help. If my search descend into a hunt for a good resource, I stop and try another way.

When you first search for link companions, you may find it confusing – no better than simply tossing keywords at a search engine. Do persist. At times, this is a very elegant way to search. It is worth our time learning when link companions work well and when they do not. I find link

companions either help quickly or not at all. I use this technique in perhaps just five percent of my searches but with splendid results.

Almost all internet information will have link companions just as all published information away from the internet usually lodges in some folder, bookshelf or specialist library. It is the magic of the internet that allows us to know where information lodges with so little effort. And where it lodges, we find other resources on the same subject.

Our next topic deals with what we actually find and echoes once more, how the real world and the internet behave in much the same manner.

FORMAT

Information weighs heavily in our lap, blasts in our ears and assails our eyes. We easily categorize information by how it reaches us; by its presented appearance. This is a printed book. That is a newspaper. Please speak quietly so I can hear the news on the radio. We easily classify information in this way.

Far more significantly, look to how information is prepared. We will use the term 'format' to describe a very specific vision of these forms – a vision that divides books from articles, from research reports, press-releases, sales brochures, memos and discussion items. Each format is different, distinct and provides clues about the information within.

There is one critical point to our definition. 'Format' is not tied to the physical presentation of the information. Format describes the logical preparation. When I mention the book format I am not referring to the physical object – a heavy bound pile of compressed tree pulp full of words printed in ink. Instead I refer to the concept of a book – a lengthy document prepared by an author, carefully edited, then published by a publisher with a profit motive.

The list of formats does not include 'internet', 'digital' and 'print'. These words describe presentation. Format is not tied to specific internet tools either. There is no email format, no database format, no WebTV format. There is especially no web format.

Because of this, format is distinct from filetype; from .doc, .pdf, html, textfile and .ppt. Filetype is one of the minor fields found on many global search engines. A Google search that includes **filetype:pdf** will reveal just

the pdf files matching our query. But pdf is not a format. pdf is the manner this information is presented, not the manner it was prepared.

Format so defined is a powerful concept. It ties our experience with non-internet information directly to the internet world. A book, whether presented as a physical object or an electronic version, embodies all the same qualities and characteristics. Both are lengthy documents, prepared by an author, carefully edited and published by a publisher with a profit motive. News articles, whether in print or online, have the same qualities. Both are short, sharp, abundant, timely and of less than certain factual quality. These qualities emerge from how the information is prepared. These qualities are inherited from the process that brings the information into being.

I mentioned that format is distinct from internet tool. We access the internet using tools with names like file transfer protocol (FTP), Telnet, Newsreader and Web browser. New tools continue to emerge with names like Internet Relay Chat, BitTorrent and RSS newsfeeds. Each tool uses a particular protocol, may have custom-built software and may share common qualities. Thus when an internet address starts with ftp:// – the hallmark of an ftp resource – we can make several assumptions about the information.

Indeed, all ftp resources share some qualities. They all look alike for instance. In the early days of internet searching, the internet tool was very significant in searching but this concept has become too clumsy to tell us much anymore. Information moves from one tool to another too easily. Shareware, once distributed almost exclusively by FTP, is today largely retrieved through the web. Shareware is a manner of preparation, a format. FTP is an internet tool; one among many we can use to retrieve shareware.

We could say an email is a format and we probably will by mistake. To be precise, the format is a memo or note. All memos bear similarities in how they are prepared.

This difference is clearer once we look at the sms (short message service), a format as well as a tool. The sms is a small text message painfully punched into a mobile phone touchpad and intended for convenient retrieval by another mobile phone user. At this time, the sms is largely tied to the mobile phone. Of course, I can also send sms messages from a webpage courtesy of my internet service provider. I wrote a

computer script that relays seminar reservations to me via sms making use of a commercial sms gateway. At the same time, my phone company lets me send an sms to an email address. With this in mind, if I send an sms message from a webpage to your email address, where is the mobile phone? Is it still an sms? The message certainly resembles an sms: a very short text coded with funny misspellings like U instead of you. I prepare the message in the sms format. It just reaches you without ever touching a phone.

This separation of tool from format occurred sometime ago with the fax. I may prepare a fax that never touches a fax machine. I scan a document, send it by email to a faxgate and from their to your fax number – which may be a virtual fax meaning the document reaches you attached to an email message. No fax machine. No paper. Yet it is still prepared in the fax format – a black and white graphical representation of a page of paper suitable for signed legal documents. The presentation changes. The manner of preparation has not.

In an earlier era, we paid great attention to tool and filetype. We classified information this way. It was part of our excessive focus at the time on the internet as an extension of computer science. Unfortunately, this had the unintended affect of distancing internet resources from almost identical non-internet resources. It failed to present the internet as an extension of the information world. Ebooks are not books. Certain webpages are not articles. Instead, attention to presentation directs us to what all web material has in common. Except that pretty much anything can be presented as a webpage or embedded into a webpage. Knowing it is a webpage tells us exactly nothing.

There is no format called the web. The web is a voracious creation, consuming and presenting information from so many formats. Newsgroup discussion can easily be retrieved through the web thanks to Google Groups – a powerful searchable database of newsgroup discussion with an esteemed lineage. Email posted to a mailing list often gets archived. We can view it as a webpage. Does this somehow change the qualities of the discussion item when we see it as a webpage? Clearly not.

The notion of format structures this story differently. Almost all information exists before it emerges on the web, perhaps as an article, a patent, a report, a book or a news item. Even when no prior existence can so clearly be identified, the words and ideas are prepared in a way almost

identical to a non-internet method: perhaps as a sales brochure, a phone call or a memo. As we follow this logic, the internet becomes an extension of the larger, grander information world. We tear down the walls attempting to convince us that the internet is something completely new, something with which we have no useful prior experience.

Format is elegant. Internet information is deeply organized by format. In research, we often start by judging the format we want for the information we seek. We may also find it helpful to predict the format our answers will most likely take. As we proceed, I will show you how the format can often be guessed from a great distance – if not from the web address than certainly from a quick glance at the page. Later, when we discuss the sociology of information, this notion of format permits us to watch the migration of information from print to the internet and how various formats compete for relative dominance.

We can argue another time if the web has spawned any truly novel formats. The blog, a special kind of online diary, certainly seems unique. We can also argue if all online information shares certain traits that make it distinct from their non-internet brethren. Immediate access springs to mind as a defining feature. However, at this moment, let us just explore this concept of format a little further.

THE BOOK

Books have such meaning. They symbolize distilled knowledge. They signify expertise. We are emotionally invested. Rip a book in public, an old book with just one interesting chapter. Tear out the pages and even strangers will loudly express their exasperation.

Yet books are just one distillation of knowledge. Books are large, lengthy tomes, heavy with depth and detail. They lack sharp criticism and cannot address emergent trends. In research they ideally suit general questions but make horrible resources at other times. Much is right about books but for each brilliant quality, they suffer a significant shortfall. They are old. They date badly. They have a limited life span yet physically last very much longer. They are long troublesome things to read. They cost money.

These qualities emerge from the way books are created. An author gives birth to a book, investing a great deal of time preparing a lengthy

tome. The size of a book allows for far more detail than other information formats; this size drives the author to present a complete, balanced and holistic perspective. The author and editor both invest considerable time preparing the best quality possible. Spelling and grammatical errors are unacceptable. Equally so are legally grey accusations, poorly constructed logic, fluffy narratives and amateur translations. The publishing of a book involves many people working together to create and present a lasting document of considerable length. There are financial considerations. The work, above all else, will be sold perchance for a handsome profit.

A book usually becomes available to read at least a year after the publisher enters the scene. And the finished product is not created in the month preceding its delivery to the publisher. Many books have their origin five or even ten years earlier. This book began with a seminar I delivered in spring 2000. Such lengthy preparation is sharply at odds with other information formats. A press release or news item may be created the very day we read it.

In summary, the book format tends to be old and detailed. It excels at presenting a considered overview.

THE PRESS RELEASE

The press release is a document either promoting an organization and its accomplishments to the media or explaining and mitigating a potential public relations disaster. The document's purpose is to entice the media to write a story. Thus, the press release tries to capture fleeting attention with a catchy title and an engaging story line. It then provides the background a journalist would need to write a story.

Catching the eye of a journalist is not easy. Too many organizations in our world would love to have more positive media stories about their accomplishments. This market is highly competitive and news writers are very bored. Tens or hundreds of press releases will cross their desk daily. From this attention overload arises a highly structured format that helps the journalist quickly scan for relevant information. Each element is highly dictated by custom and purpose. A little less than a page in length, start with the organization issuing the statement and a date. Short. Sharp. One story and one story line only described in just a few words.

Press releases are distributed in three ways. Firstly, they are faxed or emailed to specific journalists. One task of a public relations department is to cultivate press contacts, and once cultivated, bury them in press releases.

Secondly, businesses called wire services specialize in distributing press releases to the media. BusinessWire and PR Newswire are some of the largest. For a substantial fee, wire services will include a press release in a daily ritual with hundreds or thousands of other press releases made available to the media around the world. There are smaller newswires to fax or email our press release to journalists in our area or journalists in our field. Techwire, U-wire (university campus papers), VentureWire (venture projects) – the Yahoo Directory keeps a long list of such wire services. Each for a fee will distribute our press release in the often vain hope that the media scanning today's list will rest an eye on ours ... and write a story.

Thirdly, press releases are archived as an information resource. They often appear in the media section of large corporate websites. Internet search engines index some and commercially prepared company profiles may include press releases as well. The larger wire services archive past press releases in vast commercial databases that stretch back years.

The press release is a highly specialized format designed to capture the attention of an audience of journalists mildly interested at best. It is a format that demonstrates focus and tries to promote.

THE NEWSPAPER

Perth, Western Australia has a single dominant daily newspaper called The West Australian. It costs one dollar and is available for home delivery. Perth also has a collection of free newspapers delivered to specific suburbs. Where I once lived we received the Guardian Express, the Voice News and several further free newspapers like the WAX (West Australian Xplorer), OUTinPerth, SHOUT and The Parents' Paper. These four focus on events in Perth and come out less frequently.

With just over a million souls, Perth has just one prominent daily newspaper. Why? The answer is a tale of business development, a tale best told from one business coach to another describing how newspapers build,

purchase and struggle to attain primacy, then defended their position, as businesses do, ensuring other daily newspapers cannot succeed.

From a searcher's perspective, the existence of just one prominent daily newspaper is of considerable importance. If we are interested in a regional approach to information, the newspaper ranks very high as an avenue to explore. With just one prominent newspaper, we have only one door to open.

That Melbourne retains two significant daily newspapers should not cloud the importance of this generalization. A daily paper does not present the same news as a local fortnightly free newspaper. They differ significantly … and we know this already.

Consider the world news section of the newspaper. International news reaches us at the end of a sophisticated chain of supply with information swept from a far-flung journalist to our local newspaper. Journalists submit an article to a newswire. Newswires sell subscriptions and reprint rights to local newspaper editors. Editors select news and incorporate it into their world news section. Your local newspaper has many staff writers for local news and these journalists may in turn occasionally submit articles to newswires when international interest exists. Your local newspaper has few staff writers in foreign locations. Most foreign news emerges from journalists with little or no relationship to the newspaper we read. Many of these foreign articles are attributed to AP (Associated Press) or the New York Times or another of the international newswires.

A lovely twist to this supply chain occurred early in the life of the internet when these international newswires began publishing their stories directly to the internet. They essentially began competing with the newspapers they supplied. Various market approaches were tried to move news directly from newswire to consumer but they did not work as effectively as the existing journalist → newswire → newspaper system. Agence France Press, the primary French language newswire, eventually withdrew from publishing free to the internet. Today, news is readily available from newspaper websites and various portals.

We shall continue this story later but the significance of newspapers rests in how this structure occurs in all cities. Everywhere, the newspaper has a similar role to society, a role driven by the nature of newspaper news; by how it is created and distributed.

SERIAL BROCHURES

A very different mechanism is at work with the serial brochure. Here, financial concerns are not paramount, certainly not at the start. There are expectations of financial justification later down the track but when they begin, the advertiser/publisher is willing to gamble. Thus, we have 'periodicals' sent to us by our phone company, by our bank, and if we have a new child, then by our baby food supplier too.

Unfortunately, this type of information suffers badly from poor editorial concerns. The information is so completely biased by the interests of the publisher that the information is almost useless. And because the audience is characterized in such a superficial way – "all telephone users" or "all parents with young kids" – the information contained in these periodicals tends towards being general and superficial. Further, since the content is intended to represent the company, there is a serious avoidance of contentious or confronting information. Over time, as the financial returns weigh more heavily on the management of serial brochures, the quality of articles, the quantity of advertising, the acceptance of far more advertorials and similar steps reduce the package as a whole. As an information consumer, I look to these periodicals with disdain.

As an aside, this is often the same trajectory of many websites. Much of the internet is prepared in this format.

Not that such an effort need be so poorly executed. There is no need to limit such a periodical to just one advertiser. The publisher can improve the content in ways to interest further advertisers. The difficulties that plague this system of communication can be surmounted. We know this because of the inflight magazine.

The Australian airline, Qantas, publishes a monthly magazine of very high quality. They pay international authors good money for good articles – though often buying only second or third rights (the articles have been previously published elsewhere). The audience is interested in travel and is generally wealthy. The magazine carries excellent articles on travel destinations and enjoys abundant advertising from a range of top end retailers. Thus, with a captive audience, a wealthy audience, with good advertising and minimal distribution costs, Qantas has a success. Indeed, the abundance of advantages contributing to their success may highlight the difficulties in presenting a meaningful serial brochure. Most serial

brochures do not have such abundant advantages. Few serial brochures reach such lofty heights.

A MULTI-FORMAT WORLD

Newspapers, press releases, books and articles; the information world has not always been so complicated. Complexity is a gift of the post-modern world as are a multitude of new formats it makes possible. If we look way back in history, back when few people could read, the popular modes communication were few indeed: the troubadour, the orator and the sermon. Egyptian scribes allowed for popular use of the written language and we have an abundance of tender letters, banal receipts and religious writing to attest to this.

Consider the format of the etched stone slab prominently positioned for public attention. Hamurrabi's Code of Laws (circa 1780 B.C. according to the Louvre) is a black slab of basalt etched with a very early legal code. King Hammurabi arranged for such monuments to be placed in prominent and frequented locations throughout his kingdom and so reach his subjects. The famous Rosseta stone, a message in three languages, is a similar, later example.

Art conveys messages too. A lord, cardinal or pharaoh depicted in some act of devotion beside instantly recognizable religious personalities conveys greatness and piety. Such leaders obviously deserve respect and loyalty. After all, how many of your friends have archangels personally arrive to chaperone them to heaven? Many a gothic painting shows just such a scene. We see this again in the deep relief carving of an Egyptian pharaoh in the company of gods, blessed by the warm hands of the sun.

Modern paintings often tell exquisitely confronting social messages. Francisco de Goya's painting "The Third of May 1808, Executions on Principe Hill" chronicles the horror of Napoleonic Troops oppressing occupied Spain. This picture is far more eloquent than any book, statue or slab of basalt. Here in Australia I remember the modern art installation "Christ in Piss". The piss, I vaguely recall, was revealed as lemonade only a year later. Great controversy surrounded this art piece. It cleverly challenged viewers to consider their response to religious heresies. Art can be a very sophisticated medium of communication.

Works of art. Slabs of granite. What has developed over the last few millenium is a focus and clarity to our communication. Modern and post-modern formats like the press release and the blog are highly focused and efficient. Over time, formats also change. Painting, once a method of realism, was pushed firmly into the realm of art. The discussion piece drifted into the electronic world, collecting and perfecting a sense of immediacy and an invitation for public comment.

This vision of format is immensely potent on the internet for two reasons. Firstly, much of the internet is organized by format. Start with a book in mind and we certainly have a clearer idea of our path and destination. Significant book databases spring to mind like Amazon.com's book database, the Catalog of US Government Publications (CGP), the library catalogues of the great national libraries of the world and the commercial database, World Books in Print. Similarly, if we decide to chase an article on Kashmir in a political science journal, we have already sketched our journey in the sand. There are lists of such journals, established ways to find them and a certain look and feel to such resources. We know the path forward. And this path has little to do with approaching a global search engine with several keywords in hand.

Secondly, once we determine the format of a given webpage – often before we even look at a webpage – we know many of the qualities we will find. Newspapers are current, light, biased and abundant. Books are old, detailed, considered, an overview. Recognize a webpage was prepared as a news article or a book and we know so much about the information within.

Recognition is the key. Recognize the format of information we encounter and recognize the format we seek. Let format guide us to the information we seek. Before we continue this discussion, however, we also want to recognize the author and publisher as well.

SOURCE

His fame was short-lived. Strike. Parry. Thrust. Albert spent much of his fourth year hating his profession. The physical effort left him constantly exhausted as he learned just how unskilled he was with the sword and bow. Furthermore, his skill did not improve much as time passed. Respect from his peers came only later, paradoxically, once he no longer sought it. On his first military campaign, he

realized how little he did know of soldiering. Spending so much time preparing and waiting, he grew to understand how patience was as important as how he held his weapon.

After Albert returned, scarcely having seen the enemy except in the distance, he grew more at ease with waiting. He relaxed into the task of learning how to defeat an opponent with a sword. He also grew resigned to the knowledge he would be lucky indeed if his ultimate death had anything to do with a sword. Most likely, he would die of a horrible disease brought on by poor food and shelter, waiting for a battle to begin.

<p style="text-align:center">*   *   *</p>

The author writes. The author perches lightly on the edge of a chair, delicately scratching on a pad of paper. Or perhaps the author lounges along the length of a sofa like limp salami, banging away at a keyboard. The author's words appear in the finished product. The author's nuances enliven the work. The author's depth of understanding reaches us through the writing. The author's bias colours the words.

Authors wield tremendous influence on the information quality. These qualities are varied and include such aspects as spelling and punctuation, plot development, metaphors, logic, persuasiveness and more. An author may try to convince us with an impassioned plea, a quilt of circumstantial evidence or a logical mathematical-style proof. This reaches beyond how it is recorded to touch on what is recorded. Personal bias, exaggeration and the selection of statistics may generate a biased view. These influences originate in the author. Authors choose words.

These choices also reflect on the author. We make certain judgements about information based on their author. We make judgements about an author based on their writing. Creator and creation are interrelated. Is the author excessively biased or deeply flawed in some way? One simple approach to answer this question is to assume the author is professional and knowledgeable. Now downgrade this assessment each time we see poor spelling, bad grammar, leaps of logic or convoluted twists that do not ring true. As our impression drops, we move closer to just discarding the work and considering the author a crackpot or a naïve idiot. We reach a point where we are unwilling to believe their scholarship.

Under other circumstances, we approach this completely the other way round. Authors must prove, quickly, that they have the professional

experience and knowledge to be responsible. What are their credentials? Are they new to this field? What else have they accomplished? What do reviewers tell us? If not suitably impressed quickly enough, we discard the author and their work as too amateur to be of value.

Thankfully, there is more to judging authors and information. There are vast vetting systems to limit the audience of a poor and confused writer. To publish a journal article, we must first convince an editor and knowledgeable peers. To publish a book, we must usually convince a publisher to risk money on our creation. Vetting is important but in no way does this diminish the importance of the author's craft nor diminish the importance of us asking about the author.

Now let us consider the publisher. A publisher has many roles including financier, editor and promoter. Usually, publishers have a role in establishing the boundary of the topic too. Occasionally this extends to dictating conclusions. Publishers often establish certain quality standards they expect of all the writing they are involved in. Publishers also oversee the publication process including the printing, or in the case of the internet, the preparation of webpages.

In the information sense, the publisher may be a journal editor, a book publisher, an association, a company or a government agency. The publisher may simply be the author. In all cases, someone or perhaps several people, will provide the financial support, editing and promotion for the work created by the author. An important task in appreciating information involves identifying the publisher and establishing their interests.

On the web, the author may take many of the traditional roles of the publisher. The author alone selects what to publish. The author codes and places the information online. This change is neutral though. Just as a publisher may add bias to a work, so the publisher can counter bias and drive the author towards greater clarity. Self-published work often lacks the wider perspective and polish that carefully edited traditional publishing contains.

Determining the potential bias of a publisher can be vital to assessing the value of an item of information. Determining some statistics on smoking habits are published by a tobacco company should at least give us pause.

Thankfully, it is simple to gather such information about a publisher. The website we are on will have some information about the author and publisher as covered in Chapter Three. Context, endorsements or a direct search will lead to the rest. The important step is recognition. As with format, learn to recognize the author and publisher. Consider carefully before spending time reading anonymous information.

VETTING

Context/Format/Source: a foundation of library science. Here is how we often determine the value of information. A book, prepared by an accomplished author, published by a respected publisher and purchased through a fine bookstore is probably great. The same book, this time by an unknown author, published by a vanity press and delivered through mail order is likely to be far less valuable. Why? Well, if the book is that great, why is the author unknown? Why the vanity press? Why mail order? A lack of indications of high quality is often enough to judge something as lacking in quality. Such vetting occurs to all information passed to us by an intermediary. This vetting is simply a selection process based upon someone's quality assessment.

Libraries and bookstores wish to maximize the value of their collections – wish for this deeply – so libraries and bookstores vet the books they purchase. They check to see if they are good books before they buy them. An acquisition librarian may make an assessment based on the recommendations of library suppliers, publishers, best-seller lists, award winners and what is local or in the news. Many libraries hand a significant portion of this vetting role to library suppliers.

This sounds highly informed but many books are still not vetted from a superior knowledge of the book. Certainly few bookstore managers have the time to read a book before they stock copies. Those who purchase books often select books because they are on certain catalogues by respected publishers, are on various best-seller lists or because they win an award. Book publishers select the best books for publication, then put most of their marketing budget behind the most promising. Magazine editors cultivate relationships with the most interesting writers, then put the best article on the cover. Writers report only the most interesting stories. This chain of vetting clarifies the information surrounding us by

removing from view information with less interest and value. This chain of vetting spins our way information with higher quality.

Vetting is not, however, fair. Not all excellent books get published. Not all ground-breaking science gets publicized. Not all interesting stories get reported. Vetting is rarely a just assessment of the facts based on expert opinion and critical thinking. Such consideration takes time and effort. Instead, vetting is often just a quick assessment based on the most obvious features of the information. Did the book author win a Nobel Prize for Literature? Is the author accomplished and respected? Do we agree with the premise? Do we like the cover? If not, let us read something else. In a very similar way, the peer review process found in prominent scientific journals and magazines displays a clear vetting bias towards established and understood achievements.

These vetting chains are often deeply affected by an overwhelming range of alternatives. Over a million new books are published annually. Many more articles are offered for publication than are accepted. The most prominent peer-reviewed journals and magazines are vastly over-whelmed by worthy research projects seeking the publicity publication supplies. Even if we think vetting is too severe, out of necessity, vetting must choke off some worthy projects. As searchers judging quality, we should keep this in mind. The field of third-world aid, for instance, has far more worthy projects than newsprint can draw to our attention. Not having news coverage does not mark an aid project as less deserving or less worthwhile.

The process of selecting information may be sophisticated and informed. The people and organizations involved in vetting may have the finest motives at heart. However, this process still permits bias and uncertainty. Many a significant advance has failed to find acceptance in scholarly publications. Many a fine book has failed to attract critical acclaim or found success only after repeated failure and rejection. In reverse, some bad ideas are published in peer-reviewed journals. Even outright frauds occasionally occur.

The beauty in systems of vetting is not in their perfection. The beauty is that they work most of the time and usually steer us towards better information. Vetting makes the best of a bad situation.

The reason for this is that vetting may involve very little thought. Without time, most quality assessments retreat to simple questions like

"What else have they done?" (context), "Is it a book or an article I am looking for?" (format), "Do I recognize this author? This publisher?" (source). It simply takes too long to do better.

Information that attempts to circumvent our established vetting systems rarely reaches our attention. I do not read books published by vanity presses. I do not believe technological advances introduced on sales brochures. I do not believe in the quality of the generic Viagra substitute as proclaimed by the intrusive spam that arrives daily to my email box.

While vetting is not necessarily fair, a quick vetting is very effective at holding off the torrent of mixed quality information that invades our lives. Given the right training, we invest as much effort into vetting as we feel it deserves. Unfortunately, this places the onus on the author to approach us in a way that passes or slips past our vetting. A new discovery must convince an important peer reviewed journal like Nature to publish their discovery. A normal article in a lesser journal is just not important enough to slip past the vetting of most scientists. Have a new medicine for ulcers? Prove the technology and efficacy to the editors and peer-review boards of some of the most overwhelmed journals first.

Somewhere in this onus of proof is a bridge to the field of marketing. Name brand recognition is very significant to the selling of objects. We usually do not make a detailed study into which laundry detergent to purchase. We simply don't care enough about our brand of toothpaste to undertake a detailed assessment.

So it is with the information we consume. We undertake our own quick and dirty assessment. We rely heavily on the vetting undertaken by others and we throw the onus of proving value onto the author and publisher.

Let us now turn our attention to the Internet. It must not have escaped your attention that the vetting of a journalist, a news editor or a fellow reader resembles the publishing of a link or stating an internet endorsement. When I write that **SearchEngineColossus.com** is the definitive source for regional search engines, I have vetted this site and found it praiseworthy. I considered and discarded other comparable resources. What is distinct about the internet is how short internet vetting chains tend to be and how brief the vetting.

Information delivered purely on the web, and not having originated somewhere else first, tends not to be vetted at all. The web publishing process is brilliant in the way it strips away the need to have anything we

do vetted by 'the powers that be'. This was part of the exciting liberation of information that occurred with the internet's arrival. An opportunity for truly unfettered communication. Ideas and concepts flow swiftly, directly from author to consumer. In the process, the onus of proof of value is stripped from the publishing process. The author can ignore the task of proving worth and just get on with writing.

Unfortunately, within this unfettered communication lies a lack of self-censorship; a lack of self-vetting. Consider the act of blogging: a veritable diarrhea of diary entries shoveled on the web for all to see. To a degree, quality standards fall furthest when there is no reason to improve information.

There are always counter-forces at work. As the internet environment develops, a different style of vetting imposes itself on the process of finding information. Instead of long vetting chains we use popularity ranking. In the ever-growing internet, the way we reach information becomes most telling.

In a sense, we are replacing vetting chains with search tool bias. As a vast abundance of information arrives on the internet, we turn to our peers for vetting and advice on what to read. In the absence of peer reviews, we use the clues offered by search engines that recognize prominence. Beyond this, we simply choose the first to catch our eye.

We simply cannot escape from the confines of our systems of communication. We can at best just push them aside for a time as happened during the emergence of the internet – a topic for Chapter Eight.

Remember, we must somehow reduce the quantity of information attacking us. Our time demands this. We vet, we focus or we invoke search tool bias.

Fortunately, even when information is deliciously unvetted, free from any kind of interference, we can still tell the quality of information from its context/format/source. Where was it published? How was it prepared? Who published it? What else is published nearby? The internet makes such questions profoundly easy to answer. We can invoke context/format /source to filter results.

Internet information is a microcosm of information in our larger society. Many of the facets and unique elements to the internet look far less impressive if viewed in this larger context. The change swept in by the emerging internet becomes less than truly revolutionary in this light. As

we will see later, it is less a marvel of construction and more a gradual evolution from times past.

It helps to visualize this idea. Imagine for the moment that in addition to the earth we stand upon, in addition to the stratosphere (the air surrounding the earth), we have another sphere that surrounds us. This is the information sphere. Don't confuse this with cyberspace, a reference only to internet information. Consider it instead as a world or realm of information.

Like the stratosphere, we are embedded in it. We live and breathe this medium constantly. It is evident in everything that informs us and sells us our opinions.

There is a history to this information sphere. Two centuries ago, it was a history born of books, sermons and newspapers. Then it became a medium of books, articles and newspapers. Today this medium is far more vibrant and chaotic: multi-ethnic, multi-lingual yet fundamentally similar.

This medium includes the CNN we watch over breakfast, the political discussion we have with our neighbor, the magazine we read during the coffee break and the websites we view in the evening. All this joins together as the information realm. The internet represents just part of this medium.

What is the essence of this medium? It is the message. The fundamental particle is the message informing us of some point of view. This message is prepared in a particular format: a book, journal article, press release, news article or joke. This message has an author, with bias, perspective and values, as well as a publisher, responsible for editing, quality control and promotion. This message appears in the context of other messages presenting similar perspectives and similar bias.

Here we have a holistic image of a vibrant medium encompassing many formats whereby a message is crafted, communicated and understood. The internet cuts across many of these formats to offer an alternative digital path of communication. No, the internet is not a format in itself – that unhelpful distinction leads us astray. Instead, the internet participates in the communication of books, articles, press releases and so on. The internet never was the death of the book, much less the death of television as predicted in its early genesis. These formats co-exist and cross-fertilize one another.

When it comes to research, please retreat from the myopic view that we are searching a computer filled with facts. For starters, we are dealing with messages, not facts. Facts are somehow truthful, factual and beyond alternative interpretation. I left that idea behind in high school and in Chapter Three. I welcome instead the notion that bias and values are embedded in most statements; perhaps every statement.

Now that we have an image of an information realm, notice this realm includes a competition of ideas. The people behind the messages would prefer we read and absorb their messages and not others. Some messages communicate more effectively, more pervasively, than others. The first three messages we encounter may not be the messages we want to encounter. They may not be the best messages for us to base our decision upon. Looking at the internet as some kind of giant computer filled with facts misses this distinction. It would lead us to consider the first three items on a search engine results page as perhaps the most relevant. And since this is rarely so, this misunderstanding will lead us to discount the value of the internet as an information resource.

The early metaphor for the internet was a library where all the books are tossed in a pile and the lights turned off. However, with the internet full of messages, we have not disorganized information but competing information. The order does not appear as a surface feature but instead is embedded deeper in notions of context, format and source. The pile of unsorted, unclassified information resolves itself into a mix of systems of communication – each centered on a different format, each system highly organized with its own strict rules and social mores. Yes, collectively this gives us a feeling of vertigo. The internet slices across many different systems of communication so we see unrelated information placed side-by-side. Instead of books beside books, works of comparable authors group together, works on similar topics group together.

It is as if we waltz into a pubic library, approach the omniscient card catalogue then search for a particular chapter from a book whose title we have forgotten. Card catalogues index books, not chapters. What to do?

This image is not of chaos. To name such a library as chaotic is to first assume our search for a chapter by name was simple, then to give up, throw our hands in the air and cry out with blackened frustration.

All is not lost. Consider context/format/source. What author would write the chapter we seek? Will it be in a book or magazine? What is the

context? The subject? Even without a book title for our chapter, we should be able to find our way.

We may find ourselves scanning a list of potential books for something promising. We may find ourselves thinking about what we hope to find. Both these approaches work splendidly on the internet. Both stem from thinking in terms of context/format/source. As we proceed, further approaches will stem from an intimacy with further influences in play upon internet information.

I must pause this discussion here for a time. I must be careful lest I continue commenting but fail to reveal any search tactic of practical use. So let us now turn our attention to searching quickly, moving more swiftly and shortening our journey. Later, we will take up this discussion of the nature of information once more and, I trust, the search approaches that arise will seem accessible instead of impractical.

Chapter Five

_____

HASTE

The young rebel strode forward, plumed and confident. He was much too young for such a task, thought Albert, as he furiously sought a solution to this accidental disaster. Albert's small band of soldiers stood nervously to one side of the forest clearing while around him emerged evidence of a much larger band of troublemakers. Unseen voices spoke short words. Accented words. Aragonese words. These were not peasants up to mischief as he had been told. They were a band of freebooters, of highwaymen who had journeyed north over the Pyranees Mountains from their homes in Aragon and would return there come winter if not before. They would be well armed. Their Aragon masters may even have sent them.

The young boy approached to negotiate god knows what. They must feel confident indeed if they felt willing to negotiate. Archers, obviously. And as they had found him, they knew his numbers. How many surrounded him? What to do?

The kingdom of Aragon was not often a source of such trouble and this group had not long arrived. Albert had been sent at the first mention of bandits. Now, with just eight men under his command, so much was unknown. Would they be dead in an hour?

His men had weapons and discipline. His men? Albert sighed. He had yet to earn their respect. Well, they would follow his lead now. If only he could bend this situation to his benefit. Find the high ground in this forest of uncertainty. Turn this blade before it could be thrust....

<center>*   *   *</center>

As we race around the internet, no longer are we presented with information of unknown quality from unknown sources. We encounter

information supported by a list of similar documents by the same author, accompanied by comments and supportive gestures from other people and institutions, embedded in a collection of comparable resources. Internet information has a halo of supportive detail.

Let us gather this information quickly though. It is nice to say supportive details are available; so very much better to gather this information at the click of a button. This chapter is largely an information dump. Strategy comes later. So to start with a bang, here is the bookmarklet to retrieve context at a single click.

THE CONTEXT BOOKMARKLET

For Google:

```
javascript:q=location.href.substring(0,location.href.substring(0,
location.href.length-1).lastIndexOf('/')+1);if(q=='http://')
q=location.href;q=q.replace('http://','inurl:'); void(location.href=
'http://google.com/search?hl=en&num=40&filter=0&q='+q)
```

For Yahoo:

```
javascript:q=location.href.substring(0,location.href.substring(0,
location.href.length-1).lastIndexOf('/')+1);if(q=='http://')
q=location.href;q=q.replace('http://','inurl:'); void(location.href=
'http://search.yahoo.com/search?p='+q)
```

Frightening? Do not look too closely at this script since you probably will not choose to alter it. Allow me to describe it in general terms. A bookmark is a link kept in a convenient place. A bookmarklet (notice the 'let' suffix) is almost the same but instead of linking, it causes an action to occur. When we click a bookmarklet, we run a javascript like the one above. One of the more famous bookmarklets is the Google Browser Button, a bookmarklet that searches Google for the words we highlight with our cursor. Because of this, you may already know of bookmarklets as 'browser buttons' instead.

The two javascripts just above permit us to retrieve, at a single click, a list of additional publications found within the same directory as the page we are currently viewing. We covered the use of local context in Chapter Three in our discussion of quality assessment. Add this button to our web

browser's Links toolbar or Favorites list but do not bother typing it in – visit **SpireProject.com/buttons.htm** and install it from there.
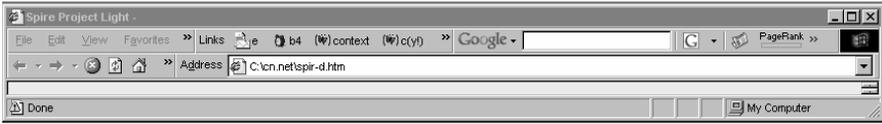
-- Screenshot of Microsoft Internet Explorer featuring my Context bookmarklet placed on the links toolbar, followed by the Google Toolbar. Reprinted with permission from the Microsoft Corporation and in line with the instructive use of Google Brand features.

Yes, javascripts pose a security risk but you can easily see these scripts are harmless. They counts the number of '/'s in our current web address, subtracts one, then ask Google or Yahoo for pages that share the address. Installed as a simple button at the top of my web browser, I often click it as I browse the web.

Bookmarklets are quite flexible. There is a bookmarklet for loading two webpages side by side using frames, another for changing the resolution of an image displayed in the web browser and another for requesting that any click on a link opens as a new window behind our current window. This last one is handy since we can click, click, click and all three pages open as separate windows in the background. There are many book-marklets. I urge you to explore this topic further for it lets us accomplish certain tasks very quickly.

I have three bookmarklets I use most frequently. The context book-marklet mentioned above, a bookmarklet I call B4 that checks **archive.org** for past copies of the webpage I am viewing and a bookmarklet to retrieve in-bound links. I also include here the bookmarklet from Chapter Four to retrieve endorsements. All these bookmarklets can be installed from **SpireProject.com/buttons.htm**

The B4 Bookmarklet:
javascript:void(location.href='http://web.archive.org/web/*/'+location.href)

The Links Bookmarklet:
  javascript:location='http://google.com/search?q=link:'+escape(location)

The Endorsements Bookmarklet:
javascript:q=location.href;q=q.replace('http://','');location='http://google
.com/search?q=link:'+q+'%20OR%20"'+q+'"'

Bookmarklets are flexible too. We can easily remake a bookmarklet to point towards a different search engine. We could adapt the Google Browser Button, for instance, into a bookmarklet that asks the Library of Congress Online Catalogue (LOCOC) or Amazon.com for their record of any book title or ISBN we highlight. We could, for instance, adapt the B4 bookmarklet to check if a website appears in the ODP directory. Plenty of routine can be simplified with a bookmarklet. However, unless you wish to learn javascripting, I recommend only searching for and making adjustments to existing bookmarklets. It can get quite complicated.

May I stress that the bookmarklet is a simple way to move more swiftly through the internet. Consider the B4 Bookmarklet above. Rather than retrieve the web page for the Internet Archive, type or paste in an address then press return, I have an alternative. I click a single button on my web browser's Links toolbar and the task is complete. It is all so very simple. Retrieving history or context or endorsements becomes a gesture rather than a task. Like waving our hand to catch the eye of a waiter – we do not need to make the coffee ourselves.

What we cannot do with a bookmarklet, we may be able to accomplish with the next technology: embedded forms.

WORKING WITH FORMS

A form is a collection of HTML tags that allow us to send information from one page to another. They are everywhere on the internet. At its simplest, a form is the text box on a webpage. It is the textbox on Google.com we use to search their database. It is the textbox on our local library webpage that connects us to their card catalogue. It is the form on Amazon.com that connects us to their book database.

HTML, Hypertext Markup Language, is the coding language that tells a web browser how to display a webpage. All forms are made from just seven HTML tags. Forms start with <form action=address>, end with </form> and forms cannot overlap.

Forms can be complex. Google's advanced search page contains a complicated form with numerous elements, many of these elements

hidden from view meaning not displayed on the webpage but visible in the HTML.

What is not commonly known is that while connecting a database to the internet is a technical feat that challenges anyone lacking a degree in computer science, moving a form is simplicity itself. We can pick up, move and indeed alter any form we find on the internet. We can take it from one page and move it to another as we desire. We can easily shift the form for Google's search box and place it on our homepage within the context of our choosing. I call this 'embedded forms'.

As part of the Spire Project, seven years ago I embedded over 130 forms within the supporting text that explained how and when to use certain search tools. Articles on patent research, government publications and searching discussion lists stacked embedded forms one over another within a discussion of what an ideal search involves. Search the various significant databases in turn, all from within the one article. Simply start at the top of the article and gradually work down the page.

It was a good innovation because a better search often needs only a few choice words of when to use a certain database, of what punctuation it accepts and what to use next. My difficulty was the mammoth unpaid task of keeping such advice current.

Using forms in this manner shaves steps from a search. No longer must we visit a webpage, then send a request, travel to another webpage to access a different database, then travel on to a third. Now we do all of this from one location. Furthermore, to help others search effectively, simply share our search page. They too can work their way down a webpage, following our steps.

The technicalities of shifting a form are simple but not intended for this book. I wrote the instructions as an article in the ONLINE magazine by Information Today[33] and since I retain copyright, I am happy to provide a copy of the text at **SpireProject.com/art20.htm**. I will, however, draw your attention to two enhancements to embedded forms that demonstrate just how flexibly we can remake the internet to our advantage.

Firstly, and typical to all computing situations, it is rare indeed when we cannot find someone who has solved a problem for us. Remember I mentioned how forms cannot overlap? This is a slight overstatement since many years ago I found and adapted an early javascript by Adolfo Quevedo that folded multiple forms together into one.

The latest version of my merged form is called the Unified Search Plus and resides at **SpireProject.com/plus.htm**. It is free and helpful. Place it anywhere your like.

Here is a picture of my first unified form as published in 2000:

As an aside, the creation of this form led to the discovery of the hidden Google inurl field search, a find that later led to the rediscovery of context in internet quality assessment. Internet skills emerge slowly over years of study. Piece by piece, the picture comes into view.

ALTERING FORMS

A second adaptation of form technology is a rather amusing reversal of database control from database owner to visitor. When a database is placed online, many standard values are set by the programmer. Values like the number of matches to be returned by a search engine are coded into the HTML or left at a default value. If we can learn or guess the names of these variables, we can change their values ourselves. We can present the information as we want it.

As our example, say our search of Google generates an address like this:

http://www.google.com/search?q=bookmarklets

The form that creates this address uses just one variable, 'q', and it equals 'bookmarklets'. See the address? It is just there at the end: **q=bookmarklets**. Now say we add a value to an unused variable called 'num' that controls the number of matches Google returns.

http://www.google.com/search?q=bookmarklets&num=40

Tack **&num=40** onto the end of our first address and Google returns forty matches instead of the usual 20. We can accomplish this by changing the address directly as we just did or by adding a short code to the form when we move it. If we place the tag: **<INPUT type=hidden name="num" value="40">** between **<form ... >** and **</form>**, then Google returns forty matches. Yes, it is that simple. We just added a hidden variable and value.

Yes, we can adapt a database to our purpose if we know its fields. We can hack a database!

To learn a database's fields – the names to their many variables – read the HTML of the most complex advanced search page available. Many variables will be hidden from view but clearly visible within the HTML. Google's advanced search page invites us to view more webpages. Within the HTML of this page, it clearly labels that variable as **num**. At other times, just guess. Google has the allinurl field and an allintitle field. In a moment of blinding inspiration in about 1999, I tried and found inurl.

In the early days of the internet, some foolish online retailers would set price as a hidden field. It was there on the page, in the form used to purchase the item, hidden from view. Crafty internet users could move and rewrite this form to have the hidden variable equal anything they wished. Yes, purchase that golf bag at $2.29. Who says internet skills don't pay?

Hacking a database, however, is not about discounts. It is about who adapts the database to suit whose purpose. In 1998, I adapted the form to the vast Library of Congress Online Catalogue (LOCOC) into a perfectly sensible search just of their periodicals collection. Viola! A free search for magazines and journals by title and publisher – a search that did not require seven steps to accomplish. It was just the search I needed to round out an article on searching periodicals and was an entirely unintended use of the LOCOC database. There was also no comparable alternative. On occasions like these, I simplify overly complex forms and apply standard values to choices I know I will make. Do not ask if I want to search by author/title or subject. I will set my choice as a pre-selected hidden variable and do away with that question.

The earliest use of embedded forms I have seen adapted the form for the PubMed database (a prominent database of medical research by the US National Library of Medicine) so as to fix our search words.[34] This creates a

rudimentary current awareness search. Click a button each month and look for any new articles with our pre-selected keywords.

If we add a little scripting and perhaps a spider we can create even more fantastic tools. I have a footnote script prepared in Perl (a popular programming language) at **SpireProject.com/spir.htm** that grabs a page off the internet and returns it with link addresses listed as footnotes. For years Chinese residents used this tool to get around some censorship issues. Reaching further, we can easily convert a 'get' form into a 'post' form. With a spider we can also get around the personal identification requirements of sites like LOCOC where we must log in to be allocated a personal ID number before we proceed with a search. With a script and a spider, we do not have to play by these rules. Type and click. Let the script do the rest. Internet technology is wonderfully adaptable in this way.

Gary Price and his pioneering work on the invisible web[35] identified a great many prominent and important resources that remain beyond the reach of global search engines. The use of embedded forms permit us to access and play with such information as we like.

PREPARE OUR WORKSPACE IN ADVANCE

Besides the use of bookmarklets and embedded forms, consider these four further suggestions. Firstly, have someplace to copy-then-paste information and addresses as we search. I use a simple text file that I keep in my Microsoft Windows Start Menu. Into this, I paste key paragraphs, the addresses I may need later and the keywords I find most useful. When I complete a search, I have not just a record of what I found but also a record of where and how I found it. Alternatively, many sophisticated programs will archive sites as we traverse the internet. Whatever solution you prefer, prepare it in advance.

Secondly, shortcut keys help us accomplish tasks on our computer more swiftly. Here is a short list for the Microsoft Windows operating system. With practice, each becomes a simple roll of the wrist. If we wish, we can also define new shortcut keys to specific tasks. I have 'paste as text' assigned to Alt+V on my system because I use an older version of Microsoft Word.

Starting list of common shortcut keys.

|	
Cntl+C	**Copy**
Cntl+V	**Paste**
Cntl+Z	**Undo**
Cntl+Y	**Redo**
Cntl+F	**Find on a page**
Alt+Spacebar+N	**Minimize the current window**

One further shortcut deserves our special attention: the hard refresh. When information we request arrives incomplete, it often lodges in a range of computer caches along the route. Caches are a technical improvement that speeds the transmission of information through the internet by keeping copies of recently retrieved information nearby. A copy of the **IBM.com** webpage I just retrieved may be held on the computer of my internet service provider (my ISP) for a time. If another person from my ISP requests the same **IBM.com** webpage, the ISP may reach for the cache copy and deliver it instead of reaching across the internet for another copy from the IBM computers. Caches like these are scattered throughout the internet.

This system of caches works splendidly. It saves time and bandwidth. Unfortunately, if a page does not arrive completely when first requested or if a page has changed, we may not get a complete and current copy. To bypass these various cache copies, do a 'Hard Refresh'.

The Hard Refresh:

Depending on our web browser, we hold down shift or the control key when we click the refresh button. **Cntl+Refresh** works for Microsoft's Internet Explorer. **Shift+Control+Refresh** appears to work on all web browsers. There is some complexity to just which cache copies are bypassed but this will almost always retrieve a fresh copy from the source. I typically need to hard refresh when I stop a page downloading, then request it again but receive only half the page.

The third suggestion in preparing our workspace is to build a useful homepage and use it as a starting point for searching. Bring together many of the resources we frequently access or should access. Add in a link to a web

translation tool like Babel Fish found on **AltaVista.com**. Add a link to **SearchEngineColossus.com**, the definitive directory of regional search
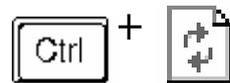
engines. Add a link to a directory of international government websites. Add whatever you like but create the homepage for the purpose of searching, not for internal business communication or as a nice screen saver. As an example, visit the page I search from, the page I jump to when I click my web browser's home button, **SpireProject.com/spir.htm.**

The fourth suggestion is to learn how to turn off advertising and turn off flash while we search. I find I do not need either. Intrusive advertising in particular can be greatly curtailed by making use of a HOSTS file on your computer. This simply stops you from drawing information from certain servers previously identified as serving advertisements. For several years now I have instructed my computer to ignore most of the banner advertisements that plague the internet. Read more on this topic from the Wikipedia (**en.wikipedia.org/wiki/Hosts_file**) and from Fravia's SearchLores site (**searchlores.org/antiadve.htm#hosts**).

CUTTING CORNERS

Moving swiftly is an admirable result. Bookmarklets, embedded forms, shortcut keys and a decent homepage will certainly assist us to shave steps from any search we habitually undertake. Cut corners from our path. Build escalators and elevators to move quickly along paths we routinely tread. However, search tools like these have another, far more significant influence. A good tool will lead us to make more frequent use of higher-level search techniques. The tool draws us into making better searches.

My homepage includes my form for searching different search engines and directories. Since I placed it there, years ago, I noticed I reach for the Open Directory Project more frequently. The tool makes it simpler to do what I should do anyway. Again, once I added the Context Bookmarklet to my web browser, I began to reach for context far more frequently. In each case, the tool led the way to a change in my search behavior. Even though we can rationalize I was saving only a simple click or three, that I could always have visited the Open Directory Project any time I wished, having that tool nearby appears to be decisively important in helping me actually consult the Open Directory Project. The tool leads the searcher. That a tool is efficient is almost an afterthought.

If we accept this reasoning, then a vast gulf exists between what we know we can gather on the internet and what we do gather. To bridge this gulf, we should create and position near us the tools that draw us to better search techniques. If we do not install a links button, we will not use endorsements or prominence in all but the most significant situations. If we do not have a language translation tool, like Babel Fish, nearby, we will shy away from searching pages in a foreign language in circumstances when we really should leave the English zone of the internet. Is it possible that so many internet citizens do simple searches only because they have so few of the tools installed that would push them to do better?

The most important lesson of this chapter is not one of the many rather transient ah-ha moments as I suggest a new tool or technique. The message is more perennial. Reshape the internet so we can easily gather the information we require. Use search tools to step from knowing how to search better to actually searching better.

This is one of three simple rules I have for searching with haste.

> Rule #1: Reshape the internet to help us find what we need.
> - Bookmarklets
> - Embedded Forms
> - Links and Bookmarks
> - Shortcut Keys
> - Hacking a web address
> - Homepage
>
> Rule #2: Never sit waiting for information.
> - Juggle windows
> - Get a faster connection
>
> Rule #3: Create a search style all our own.

Before we move on to rules two and three, I have placed a range of search tools on the Spire Project website and often adapt existing tools to suit specific needs. If you find or create something helpful, or need a little nudge, consider sharing the idea with me. Reach me via a form at the bottom of **SpireProject.com**

JUGGLING WINDOWS

For anyone frustrated that the internet is too slow, the solution is to juggle windows. Juggling alone will more than double a dialup connection and make delays in retrieving information from any connection of little significance to the speed of searching.

You may juggle already. It is fairly common to have several programs running at one time. When I ask at my seminars, three-quarters or more of my audience already work with multiple programs – leaping from one program to another at the flick of the wrist. Perhaps we run a word processor in the background while we surf. Periodically, we grab a para-graph or web address from our browser, jump over to our word processor and paste it into a gradually developing report.

On this occasion, I want us to develop a very specific habit of working with several windows to the internet at one time. We will first learn to juggle, then to juggle in different ways. For want of a better metaphor, we will juggle not just more balls in the air but juggle off the wall and floor as well.

Eventually we will use the internet in a way that continuously opens new copies of our web browser as we search and closes them as we leave. Instead of viewing the internet through one window, we may work with five windows open to the internet, constantly opening and closing windows as we search.

We do this for two reasons. Firstly, the internet connection plugged into the back of our computer usually idles, waiting for instructions. Information moves swiftly around the internet but it moves in spurts and flashes from the internet to our computer such that much of the time our computer sits waiting for our next instruction.

Let me explain. Type a web address into our web browser, then press return. A request for a webpage spins off across the internet. Pause. The distant computer sends back our requested webpage – just the HTML text document. Our computer reads this file and uncovers five pictures required to display this page. Five requests are made to the distant computer, one for each image. Pause. Our software is set to request just a few images at a time, perhaps four, so the fifth image probably waits until the first completely arrives. Another pause. Finally, the page is assembled on our computer, images and all. We read this page. Big Pause. We click a link and the process starts again.

Notice the many pauses? Much of the time, nothing is happening. The computer waits for information to arrive. The computer waits while we read. We can see this waiting in numbers if we watch as bits come and go from our computer. Working with several windows fills these delays simply because we keep adding requests to the queue. The computer always has something to retrieve.

A second reason to adopt juggling involves the way we search. Novice searching tends to be logical and linear. We step forward, then step back to try another direction. We view the internet as we view a road map. Drive towards a destination. If a route is blocked, we re-plan our drive and approach from a different direction. Experienced searchers search in a different way – rolling forward in several directions at once – moving in a way we cannot move if we have just one window onto the internet.

HOW TO JUGGLE

Let us place some of the tools we will need on our desktop.

> 1_ **ShowDesktop** is a simple Microsoft Windows program that minimizes all the windows currently open. We find it down in our **/Windows/Systems/** directory (not **/windows/**). Create a short-cut for this small program. Bring it onto you desktop, then add it in your toolbar on the lower left corner. Set the toolbar to 'Always on Top' + 'Auto Hide'. (Right click your toolbar, select properties, then check 'Always on Top' and 'Auto Hide'). As an alternative, learn to use the keyboard shortcut, **Windows+D** that uses a small key with a window image found only on some keyboards.

> 2_ The **Alt+Tab** shortcut jumps to the next window we have open or minimized. When we have several windows open, hold down the Alt key, press the **Tab** key, then press **Tab** a second time. See how it cycles through the available windows? We can keep thumbing **Tab** to reach the program we want.

> 3_ Also of interest, the **Alt+F4** shortcut will close the current active window. I use both the **Alt+Tab** and **Alt+F4** shortcuts so

frequently that it has become a well-practiced gesture, a mere flick of my wrist.

4_ Lastly, we have four ways to open a new window. Select the web browser menu: File → New → New Window. The shortcut for this is Cntl+N. It opens a new window showing our current page. We clone our page. To open pages as we browse we can right click our mouse on a link, then select the menu items that reads, "Open Link in New Window". The shortcut for this, my personal favourite, is to hold down the shift button when we click a link. (Some web browsers like Firefox offer tabs, so the practice of juggling is a little different.)

List of additional shortcuts.

Close Window	**Alt+F4**
Jump Windows	**Alt+Tab (then Tab again)**
New Page/New Window	**Cntl+N**
Show Desktop	**Minimize all windows**
Shift+Click	**Open page in a new window**

Where are we going with this? We aim to liberate ourselves from looking at the internet one page at a time. We will search the web in many directions at once, rotating through several windows. We will open and close windows as we search. In this way, not only will we move faster since information constantly downloads in the background, not only will we not have to retrace our steps, we also begin to search in the way our mind works – in a scattered manner full of momentary inspirations and clever asides.

MOVING SWIFTLY IN PRACTICE

View a page with a large pdf file. Do we click the link and wait, and wait, and wait while the pdf file downloads? No. Hold down the shift key and open the file in another window. Minimize the new window just as it appears, then continue reading from our first webpage. As we read, the pdf will download in the background.

Faster? As we read, we encounter a link that interests us. Hold down the shift key and open the link in a new window. Now return to the first

window with the Alt+Tab shortcut and shift-click on another link that looks interesting. Keep going until we reach the end of the page, then close our current active window. By now, we should have several pages downloading in the background, the first almost certainly ready for us. Go there and start again.

Two advantages emerge. Firstly, this is a far more fluid way to search. None of that stopping and starting that takes our train of thought in different directions, constantly showing us new pages organized in different ways. Instead, we read the page we are on completely, then jump to the next page. Secondly, having several pages awaiting retrieval at any one time, we improve the speed we download from the internet – valuable if we do not have broadband access.

Faster still? Say we come upon an interesting page that refers to a word we do not understand. We should look it up in a dictionary, right? Type CTRL+N to open a new window. Press home to reach our homepage where we have a Google search form ready and waiting. Now search for a definition for our word. We can come back and read the answer later or just wait. Once we learn its meaning, close the window with the ALT+F4 shortcut and we are back where we first encountered the new word.
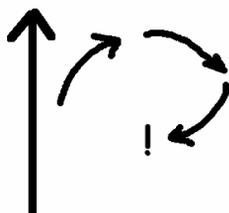
Say we visit a page and wonder if the author is an expert. A quick Cntl+N clones the page we are one. Next, click the context bookmarklet on our web browser (the bookmarklet we installed at the start of this chapter). This retrieves a list of further documents located nearby. Skim through this list. Judge its value. Now close the window. We continue our search better informed but essentially undisturbed.

If we draw this motion, it would be a little circle on the side of our search that does not disturb the general flow of our search.

I was buying Adobe Acrobat recently and I found an online store that sold a copy very cheaply at just US$79 – far less than the US$299 that Adobe itself charges. It seemed like a good buy. **eSmartsoftware.net,** found at **esmartsoft.cjb.net**, would get my business. I was about to buy when it

occurred to me I was not yet comfortable providing my credit card details to this unknown company. I wanted to know this company was legitimate. Perhaps I should check the prominence of their website. Have they been in business long?

Cntl+N clones my current page. I next click my Link Bookmarklet and retrieve a list of other websites linking to this page. Hmm, only internal links. Disturbing. This is not a popular, widely recognized retailer. I change my approach and search for endorsements. I search for sites that mention **eSmartsoftware.net**. I find little more than a page from a university law site describing the sale of 'legally grey' software, explaining that some online retailers sell software without a license on the assumption the purchaser previously purchased a license but lost their legal copy of the software. This is not for me. I need the normal version. Internet Informed, I close both windows and buy from a retailer I respect.

Can you see how this way of using multiple windows simplifies the task of gathering halo information? With only one internet browser window, we stare at the internet through a single looking glass. Any side trip involves moving our attention away from the page at hand, then retracing our steps later. This makes for a very staccato, shaken search style. With juggling, checking the history of a website behaves more like how we think. For a simple aside, we simply open a new window, retrieve some details, close, then continue as before, undisturbed.

A SEARCH STYLE OF OUR OWN

Bookmarklets, embedded forms, shortcut keys and a homepage full of helpful search tools and links to useful websites – all this comes together to create a search style all our own. If we do this right, we cut many of the corners from many of the routine tasks we undertake online. We simplify our search. We develop a fluid process that mirrors how we think.

I have my own style, a style built for searching a diverse range of topics in a wide range of depths. The habits I have formed, the flick-of-the-wrist familiarity with certain shortcut keys, my bookmarklets, bookmarks, links and homepage, all serve to smooth out my search style. This style is fluid, continuous and unlikely to suit you.

There is a style for you. Find it. Find it and searching becomes more relaxed, more enjoyable and faster. Instead of taking five steps to accom-

plish something, a search becomes a blur of finger movements and jumping windows. Several steps dissolve into one action we can do in our sleep. At this point, our focus moves from doing five distinct steps to searching as a series of actions – of grouped steps that lead us to the information we seek.

Said another way, a good search style liberates us from thinking about which links need clicking and which buttons need pressing. Instead, we think at a higher level about where we want to go and what we want to see.

As I login to my webmail service, I remember that I wanted to look for a list of technical bookstores in Australia. I clone my present page, click my homepage button and with the cursor waiting in the search box, I type: **technical bookstores inurl:.au**. I press return, flip back to my webmail page (**Alt+Tab**) and open four emails in quick succession so they load in the background (**Shift+Click**). I flip back to my Google search and start refining the search into something useful. My mind hums with activity, not delay.

Another occasion and another search style: I chase the contact details for astronomer Forrest Hamilton. I have eight webpages open in a spray of divergent searches. With one window, I chase an address on a website I am told he manages as webmaster. On another window, I try to find his contact details near his photo where it mentions he is a member of The Hubble Heritage team. From another window, I seek his details with a global search engine, launching several pages, closing each in turn as they prove unsuccessful. Finally, I read yet another reference to the Space Telescope Science Institute and it occurs to me he may be on their staff directory. Two minutes later, I copy-paste Forrest Hamilton's contact details into my leads text document, close everything, then write Forrest an email.

Before our enthusiasm reins supreme, there are three negatives about this style of searching. Firstly, having many windows run at the one time may have the unfortunate side effect of causing a computer to crash more frequently. Make sure you have a pop-up stopper installed on your computer. I use the Google Toolbar that includes a good one. The internet software archive Tucows (**tucows.com**) has plenty of alternatives. This is an absolute must not because we juggle but because pop-ups trouble all internet users.

Secondly, Microsoft's Internet Explorer has a difficulty with memory, I am told. Every time a program is run, then closed down, Microsoft Windows loses the use of a little bit of the computer memory – not much but gradually enough to slow my computer. Every hour or two, I close down my Internet Explorer, then launch it again.

Lastly, we need to remember to keep shutting down windows as we search or this becomes an exercise in desktop clutter. This is why I have my text file open to record interesting but divergent search directions.

Why bother? I mean, memorizing shortcuts, installing bookmarklets and a new homepage for searching, running the risk of more crashes.... Why? Because it works! You search much more effectively in this way and the experience is far more enjoyable. You should see me do this one day.

Searching this way is fluid and graceful – adjectives usually not applied to computing. This style of searching looks like an amoeba moving through pond water. We flow forward along many different directions at once. We occasionally taste the supporting halo of information nearby. We reach a place of interest, then push forward from there, again in many directions at once. Amoeboid searching is very different from a search conducted with just one window, a search wandering from page to page, backtracking to past pages before reaching forward again. We may visit the same pages as the amoeba but visually we see a single thread, a string. Such a path does not foster a search strategy. It often fosters a kind of mental turmoil where the overall picture, the elevated vista, is the very last thing we think about.

Did I mention amoeboid searching is intensely fun? Information piles on us as in a rugby game: fast and aggressively informative. We do not derail our train of thought. We stay on path. We search with purpose. Just what purpose will begin to unfold in the next few chapters.

Chapter Six

STRUCTURE

R eturning from Aragon, having first talked his way into accompanying the highwaymen as they returned, then having discussed a trustless pause in hostilities, Albert found his reputation greatly enhanced. Invited to court in celebration, he chanced to dance with a most enchanting lady in waiting.

Enwrapped in the newness of court, and women in general, Albert's feelings flowered most fragrantly. He yearned. He adored. He sought salvation in a romance that could never be. Indeed, any appearance of physical intimacy would greatly harm the lady's reputation. Courtship was only that – courtship for the sake of courtship. This slow circling, these expressions of adoration, served only to celebrate the soul – the lady's most certainly but ultimately the man's as well.

Albert's initial devotions were clumsy indeed; of only fleeting interest to the lady he sought. His second love, the petite daughter of a wealthy Basque cloth merchant, responded in a way that drew from Albert a hidden reserve of passion. With deft attention and distraction, she helped Albert reach deep within his soul.

Months of ecstatic, tender attentions amid the crafting of much awful poetry ceased abruptly when her family married her to a wealthy suitor and she moved away.

Such painful parting. Such unrequited love. Her last, most beautiful gift.

*   *   *

At the scene of a messy train wreck, two investigators search for clues. The first searches for that one clue that reveals the way forward; that critical piece of evidence that explains why the crash occurred. A beer bottle beside the driver's seat. The twisted stretch of track. The rotten

wood of a collapsed railroad tie. Wandering among the rubble and broken freight cars, our first investigator casts a sharp eye about for possible clues.

Our second investigator looks instead for the best vantage point; that one critical and distant position that reveals everything that happened and points to the place where evidence must exist. The train jumped the tracks here. It slammed into the other train there. We cannot see the crash clearly from this mound here on the right, so let us move to that hill on the left and look again.

Both approaches have merit. That beer bottle will not be found from atop that hill on the right. That first point where the train jumps the tracks cannot be found standing amidst the wreckage. Both approaches have merit and we practice both approaches when we search the internet. We know how to search in a precise manner using the five critical tactics of "", -, OR, inurl: and link:. We know something of the distant overview and its reliance on context/format/source.

Structure belongs to that second investigator standing atop the hill. Grand structures are best seen from a distance. Certain information is already organized for our benefit and does not require us to descend into the mass of train wreckage to sort by hand.

The principal internet structures we will discuss are:

- government hierarchies,
- geography,
- associations,
- directories and nexus points,
- commercial-quality databases
- and the use of a thesaurus.

We will also address three structures found within a website:

- website search functions,
- staff directories
- and website hierarchies.

Let us move through these structures one by one. Eventually, we will hold these structures in our mind awaiting a time when, like a grappling hook, we can throw a line and attach our question to one of these existing structures, then climb up to our destination.

GOVERNMENT HIERARCHIES

Every state and national government website prominently displays a list of government departments and agencies. You will find this list linked directly to the state or country homepage. If we can reach this agency list, we can easily find the website for the Department of Education in Western Australia, South Australia, Alaska and Argentina.

Furthermore, we can locate state and national government websites easily with a directory like Foreign Government Resources on the Web (**http://www.lib.umich.edu/govdocs/foreign.html**) from the University of Michigan Documents Center. Yes, thanks to ranking technology we can also find the state website for Alaska by simply tossing **Alaska** at Google or Yahoo. **Alaska inurl:.gov** works with more certainty. Knowing something of states and web addresses, the address is probably just **Alaska.gov.**

This means that if our boss arrives rather excitedly one morning raving about the childcare reform package recently released by the state of Virginia, and asking for a copy just as soon as a cup of coffee can be found, we know what to do.

What do we do? We have three ways to proceed.

> 1_ We search a global search engine like Google or Yahoo for **virginia childcare reform** in the hopes that the search engine has already indexed it and found it worthy of visibility.

> 2_ We search a government-only search engine like FirstGov (**FirstGov.gov**) or the Google US Government Search (known as UncleSam at **Google.com/ig/usgov**). Again hopefully they indexed the page and consider it prominent.

> 3_ We find the website for the State of Virginia, move to its agency list, select the agency responsible for childcare, then look on that agency website for the reform package we know must be there.

This third approach works every time. Simply follow the thread through the government hierarchy to where the report resides. Instead of searching all across the vast internet, let us search only the website of the appropriate agency. The key is in recognizing that the any childcare reform package MUST sit upon, or the very least be clearly referenced by, the appropriate government agency website. The train MUST have jumped the tracks here and slammed into the oncoming train there.

Using this approach, we could quickly gather a hundred childcare reform packages for review by simply visiting the appropriate agencies for a hundred states and countries from around the world. Most significantly, such a search would be comprehensive. It would be complete. Unlike a search with a search engine, we will not miss a report because we search for **child care** instead of **childcare** or because the report was released two hours ago or because the report is forgotten or considered lacking in prominence or has an unusual title or is buried under a hundred thousand other related reports. Even if the report is not present, we know who to ask. Call the agency on the phone. They must know where it is. This is the best way to find a government report.

There is the temptation to search for government material as we might search for a book. Find any book on how to play the piano, then look on the shelves nearby for the book we want. On the internet we might throw some basic keywords at a search engine to find a regional government document on childcare. This should tell us the agency involved and land us on their website. Thus, search for **childcare legislation western australia** and hope this leads us to something on the website of the Department of Community Development, Western Australia.

Unfortunately, this rarely works smoothly. We nearly always must scan long lists of possibilities and we have no certainty we will find the right agency. This can then put us off using the government hierarchy all together. For a better approach, work like the researcher who simply reaches for a published directory of government agencies. Use the list of agencies found on the state government website.

The other trap is like trying to search a library for books by illustrator Stephen Cartwright. As illustrator, Stephen works with various authors. His books are not shelved under his own name. They are scattered all over the library. The government hierarchy works when we know, or guess, the information will be found on a government website. If an agency is only peripherally involved, if they are only an interested bystander, reaching for a government agency website will not help us.

I pose the example of the childcare package in Virginia in many of my seminars and very few attendees immediately reach for the state government agency list. As with other structures on the internet, we must learn to think where the information most likely resides. Will any structure like government hierarchy or geography lead us part of the way?

GEOGRAPHY

Many questions have a geographical dimension: questions that pertain to a place, an area, a region or a country. At other times, we should ask questions with a geographical limit. Different communities will overcome a particular problem in different ways and at different times. How does Argentina deals with child care reform? How does Venice deal with having too many pigeons in Piazza St Marco? (Answer: contraceptives in tourist-bought pigeon food.[36]) As soon as we can link a question to a specific region, a new set of very specific tools leap to offer assistance.

We have about five regional resources to consider:

1_ regional search engines,
2_ Babel Fish language translation,
3_ regional English language newspapers,
4_ local newspapers
5_ and global search engines restricted to regions by using **inurl:.au, site:au** or restricted to a certain language.

SearchEngineColossus.com is the definitive directory for regional and national search engines and directories. It is very comprehensive. Want a local search engine for Spain, Sri Lanka or Singapore? Come here. Hunting for a European take on childcare? Reach for an all-European search engine.

Regional search engines were once thought certain to blossom into big business and thought certain to offer a clearly superior search. Sadly, the whole regional approach never really took off. Many of the regional search engines are not vast nor quick and regional searching never became a common habit, though I hope it will some day.

Many a regional search engine will, of course, lead us to a webpage in a different language. The solution is Babel Fish, named after a most amusing creature found in The Hitchhiker's Guide to the Galaxy, a book by humour writer Douglas Adams.

I have an example I enjoy delivering in seminars that rather perfectly captures the value of a regional approach. During the Jubilee Year in 2000, I took my wife to the Vatican. While planning the trip, I wondered if we could attend a nice opera. I naturally turned to the internet for help.

A global search for **rome opera schedule** seemed to recommend only travel companies focused on taking east coast Americans on European

cultural trips, trips that included the occasional opera in Rome. A global search for **rome opera tickets** seemed to lead to only British companies selling tickets to events across Europe, events that included the occasional opera in Rome. No, none of these companies were linking to schedules of Roman operas.

We need to go to Rome. Our question has a regional dimension so we can use a regional search engine like Arianna (**arianna.libero.it**) to find Italian pages on Italian operas. After all, the Italians probably know best what operas are playing in Rome. To overcome the language barrier, I open a second window to my web browser and point it at AltaVista's Babel Fish and translate each page as I go. First, I translate the search engine's Italian answer to my search for **opera roma**. I can clearly see Opera di Roma on the list. Next, I translate the homepage for Opera di Roma and see I want to visit the Biglietteria, the ticket office.

The website of the Opera di Roma did not help me but searching further in this manner, first in Italian, then translated English, eventually led me to a Jubilee page[37] by a government agency listing all the events in Rome; operas, plays and concerts. I would not have found this list if I stayed with a global approach to searching.

Today there is an English site dedicated to events in Rome but this style of translation allows us to move beyond the English zone and reach information not yet prepared for us in English.

In addition to translating foreign language websites, we also have English-language newspapers to consider. Numerous English-language newspapers dot the globe and are available to read online. There are even English newspapers in Pakistan, Cuba and Korea. Just approach a directory of English newspapers like those at **DailyEarth.com#directory**. Years ago I made a tool for English-language papers at **SpireProject.com/spnews.htm**.

Using regional newspapers is a time-honoured approach in information research. Most cities have just one significant newspaper, so any issue of interest to a local audience will appear in this one clearly visible resource. In competitive research, we may scan a local paper for employment advertisements and search through a local newspaper's database of past articles for business or factory-related news. This will cost us and will require we find access to the archived news database but regional news of this nature is difficult to find another way. On a global topic, consider the New York Times or indeed a translated news database like World News

Connection, a US government database once freestanding but now accessed through Dialog.

Global internet search engines allow us to limit information by country code and by language. This is our fifth geographical tool. For countries, merely add **inurl:.[country code]** as in **pigeons inurl:.au** or **pigeons inurl:.it**. The alternative **site:au** works just as well. Language is selected from the advanced search page for each global search engine. This kind of simple limiting of results can be powerful in part because it takes so little time. Search for pigeon population control. Now show me just those in Italian. Keep in mind, however, the larger regional search engines probably offer better coverage than global search engines restricted to a specific country code.

ASSOCIATIONS

Unlike government agencies, associations have little over-arching structure. It can be difficult to find the association we need. We only have the certain knowledge that foundations and associations exist for every occupational group and weigh in on every issue of public concern. If we wish, we can generate a decent overview of a contentious issue simply by listening to the opinions of two or three associations somehow involved.

Indeed, after lengthy research, we often find the associations have a central role in explaining various perspectives. My exploration of the value of wind farms, for instance, centered upon the perspectives of two associations: a community group against wind farms and an association representing the British wind industry. In regards to teaching Intelligent Design in public schools, vocal perspectives emerged from the US National Center for Science Education and The Discovery Institute. These are two associations despite what their names may suggest.

It is not always clear which associations will offer the most help. While there are patchy internet directories of associations, they list associations, not those associations actively publishing internet material. Associations were generally late to internet publishing. Many publish very little on the internet for public attention. However, by virtue that their purpose matches so well with the opportunities the internet offers, we can look forward to a rise in association publishing and more effective promotion of what they do publish. Perhaps one day, the structure associations add

to public discourses will be a more effective tool. I usually avoid using this structure, then notice too late that the best information came from two antagonistic associations. I look back wistfully and wonder if I could have saved time by searching for associations first.

I am also mindful that government and political research often involves little more than listening to the perspectives of relevant associations. Much of lobbying is merely associations speaking their mind, presenting their perspectives directly to policy makers. In another setting, away from the internet, we might simply reach for a directory of associations and start talking.

DIRECTORIES AND NEXUS POINTS

The directory forms one of the most fundamental structures on the internet. The information we seek is already organized for our benefit; the scaffolding already erected for us to climb to our answer. It seems everything has its directory these days. Can we rephrase our question to search for a directory instead of searching for an answer directly?

If we seek a list of universities that teach information skills, we can:

- search a global search engine by keyword,
- start with a directory of universities, then hunt though each for an information school
- or search for a directory of information schools.

I know the internet overflows with directories so I search first for a directory of information schools and I find three suitable ones. My quest is almost complete. Furthermore, these directories make it very simple to repeat and extend my search at a later time. I also have a comprehensive answer, something I would not achieve with a global search engine.

Yahoo and the Open Directory Project (ODP) are the two largest global directories. Unfortunately, I often find these directories not specific enough for my questions. They can be relied upon to guide me towards university websites but perhaps not to information schools. I would not expect them to list Brazilian football clubs and yet I am certain lists of Brazilian football clubs exist. Yahoo and the ODP do not list international English-language newspapers but both clearly list sites that do. The ODP has a list of information schools but I doubt it is comprehensive. Compare

it to the World List by Professor Tom Wilson, a list both longstanding and famous. When it comes to finding that perfect directory, Yahoo and the ODP are probably not what we need but may well lead us there.

Paradoxically, the more time-consuming to build a specialist directory (yet possible), the more likely a specialist directory will exist. If several other people have searched for something difficult to find in the past – difficult as in effortful not difficult as in confusing – then we will probably find the fruits of one of these searches presented as a directory.

Searching for a directory is different to searching for the information they lead us to. Firstly, directories look different. They tend to have clear titles that say 'guide' or 'directory'.

Secondly, we can find a directory through the resources they link to. That is, we can triangulate their existence. Want a directory of cement producers, for instance? Just craft a search for pages that link to or mention two large cement producers. The list of matches will mostly include any directories.

Triangulation:
Link:#1 Link:#2 finds 3,4,5.
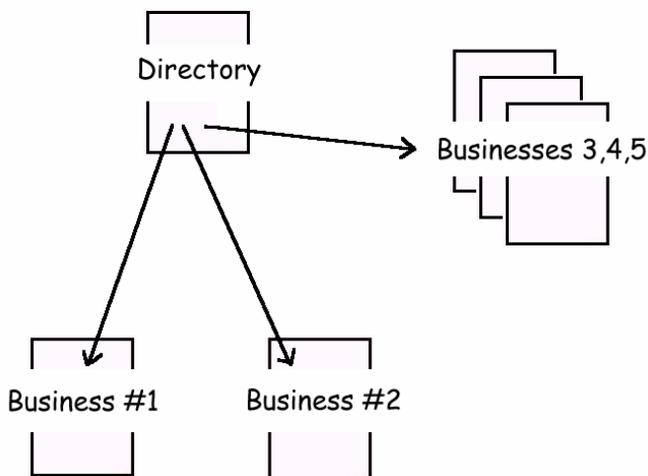
Better directories tend to have prominence too, so if only one cement producer springs to mind, we will probably stumble across a directory when we retrieve a list of pages linking to that one cement producer. Just on this topic, many directories only mention an address instead of link so allow for this in your searches by including the address in quotes as in: **link:address OR "address"**.

A directory is a comprehensive list of resources that match some criteria. This criteria and the experience of the person judging it will be significant to the value of a directory. At other times, as in a search for small meeting rooms and town halls, we may appreciate whatever we get. A directory listing may be the only information available online. Many small meeting rooms simply do not have websites.

To use this structure, ask ourselves if we can rephrase our questions into a request for a suitable directory. Many a search that at first glance looks like something we should throw at a search engine is in fact a situation where a hunt for an easy-to-find halfway point would suit us better. Directories are often these halfway points.

I use the term 'nexus points' to describe other halfway points that are not directories but still strive to bring order to the internet. The largest nexus points are sometimes called portals. These websites link to and describe resources of interest but not with the indexing and size of a commercial-quality database nor with the focus and comprehensive nature of a directory. Nexus points merely link to many resources on a particular topic and do so in more detail than usual.

The best nexus points have prominence, peer respect and good coverage. Such sites make no effort to index all the relevant material but do bring together a great many resources we need to accomplish something.

If we display the mass of internet information in a graphical manner, we would see certain sites radiate links to many of the most important resources out there. Such sites are not resources themselves – they will not answer our questions – but they bring together the resources that do. Consider these as the grown-up version of the humble guidebook.

Naming them nexus points does not represent a new way of searching – we stumble upon various efforts to organize the internet all the time. However, do we recognize them as structure? Do we ever search for them directly?

Nexus points attract searchers since they are great places to gather an overview of the type of resources available on a topic. They also assist publishers to build awareness for their information. Better nexus points act like magnets to people actively organizing the internet. They become collaborative efforts or at least incorporate ample volunteer participation. Such nexus points work as meeting points for researchers, publishers and organizers.

I remember back in the early 1990s, I was commissioned to organize relevant information for a housing authority. I searched and searched for information about housing and urban development but could not find more than a few items very separate from each other.

In my mind, I knew that somewhere there should be a nexus point of communication about housing information. Eventually I found it, in the HUDusers group website centered around one of the US Federal Agencies with an interest in housing and urban development. Once I found that resource, only then could I say I had completed my work. Many of the more prominent articles and resources were referenced in this website and I could reasonably hope new groundbreaking resources would appear there in time.

An individual can create a nexus point but will find it very hard to sustain the work. In earlier years, the success of a nexus point pivoted delicately on how actively it harnessed the support of members willing to share around the tasks involved or on any elusive financial support it could secure. Later, nexus points became a prized business model adding a layer of advertising and commercial marketing. This layer often destroyed the subtle synergy born of volunteered group effort but when successful, advertising provided the financial foundation to reach further.

Nexus points have a particular feel. They are fairly easy to recognize and distinguish from other kinds of web projects. They have far more links than most pages. They are very topic specific. They put thought into describing and categorizing links. They are also usually full of voices directing us to destinations that answer questions. The importance of the nexus point will emerge later in this book but for the moment, they are a destination, a halfway point we can use to help answer questions. They provide one more element of structure within the internet galaxy.

COMMERCIAL–QUALITY DATABASES

The most refined internet structure emerges in commercial-quality databases. Resplendent with multiple fields and detailed descriptors, these databases exist as islands of tightly organized resources amid an ocean of less-organized material. Some commercial-quality databases are free, databases like PubMed (medicine), ERIC (education), PatFT (US patents) and LOCOC (Library of Congress). Far more are not.

Gary Price and Chris Sherman published their book, The Invisible Web, in 2001 on this topic. Variously known as the invisible web, the hidden web and the deep web, these realms are dominated by focused databases often pivotal to a search. One of the central tasks of all searchers involves resource discovery – and we must keep a particular eye open for extensive significant databases.

The financial success of a commercial-quality database depends either on building a paying clientele or in securing government funding. This means that while the idea of creating a definitive pocket of organization appeals to many, the task is often expensive and fraught with financial risk. However, the best free databases are stellar achievers and do much to improve our access to information.

Because of financial challenges, usually only a few free commercial-quality databases will exist for each individual field and we can guess many of them once we have a little experience with databases. Of course, we will also encounter many databases too small, too patchy or not prepared to a commercial quality.

For paid commercial databases, there is a definitive global directory by Gale Publishing and separate directories for each of the large database retailers like Dialog and LexisNexis. Firms like these on-sell a range of databases much like a supermarket sells food. If it were not for the financial difficulties involved in presenting improved information on the internet, we would have many more of these islands of tight organization. As it stands, a delicate cost barrier separates the internet from most of these commercial databases. We will explore this dilemma further in Chapter Eight.

THE THESAURUS

Tagging recently ballooned in popularity with services like del.icio.us and Flickr capturing media attention. The first is a service for sharing bookmarks; the second for sharing photos. By attaching descriptive keywords to each resource, other users can more easily find something on a particular topic. We could ask, for instance, for a list of all the pictures described as volcanoes.

This idea of tagging is simply another version of a decade old habit of adding descriptors to articles or subject categories to books. However,

unlike the internet, library catalogues and especially commercial-quality databases put a great deal of consideration and effort into selecting the most useful keywords and descriptors, then using them in a standard way throughout the database. The subject headings of a library catalogue are used uniformly throughout the library. There are no synonyms or competing words. No overlap. The Dewey decimal system has just one number for each book. Such a refined use of indexing terms is definitely missing on the internet.

For a database, the definitive list of descriptors is called a 'thesaurus'. Yes, this is slightly different from the more common definition of thesaurus as a book listing similar or related words. The thesaurus of a database is a list of descriptors used to index material in that database. For example, MeSH is the definitive list of tags used by the PubMed and Medline databases, databases prepared by the US National Library of Medicine. MeSH is available online and can be very helpful in finding the right term to describe juvenile diabetes (Diabetes Mellitus, type 1). All articles on juvenile diabetes in PubMed can be found by searching for that term.

On the internet, however, articles on juvenile diabetes will use various terms. They may use childhood onset diabetes instead. A quick search on Google found:

"diabetes mellitus" "type 1"	= 2.4 million matches
"juvenile diabetes"	= 2.6 million matches
"childhood diabetes"	= 272 thousand matches
"early onset diabetes"	= 40 thousand matches
"childhood onset diabetes"	= 12.8 thousand matches

What variations are we missing? All these matches discuss the same condition but if we do not search under each term, we will miss something.

If we miss only more of the same, we could generally overlook less popular terms. Unfortunately, the terms we use to describe something tend to vary by field or discipline. A search for **juvenile diabetes** may lead us to more introductory, more popular resources. **diabetes mellitus type 1** may lead us to the more medically sound and research based resources.

I encountered this difficulty in some searching I did into staff loyalty programs. Business people tend to use a variety of terms including staff loyalty, employee relations and labour relations. However, searching with these terms reveals very few pages with anything to do with retaining

nursing staff – a big issue for hospitals. Apparently, within the medical and hospital environments, the terms I just mentioned are not in popular use. A look in MeSH reveals the preferred term of 'personnel loyalty'.

If we do not look at a thesaurus, we can easily and mistakenly assume we have little to learn from how hospitals retain nursing staff. We miss this because we do not search with their preferred term.

Global search engines offer to correct our spelling on search words and sometimes will suggest alternative words. A good generic thesaurus like the one built into Microsoft Word or a published thesaurus can help too. Indeed, merely keeping our eyes open while we read articles will often alert us to similar terms that could lead to further information. However, to do this right, we may need a specialized thesaurus like MeSH. This is, after all, the purpose of a database's thesaurus. In a sense, they offer us a structure that assists us to find the terms used in that discipline.

Finding the right terms can be very tricky. Many a search pivots on this point alone. Once we find the right term to search for, the technical term, the term used by professionals active in that field, we will find the information produced by those professionals. The thesaurus is one of the few tools and techniques available to reveal these critical terms. This is a most welcome structure and I would encourage you to bookmark a significant thesaurus that covers your field of expertise.

INTERNAL STRUCTURE

Internal to large websites we will always find a search function, a staff directory and some sort of directory structure intended to help us find information we need. These three structures can help us answer certain questions and should certainly not be overlooked as potential halfway points.

Astronomy Pictures of the Day (APOD) author Robert Nemiroff works for Michigan Technological University while his partner Jerry Bonnell works at NASA's Goddard Space Flight Centre. Both organizations have staff directories so we should have little difficulty finding an email address. No, we do not search for **Jerry Bonnell goddard space flight** and hope it reveals a page with his email address. Such searches are clumsy and take a long time. It is easier to reach for something specific we know exists than to find something nebulous that may not.

The structure of a large website often mirrors the internal structures of an organization. Directories follow the organizational tree. Information on a specific project will be found under the school, agency or institution working on it.

website
→ department
→ school
→ professor

or   website
→ department
→ sub-department
→ project
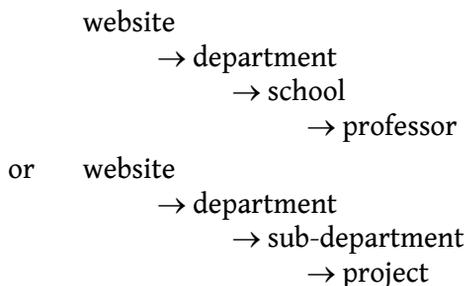
Alternatively, directories may be separated by target audience and function. Different directories address the needs of investors, customers, students & staff.

website
→ investors
→ institutional investors

or   website
→ investors
→ annual reports
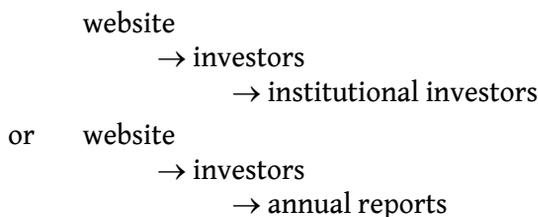
Websites structured as organizational trees are far more common. As we search, we can move up a directory to learn more about which part of the organization is involved. We can move down to a specific page by selecting the departments and individuals involved. If we know a specific department publishes a report, we can find our way there using the structure of the website. Yes, this is obvious but surprisingly easy to overlook.

THE INTERNET MESH

Let us chart some of the structures we have discussed up to this point in this book. This exercise has the advantage of being visual, memorable. The basic vision of the web looks like this:

The Basic Model:

Search Engines

Further Related Information ← Links — a page
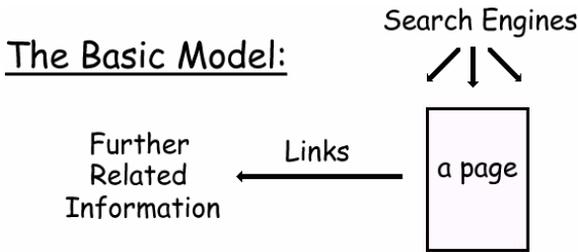
To this image we can add several further pieces. Firstly, the web works in both directions with both inbound and outbound links. Remember, we find inbound links with the link field search but these links are just one of three types of endorsements. Our new image looks like this:

Further Related Information ← Links — a page ← Inbound links — Referencing Information (link:address)

Other Endorsements (search for a name) (search for an address)
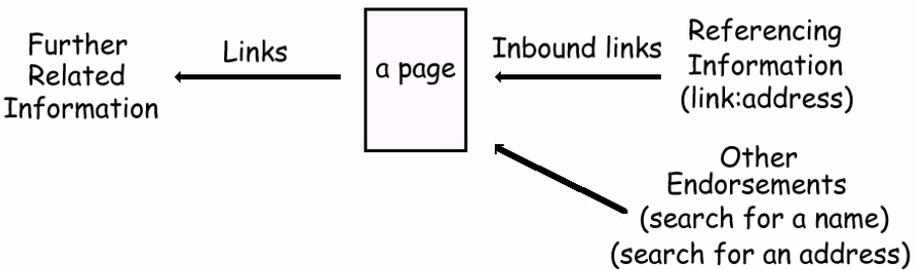
Now let us add information held nearby, on the same computer. Such information can be found with the URL field search or less easily by hacking the web address or surfing.

Further Related Information ← Links — a page ← Inbound links — Referencing Information (link:address)

Local Context (inurl:directory)

Other Endorsements (search for a name) (search for an address)
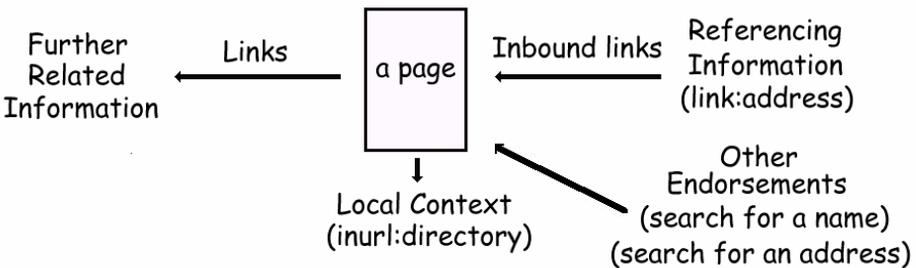
Now, let us add link companions, information by the same source archived elsewhere as well as information in the directory above.

The Internet Mesh:
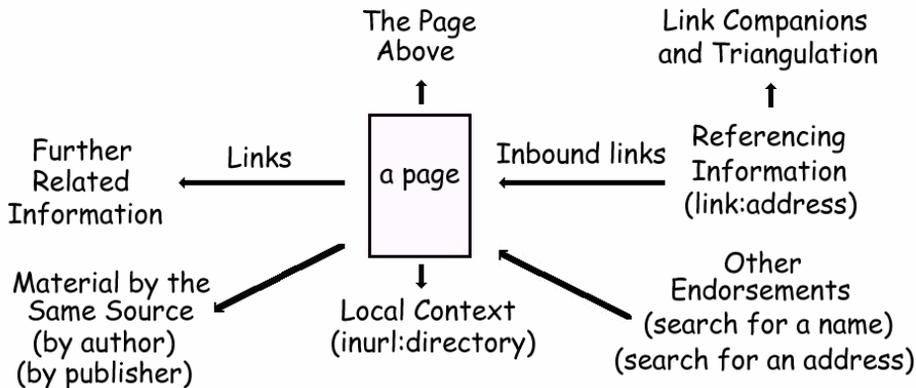
We can further extend this picture back in history by adding recent search engine cache copies of a website and adding the much older copies held in Archive.org. Remember a bookmarklet I introduced in Chapter Five? History is also just a click away.

So, here is the internet mesh. At the page level, far from caught in a web, a webpage is woven into a fabric of interconnections and relation-ships. Many of these connections help us with quality assessment and with finding related information. These are all connections. We should become adept at moving along any of these lines of connection. Furthermore, seeing the internet as a mesh frees us to approach information from a range of direction. Remember, until we break our reliance on search engines, we have not truly touched the heart of searching.

Suppose our small local theatre has a concert we wish to attend. When does it start? We can:

· search a global search engine by keyword,
· search a regional or local search engine,
· reach for a directory of plays and concerts here in town
· or visit the website of the local council that runs the theatre, then move to the part of their website that describes the theatre and its upcoming events.

See? The internet has structure. This is not a cloud we are working with. Clouds do not have strings or roads. Clouds have no scaffolding. Yes, we can filter the internet as if it were soup but we can also grasp these structures and use them to hoist ourselves close to our destination.

Our critical step is simply recognizing situations when government hierarchy, geography, critical databases or a thesaurus can assist us. We stay alert to these opportunities, a topic continued in the next chapter.

One trap is to consider structures like these as the essence of fine searching. Do we read this chapter, then decide the most significant division in internet search technique resides between using a directory and a search engine? Directories are just one more element of structure, order and organization to add to our growing collection. There is no fundamental difference between these structures and the keywords of Flickr or a webring on Narnia or a directory of library catalogues or the way format organizes the internet or the way source can direct us to places where information is likely to reside or how publisher motivation can guides us to likely publishers, a topic still to come. These are all elements of structure, order and organization.

From the start of this book we have painted an image of the internet as a galaxy. We have just seen several more pieces of our galaxy fall into place. This galaxy of ours has regions set aside for Afghanistan, for Seattle and for professionals discussing juvenile diabetes. Tight clusters of stars sit amid less structured space. Spokes like the spokes of a wheel radiate out through the internet connecting places that share a critical keyword descriptor – spokes that come together in a thesaurus. Visualize structure within the galaxy of ours: its galaxy arms, its dust lanes, its bubbles and spokes and swirly bits. A galaxy is much more than a mass of scattered stars slowly drifting round.

ATTENTION

O bserve everything. See detail where lesser mortals merely trample by. El Capitan's advice rang though his head as he mentally prepared for this challenge. Albert led his small party to the door of the Cistercian Abbaye de Fontfroide to gather aid. Left unresolved, he feared a simmering conflict of words would escalate. Minor incidents of interfaith violence would grow more serious. More frequent. Perhaps into armed conflict. He prayed for peace.

Through careful questioning, Albert learned the new emissary from his holiness Pope Innocent III inspired some of this violence. Meeting him would be a most delicate confrontation.

Barging into the long sixty-foot dining room, Albert strode purposely to the head table to confer with the abbot. Would the abbot please help persuade the visiting dignitary to act with more discretion? Help convince the papal representative to mind more carefully the potential of conflict between the Holy Roman Catholic Church and the emerging faiths of the Languedoc? Nothing could be gained by conflict. Much would be lost.

Few in the room held significance. Some present were not nice indeed. The church abbeys and their extensive estates had long been run as a separate nation within a nation where the abbot could pardon criminals, disregard taxes and certainly ignore the demands of a civil representative like himself. Albert stood on uncertain ground.

After a lengthy and animated discussion, the abbot agreed only that Albert indeed stood on uncertain ground. The abbot declined to help.

<center>*　　　*　　　*</center>

In a brutal time when concerns for community safety quickly overrode any sense of personal rights, Albert worked as a police detective unencumbered by rules of conduct. Perhaps fitting, then, that we should now delve into the detective's talent of observing the world and recognizing truths others would overlook. Some would call this the Sherlock Holmes approach to searching but this skill is much too simple to suggest only a detective of Holmes's calibre can perform it.

I gaze out the window and across the countryside as I drive west from Melbourne. There is such variety in the landscape and I wonder why I so rarely look at the scenery. Perhaps my mind is too clouded. I think of the destination, not the journey.

I once disregarded the internet the same way. I sought my destination caring little of the journey. In the process, I overlooked the hallmarks of quality, how information has history, how information competes for my attention. I overlooked so much of importance. When we focus only on our destination, we rip away the foundations and isolate the surroundings. We lose much of what makes a fact valuable.

The image of this is a page ripped from a book, placed before us. Starved of detail, all we see are the words on the page. We see nothing of the author, the context, its place in a book or indeed if it came from a book, magazine or newspaper. We encounter this information in a raw state, stripped of all context. Do we read on? Do we pause to gather background?

In the words of Alain De Botton, author of many a book on modern philosophy:

> "n'allez pas trop vite. [Don't go too fast.] And an advantage of not going by too fast is that the world has a chance of becoming more interesting in the process."[38]

Busy people do not have the time to do what they are doing, Alain laments. Search the internet. Plow towards a destination. Race for some paragraph or fact. Do we leave time to do what we are doing?

Allow ourselves the luxury of an interest in information and we have a chance to notice the internet's structure and organization so important to gathering comprehensive guidance and judging trust. These emerge only when we give the information an opportunity to tell us; when we take the time to notice.

Alas, if we go too slowly, searching becomes recreation. Appreciation for appreciation sake is enjoyable but not productive. While we may wish to divert our attention to investigate a publisher, to ruminate on the popularity of a website, to forage for link companions and synonyms and such ... what can we do quickly?

I have shown how context and prominence can be simplified to little more than a click on a bookmarklet. Let us now address the many clues that already flash before us demanding only our recognition. We need only to make them explicit. Act like a detective. Open our eyes. See what is already before us.

We will cover two particular skills in this chapter: URL interpretation and a constant awareness of our question. Together they reinforce an attitude of attentiveness to the subtleties and nuances of the internet world.

DEEP URL INTERPRETATION

Every item published anywhere on the internet has an address. This address combines words and symbols, country codes and names of organizations, acronyms, conjunctions, filenames and perhaps even coded information to be interpreted by software. It is all there for us to see.

Any internet user can gaze at an address and notice the two-letter country code and perhaps the name of the organization responsible for the website. To the experienced searcher, the web address may suggest quality, format, date, publisher, type of author and more. Let us tease apart this address and learn as much as possible, keeping in mind that with practice, this will come to take all of four seconds.

The address also serves as a foundation around which we lacquer our experience with internet information. Over time we build an accurate expectation of the types of information residing on .com sites. We begin to know what we will find on personal websites, on serials and on webpages named index.htm. Since we see the web address even before we visit a resource, we have the added incentive that an address may convey something of the quality, topic and depth before we spend any time retrieving the information. Watching addresses, we will learn to skip over a great deal of information unlikely to be valuable to us.

In all cases, the address becomes the steadying staff we lean on so as never to feel quite so lost, quite so far from home. As we travel, read the signposts. We are not just 'somewhere on the net'. We are in a given country, on a website of a given organization, reading a particular perspective as established by a given author and publisher – all of this encoded in the address. We are not lost at all. We are precisely here.

What do we mean by web address? In this book, we continue to use the term 'web address' and URL almost interchangeably. Strictly speaking, there is a difference in that the URL – the Uniform Resource Locator – can point to information placed almost anywhere on the internet including newsgroups and ftp sites. The web is just one of several kingdoms on the internet, though a most voracious kingdom with extensive ties to all others. A discussion list is strictly not part of the web. It is not accessed with the web's http protocol. Except that discussion list messages often escape and become lodged on the web for all to see and search. Almost all messages to newsgroups have web addresses if we invoke Google Groups, the grand search of newsgroup discussion.

The tool invoked to access an item of information is no longer so significant, with the exception of some of the newer tools like BitTorrent. Once, tool was of paramount importance. The distinctions between ftp, web and newsgroup were many and meaningful. As we have discussed, the concept of format now holds more value. "How was the information originally prepared?" is more meaningful than "What does it look like now?"

To my mind, the term, 'the web', has lost much of its precision in general conversation. It has largely absorbed the term, 'URL'. This is not to negate all differences, for the internet can, at times, mean any of three different concepts: a physical network of computers, a logical cyberspace realm of information and a social community engaged in communication. The web only ever refers to cyberspace.

In this book, we continually use the terms 'web address', 'URL', 'the web' and 'the internet' interchangeably. From a searcher's perspective, the differences are simply no longer significant. I labour this point because many search guides take pains to clarify these differences. Dwell on more significant matters.

What does the web address look like? The URL consists of many parts. Each part has meaning.

http://www.spireproject.com/cn/bella4.jpg
|⇐ 1 ⇒| ⇐ 2 ⇒ |⇐3⇒|⇐4⇒|⇐5⇒|

  (1) internet protocol (tool)   http
  (2) domain name   spireproject.com
  (3) directory   /cn/
  (4) filename   bella4
  (5) filetype   jpg

The domain name (#2 above) is further broken into four pieces,

www.spireproject.com
|⇐2a⇒| ⇐2b⇒ |⇐2c⇒|

  (a) hostname   www
  (b) domain   spireproject
  (c) type of organization   .com
  (d) country code   [absent]

At the far left, we have the tool (1). This is usually http and means a webpage. It can also be ftp://, telnet://, news:// and https://. Each tool defines the way the information is presented. Thus, https establishes a secure connection. ftp presents the information as a list of directories and files. If the tool is absent, it defaults to http://.

The next in line is the domain name (2) broken into several pieces and read from right to left. At the highest level (most right) we may have a country code. If the domain name ends in .ie, the page comes from Ireland. .fr means France. .au means Australia. When there is no country code, as in the address above, it indicates the US – unless it is a .com in which case it could be anywhere since we see them scattered all over Ireland, France, Australia and the US. SpireProject.com is hosted in the US but I could host it in Australia easily enough.

The next element to the left of the country code, 2c in our diagram, describes the type of organization involved: .com, .edu, .gov and .net translate as commercial, educational, government and network resources respectively. The use of .edu and .gov are strictly controlled. Others are less formal. Anyone can own a US .com for instance. Newer extensions like .info and .biz also exist and have a meaning. They are not common yet so for a time they primarily mean the .com site was occupied.

Some of these rules vary by country. In New Zealand, we see .govt instead of .gov. In Ireland, .com is dropped entirely. In England, .ac (for academic) replaces .edu (educational). Rules governing the purchase of domain names vary too. The Australian .com.au addresses go to only Australian businesses with a claim to that name.

Remember we can easily restrict our focus to sites with a specific organizational element. Add **inurl:.gov** to any search request and we will only review pages with .gov in the web address. "Show us just the government webpages," we ask. **inurl:.au** reveals Australian websites or rather those websites with .au in the address. A search that includes **inurl:.gov.au** is certainly allowed. Subtraction works too, though removing all .com sites by including **-inurl:.com** is rather clumsy and I recommend against it.

The next element in the domain name, 2b just now, should tell us the organization involved in hosting our webpage. For **SpireProject.com**, a commercial effort calling itself 'SpireProject' is somehow responsible. Trust this name. The website is found on a computer assigned to them. On rare occasions when this name is altered by hacking and spoofing, the change is usually obvious. If we find a file on **worldbank.org**, we are, without doubt, reading a World Bank document.

With one major exception. The use of .com sites can be confusing. Worldbank.com is not the UN's World Bank institution. Harvard.com is not Harvard University. Whitehouse.com for years was a porn site. This affects .com sites almost exclusively since even the kid down the block can register the generic .com. By the way, there are simple and established methods to change the ownership of misleading domain names. Just buy the domain name or, with only a little more effort, assert trademark infringement through the Uniform Domain Name Dispute Resolution Policy (UDRP).

The part of the domain name that appears first, 2a just now, is called the hostname and tells us the specific entryway to the computer involved – the name of the host computer. Often there is only one host computer, called **www** standing for World Wide Web. This standard emerged back before it became clear almost all information would migrate to the web platform. Once we had **gopher.well.com**, **ftp.well.com** and **www.well.com**. Today we can usually drop the **www** though the occasional address still requires it. **SpireProject.com** is the same as **www.SpireProject.com** and any

publisher with a domain should ask their internet service provider (ISP) to make the hostname unnecessary – almost certainly a free service.

On occasion, we may see different hostnames. For example, **ww2.xyz .com** translates as the second world wide web computer – most likely because the first became very busy. Sometimes the hostname corresponds to a department within a given organization. **health.wa.gov.au** translates as the Health host computer of the Western Australian Government of Australia.

By the way, the domain name is always case insensitive. Capital letters make it easier to read but do not affect the address in any way. Publishers should use this to their advantage. Directories and filenames, however, are case sensitive. **Index.htm**, **index.htm** and **INDEX.htm** are different files.

Following the domain name may appear one or more directories (#3 in the last diagram), then perhaps a filename (4) and filetype (5). These elements may not be included and if absent, simply lead us to a default page for that given address. Thus, pages may have multiple equivalent addresses. **SpireProject.com** is the same page as **www.spireproject.com /index.htm**. At other times, each frame and graphic has its own address so we can reach a page yet view it in a way not intended by the publisher.

DIRECTORIES AND FILENAMES

As we reach beyond the domain name, the first important lesson is that a real living person named, arranged and designed the website in a way useful to them. It is someone's hard drive. We are not working with computer generated names. There are standards, habits and conventions in how websites are designed. Directory names, for instance, are used to organize information by topic, by project or, on rare occasions, by date or alphabetically. Never randomly chosen, directory names help a real person remember their contents two years later. Thus **/database/statistics/today .htm** almost certainly means today's statistics for the database. Just read right to left. We often encounter acronyms and contraction within an address. A contraction like FinPub is simply two words smashed together losing many of its letters. Found on a financial page, FinPub almost certainly stands for Financial Publications. Remember, it means something to the person who named it.

The number of directories tells us something too. Find a webpage buried five directories deep, and we have reached a very specific project. With few directories, we look at something more general and of general importance to the publisher. There is a special case for the page at the very top of the directory tree. With no directories and no filename, this page is the initial doorway to the organization. It is a brochure intended to present the organization in its best light and in beautiful graphics. Think introduction. Think sales literature. Do not think content.

Another special case involves the tilde symbol (**~**) leading the first directory. This is a long-standing default for web addresses provided by an internet service provider and roughly indicates a personal webpage. More precisely, it indicates a loose connection between the author as titled after the tilde (~) and the organization responsible for the domain name – a loose connection not a tight connection. For example, my ISP provides me with free webspace at **iinet.net.au/~spire/**, a part of iinet's computer but set aside for 'spire'.

Next, consider the filename. This name means something to the publisher. We name files to help us remember their contents. If a filename has a number in it, like **page03.htm**, then certainly **page02.htm, page01.htm** and perhaps **page04.htm** exists. The top page of a directory is usually named **index.htm(l) content.htm(l),** or **home.htm(l)**. This page links to most if not all of the pages in that directory and also links to any subdirectories. We may not know what we are looking at but we know we will see the contents of that directory from there.

Filetypes like .htm or .html (both mean webpages), .jpg (image), .ppt (powerpoint presentation), .doc (document), .pdf (secure document) and more tell us how the page is presented. This in turn may tell us something of its contents.

Finally, some trailing information is for interpretation by software. As I understand, this pertains only to text following a question mark. This information is broken into separate variables, each separated by the ampersand (&). A Google search, for instance, may have this address:

http://google.com/search?q=inurl%3Acom.ie&num=40

The program 'search' at Google.com is handed two variables: q and num. The **%3A** replaces the colon (:) symbol. The **&** symbol separates each variable.

With redirection scripts, many times the true address is one of the variables. Yahoo often redirects addresses as seen in the following rather cumbersome address. Notice the destination address at the end?

http://us.rd.yahoo.com/search/diy/c416_f318/directoryreference +/common_link3/SIG=11i3k32ue/*http%3A//education.yahoo.com /reference/factbook/

In summary, the address is a string of pearls identifying the contents of a given webpage for the benefit of the author and publisher. The domain name suggests the publisher's identity and through that, something of the content and quality. Directories and filename tells us something of the content and sometimes format. We read the address like a sentence, from right to left.

Most addresses are incredibly easy to read. In 1997 I first published the Information Research FAQ, accepted as a recognized network FAQ and archived at:

www.faqs.org/faqs/internet/info-research-faq/part1/

Do not skip over this address. It is all there. We are reading something from the **FAQs.org** – an organization devoted to FAQs no doubt. It deals with the internet since it is in a directory called internet. Info-research-faq probably stands for Information Research FAQ and this address leads to part 1, so we know the FAQ is in pieces. We read this address like a sentence.

Try another:

www.defence.govt.nz/public_docs/defencepolicyframework-June2000.pdf

Again, like a sentence, it is all there. The New Zealand government agency for Defence is publishing a public document probably titled: Defence Policy Framework. It is a PDF document prepared or released in June 2000. I am reading this only from the address.

The essence of Deep URL Interpretation includes one final task: speed. Remember we are attempting to read the address in four seconds. Quickly digest the web address and describe the contents. Try to be precise but also try to reach beyond what we know for certain to state what we only suspect to be true. Persevere with this and it will be a great ally in our search for information.

Lets try our hand at some of the web addresses I use in my seminars. Look closely at each address and describe the resource before you read on.

1_ www.worldbank.org/data/dev/devgoals.html

Looks like the World Bank to me. Good publisher. Reputable. This file seems to be talking about data, and dev... dev by the World Bank probably means development. dev goals (development goals) but particularly data for development goals. I say it is a document with statistics describing the development goals of the World Bank.

[This page contains developmental indicators for poverty, education, gender equality and the like, arranged by country. In short, it is a page of standardized statistics describing country development.]

2_ www.eia.doe.gov/emeu/cabs/contents.html

Hmm, a tough one. The country code is missing, so this must be from the US government. DOE would be the Department of Energy or the Department of Education perhaps? EIA is a sub department of the DOE. We know this because they have their own computer. We are looking at in the directory of the EMEU, probably a project of the EIA, and CABS is probably a project of the EMEU. Thus, the CABS of the EMEU of the EIA of the DOE is doing something and we will look at the contents page. I have no idea what is involved but this project is specific, very specific since it is buried four levels deep in the US government hierarchy. It will interest us or not.

[This page is a collection of Country Analysis Briefs (CABS) describing energy needs and production statistics by country. It comes from the Office of Energy Markets and End Use (EMEU) of the Energy Information Administration (EIA) of the Department of Energy (DOE).]

3_ www.geocities.com/Colosseum/Arena/4336/fire.html

Geocities is a free webfarm, a place that offers free web space to anyone who asks. Fortunecity, Tripod and Angelfire are others. The directories Colosseum, Arena and seat 4336 just describe how this web farm organizes their directories. All we really know is that the author has placed this information on free webspace, that they do not have somewhere else to put this information and that this information has something to do with fire, the chosen filename. Having said this, it must be unimportant enough

for the owner to risk losing it since free webfarms sometimes implode and disappear.

[This webpage is dedicated to the memory of Bandit the fire dog who worked at the Chicago Fire Department and is accompanied by a very personal and passionate look at local fire fighting history.]

4_ **www.shef.ac.uk/~is/publications/infres/6-3/infres63.html**

This is a serial, a repetitive publication like a magazine or journal. Notice the 6-3. This cannot be a date. It shouts serial. Volume 6 Issue 3, especially since it somehow tied to Sheffield University in the UK. The **/~is/** may suggest a personal website but I know of few personal websites with directories called publications. The contraction **infres** probably comes from the title, too.

I say this is an academic e-journal tied to the Sheffield campus but not an official Sheffield publication. If it was a Sheffield publication it would have a directory that does not start with a tilde (~). It will almost certainly contain quality peer reviewed articles with an academic edge.

[The Information Research ejournal, an international electronic journal, is published by Professor Tom Wilson, of the Dept of Information Studies, University of Sheffield, in association with further professors in Finland, Singapore, the US and Lithuania. The address now forwards to a new domain **informationr.net** established for this publication and other work involving Professor Tom Wilson.]

5_ **members.iinet.net.au/~holmgren/truth.html**

Holmgren, either a person's name or an organization, has a website on an Australian ISP – a network resource with members – yes, certainly an internet service provider. This probably means this is a personal webpage or a small business not yet owning a domain name. We will shortly read 'the truth'. If we found this page through a search engine, the title "The Truth About Sept 11" would accompany this address. Clearly now we are looking at a highly personalized interpretation of those events. I find no intrinsic reason to trust such advice – no reason yet to consider this anything but biased hearsay so I would avoid it. Having said this, it offers strangely compelling reading.

6_ **www.chooseindia.com/tourism/gujarathistory.html**

This page probably comes from either the US or India since a generic .com site does not identify a country. The domain name suggests to me either a commercial travel agency or just possibly, a government site that promotes tourism – a site like **VisitVictoria.com** or **NewZealand.com.** If **chooseindia.com** is not the nations official tourism site, then I expect little depth, little historical accuracy and an uncomfortable bias.

[Choose India Travels published a fine and beautiful brochure much like any brochure about India found at any travel agency. This page contains about 10 paragraphs on three tourist sites near Gujarat.]

HACKING A WEB ADDRESS

Before we leave this idea that web addresses tell us something of their contents, let me show you a simple technique of hacking a web address to reveal nearby information. Usually hacking means taking an axe to something. Chop off a chunk of wood to burn in a fire. Here we chop off the filename and perhaps a directory to reveal a page located nearby. On other occasions, the address to a nearby page is simply obvious. Our aim is to alter the address directly.

Say we land on a file that reads:

http://wicked_ideas.com/jokes/page14.htm

Obviously there is a **/page13.htm** and probably a **/page15.htm**. Page two will look like **/page2.htm** or **/page02.htm.** We can further say that since there is a directory set aside for jokes, there are probably other directories set aside for other wicked ideas. We can have a look by visiting the directory above us.

When we hack a directory, we may find chopping off the end of a web address directs us to the default page for that directory – usually the index.htm (.html) or home.htm (.html) – a page that describes the contents of that directory. On occasions, this will generate an error message – in which case we may simply wish to guess the page we want is called index.htm

This next address points to a directory of International Governmental Organizations on the website of the Northwestern University Library:

www.library.northwestern.edu/govpub/resource/internat/igo.html

Let us find other documents within this website by directly hacking this web address.

www.library.northwestern.edu/govpub/resource/internat/

Chop off the filename **igo.html**, then press return and we reveal a page describing the purpose of this directory. It links to **.../internat/igo.html** as well as **.../internat/foreign.html,** a directory of foreign government websites.

www.library.northwestern.edu/govpub/resource/

Chop off another directory and we reveal a webpage that lists the many government resources this library has prepared for us, including various electronic collections and information about staff and opening hours.

www.library.northwestern.edu/govpub/

This webpage titled "Government and Geographic Information and Data Services", introduces this depository library.

www.library.northwestern.edu

Chop off all the directories and we reveal the homepage for the library of Northwestern University.

We travel this path back up the hierarchy of organization simply by making short sharp chops with our axe. Highlight the portion of the web address we want to remove, type **delete** or **Cntl+X,** then press **return.** Little effort. Little fuss. We hack the web address.

Sometimes we will hack a directory and receive an error message that says we are not allowed to view the page in question. This is the time to guess there probably is an **index** or **home** page in that directory. If all the pages are ending in **.htm,** so will the index or home page. If all the pages end in **.html,** so will this one. Few websites mix **.htm** and **.html.**

This scene gets a little more complicated if the web address includes .asp or some computer variables but the concept holds throughout. We can alter a web address ourselves to find nearby information.

I wanted to see a movie last year so I looked at the reviews for the Da Vinci Code and Slither on my favourite review site: **RottenTomatoes.com**.

I first look at:

http://www.rottentomatoes.com/m/da_vinci_code

In a simple guess, the next page I want will be:

http://www.rottentomatoes.com/m/slither

Filenames are selected to mean something to the person who names them. Directory names are selected to organize information in a way that will be obvious two years later. Use this to our advantage by hacking the web address. Keep in mind, the context bookmarklet is more useful under most circumstances but this is just one more way an understanding of the URL can assist us.

PREDICTING CONTENT WITH URLS

Let me state clearly, Deep URL Interpretation is an excellent search skill. It is sound, fast and very helpful in directing our attention profitably. If we do not want to look at a travel brochure to India, we do not visit **www.chooseindia.com** or we visit only long enough to determine it is not published by the Indian government.

Certainly, you may be clumsy at first, wrongly interpreting web addresses and projecting far-fetched expectations where they do not belong. However, with time and a little practice, Deep URL Interpretation works splendidly. We look over a list of resources and instantly see which ones are most likely to have the information we seek.

A new image of the internet now emerges where we see the internet ahead of visiting it. Suppose we ask Yahoo to recommend some prominent material on a topic that interests us. We can now predict our way to information we seek, discarding vast quantities of information merely because we suspect it would not treat the topic in a manner to our liking. Instead of reading titles, when I look at a search engine results page, I read web addresses. Addresses tell us more.

Did you know we can ask our web browser to flash a link's address in the bottom edge of the web browser whenever we move our mouse over a link? In explorer, select View→Status bar. This allows us to hover our mouse over a few links and see their destinations. Oh, this links to an article in the New York Times, this links to a personal webpage and that links to the organization's website. Even if the page we are on tells us nothing of this in words, we can read it from the web address of the link.

A central theme in this book is the gradual reconnecting of elements found in cyberspace with their corresponding elements from away from the internet. Field searching is the author/title search of the library catalogue recast for the internet. Format applies to the internet just as it applies to books and articles. URL interpretation is another such situation; another bridge.

This predictive power tied to the address greatly resembles our predictive power for information in the 'real' world – a world of magazines, brochures and phone books. We know what we will find when we visit the reference section of a library or approach the help desk of our local government council or call the media department of an international corporation or reach for a local paper in a coffee house. Each 'place' is a touchstone around which will build our expectation. If this resembles our earlier discussion on format, you are most certainly correct. Experience extends beyond format to destination, author, publisher and whatever else we find. Anyone who ever reads a tabloid will never forget the experience.

Welcome to the Sherlock Holmes style of attentiveness. Holmes counts stair steps. He listens to the footfall. He invites Watson in just as he pauses to knock on his door. Sherlock Holmes is addicted to heroin and has traits of an autistic savant. To each their own. However, if Holmes were to visit the web, he would certainly notice the web address.

Holmes also displays a second skill we need to rediscover – an ability to ask the right question at the right time.

ATTENTIVENESS TO OUR QUESTION

> The question is not what you look at,
> but what you see.
>
> Henry David Thoreau

We reach for the internet and begin a search. Sometimes we have a clear question in mind. Sometimes we have only a vague need to explore. Till now, we have assumed we generate our question with ease. This is a typically naïve approach and one I shared for many years. Unfortunately, in assuming we know our question, we miss four generous opportunities.

Firstly, our question is our flashlight illuminating what we see. Ask a better question … we get a better answer. The gulf between a good search and an excellent search often simply depends on the elegance of our question.

Secondly, our search tools respond best to certain kinds of questions. We can phrase a question to gather better information from a search tool. We have already encountered Boolean, proximity and field searching as a way to draw better answers from a global search engine. We can also frame our questions to make use of prominence. The commercial database responds best to lengthy precise search queries rich in the use of descriptive fields and a thesaurus. On a discussion list, demonstrate we have searched and ask closed, answerable questions. In each case, we ask our question in a way suited to our search tool.

Thirdly, as we build our questions, we establish a dialogue with the world of information. This is the elevated vista we will discuss further in Chapter Nine. What do we think exists? What will it look like? Where will it pool? Answers to questions like these allow us to refine our image of the internet and to answer further questions; questions like: "Is this all the information we will find?" "Should we give up looking?" and "What will the best information look like?"

Fourthly, questioning is the essence of a search – the most enjoyable aspect of a search. There are few prescriptions for crafting questions. This is pure art. We follow hunches, explore the information wilderness and work with nebulous concepts, delicate nuances and notions of the possible. Jump over this step and we give ourselves little time to be artistic.

A non-trivial search is a journey. We draw from our experience, from our knowledge of available resources and from our artistic soul. This voyage has a start, a middle and an end. In the process, we establish a dialogue with the internet.

As a search develops, our understanding of the topic, our expectations of what we will find and our frustration will gyrate wildly. As we journey, we continually ask many questions both of the information we encounter and the destination we draw towards. This questioning is also central to quality assessment. Who is this author? What else have they published? Is this work prominent? These small questions fill the interlude between grander questions we ask not of the information but of our voyage. Are we

getting the sort of information we prefer? Are we missing something critical?

A wooden sailing ship of years gone by tacks back and forth towards a destination making the most of favourable winds. We decide the night's anchorage en route. We negotiate the open water, the inshore reefs and the storm that breaks upon us. Only the armchair admiral drinking port by the fireplace on dry land would dare simplify instructions as "Sail to Gibraltar".

A non-trivial search behaves the same. We have many thoughts to consider as we move towards our goal, especially when our goal remains uncertain. Our goals may simply be "Find some monkeys in Spain" or "Plan a lovely vacation near Barcelona." When we start our journey there is no way we can state our question precisely. We do not yet know that Gibraltar, with its many Barbary Macaques, its many rock apes, will be our destination. To suggest we should not change our question mid-search is like nailing Jello to a wall[39] – rather messy and not particularly productive. Our search will move beyond Spain, certainly well south of Barcelona.

Journeys are lived. It is true a brilliant general may try to plan a battle before it begins but they would never dare leave the battlefield without leadership as the battle rages. Too many decisions must be made that cannot be envisioned in advance.

Librarian friends of mine always recount how we start with a general question and progress to more specific questions. I am hard pressed to express my dissatisfaction with this approach but I remain uneasy. Sometimes we start with something that only at a stretch would we consider a question. We may start with a vague interest. During the search we continually alter our approach, asking wildly divergent questions and questions about questions; meta-questions like are we finding the information we want? Are we wasting our time with this approach? Is there a simpler way to proceed? And then there is my personal favourite: Shall I ask a librarian for help?

A little girl in pigtails approaches a librarian at a public library. "Excuse me…" she says in a small, squeaky voice. "I'm looking for something about trees." Please, someone hand her a picture book.

What just happened? The little girl's question is simply best answered this way. We look at her needs and skills, then make a simple judgement.

We correct her question with a better question. "Where do I find a picture book about trees?"

Next an aging carpenter approaches our desk asking for help saving a tree he just mistakenly cut in half. Hmm. We can correct this question too. Where do we find a helpful tree surgeon to converse with.

This first step so totally defines our search. We no longer wander in the dark, looking at the whole gamut of resources: commercial, public, internet, from books to articles and beyond. We have narrowed our search rather precisely to just that area we expect to find an answer that will satisfy us. We want a colourful picture book and a talkative tree surgeon. This happens time and time again in searching, at so many stages in our journey. Our search becomes a string of questions. Each question hopefully brings us a step closer to our eventual destination, or if not, then confirms an avenue as unpromising.

About five years ago, I demonstrated some search techniques for Trade New Zealand. We sought marketing background on an unusual resource. Side by side with another researcher, we looked through the internet for information that would help.

My searching was not particularly better and pressed for time, my results were severely limited in scope. However, I asked the right question. At one point during the search, my mind wandered from the quest for information to the quest for who had such information. Could another New Zealander have tackled this question before? Could we make contact? Adding a quick **inurl:.nz** limited my current search to just New Zealand domain resources and a **.govt.nz** resource easily caught my attention. It demanded my attention, actually. Yes, another New Zealand government agency had completed an extensive project on this very topic just a year earlier.

Ask the right question, reveal the right information. There is a degree of mental dexterity and horizontal thinking that lends itself to this skill, as Sherlock Holmes would surely instruct us. This skill starts, however, by being attentive to the questions we already ask. A complex search is not just a quest for the right information. It is also a quest for the right question.

Notice what I am suggesting. Above our search for information rests another search we should take equally as seriously – a search for the right question. Search without concern for our question and we will probably

find the information we initially desire. However, we will completely fail this second search – the search for the right question. I may well find something of interest on an obscure commodity for the New Zealand Department of Foreign Affairs but I would never tap the far greater experience of the government-funded expert I eventually uncovered. For that, I had to ask a better question. I had to ask if someone else was at hand with the experience I needed.

We must make this second unspoken search audible.

FEEDBACK

Let us revisit surfing for a moment from the perspective of a search for the right question. We approach a search engine and allow it to bring to our attention what it computes will interest us. The search engine returns a list of twenty or so webpages. If we do not find the information in the first twenty, we are invited to look at the next twenty based on the same flawed selection criteria.

Say we notice no useful matches. Perhaps the list is skewed in the wrong direction. Perhaps we are missing the voice of a type of publisher we thought would have more to say. Our search failed. Do we really want to start adding additional words and search engine punctuation? It might help but we do not want a better failure. We want success.

Feedback offers this possibility. Find one page that interests us more than others. Look at this page and try to decide what it is that interests us. Look for key concepts and phrases that might describe what we found. Look at the publisher and author. Perhaps the format sets this page apart. Now use this information to craft a better search or to justify searching somewhere else. We 'feed back' what we learn into the next question.

Feedback picks up on this idea that a search proceeds on two levels. The first level – a hunt for an answer – masks this second level – the hunt for the right question. If our search fails to reveal something of our answer, why not try to find the right question instead. All that is required is that we notice the questions we are asking, then try to ask better questions. Further along the journey, as our search gains resolution, we will go back to seeking an answer armed now with a better question.

I have a delightful demonstration of how a search proceeds on two levels and it goes like this. How do we find the cheapest airplane ticket to

Paris? Think about it. This is a tough question. The answer: Ask how to find cheap airplane ticket to Paris, then work from there. The best answers often come from asking the best questions.

INTENTIONALLY IMPRECISE

Two further search techniques make good use of this newly found attentiveness to our question. The first approach involves varying our questions in small steps instead of trying to leap directly into a very precise search.

Need help with a damaged tree? We need an expert. Experts congregate in mailing lists. We know there must be a directory of mailing lists. Let us start there.

We sidestep some of the difficulties of using either a tight telephoto or a wide-angle lens by slowly zooming into our target. The advantage is simply that by stepping slowly, we have a better grasp of the questions we can ask. We should never be in a situation where we dump words at search engine and say, "Tell me something about fish breeding" – a hopeless question of little value. Equally, we are not ready to ask the most specific questions until we know more about the information landscape we will search in.

We can also be intentionally imprecise in such a way that we capture information we know only a little about.

· A page is missing. If the page moved, we could search for just the middle part of the web address on the assumption that the page moved to another computer or the computer was renamed for administrative reasons.

· A person's name in quotes does not reveal who we thought we would find. Perhaps we can allow for a middle initial.

· An Australian search for a website fails. Perhaps its web domain does not end with .au? Can we be imprecise about the country code?

· We search for a Brisbane hotel in the suburb of Redcliffe. **Brisbane Hotel Redcliffe**. No luck so we drop Redcliffe and

search again. Too general. We drop Brisbane and learn that Redcliffe is not, in fact, a part of Brisbane. It is a nearby town.

Lastly, we have an intriguing way to be imprecise which we will explore further in Chapter Nine. Instead of searching for a critical page with the information we seek, perhaps we want to search for the neighboring page, the 'page next door'. In an internet where perhaps 10% or 20% of the web is indexed, we may never find the information we seek because the page is not indexed by our search engine in the first place. However, turn our attention to the page next door, the page that describes the report or introduces a directory or describes the many projects produced by an association. These pages often attract more links, are more likely to be indexed, and on occasion, are easier to search for.

In the simplest of examples, some full text articles referenced by the PubMed medical database can only be approached through PubMed, not directly. Being imprecise may be our salvation.

We will take many steps as we ask questions and our questions will lead us first one way, then another. We will backtrack and try again when we get lost. We will clarify our search as we build better expectations of where we want to be. At the end of a complex search, we usually see just one or two questions were pivotal in our success. Our other questions merely set the environment for these pivotal questions to emerge. However, none of this will happen smoothly if we do not listen attentively to the questions we are asking.

PAY ATTENTION

We have covered a great many aspects of information in this book. Many subtle and minor relationships exist and add structure, colour and clarity to the information we encounter. By this point in this book we have good reason to consider the internet as resplendent in structure, colour and clarity.

To help reveal this, we have continually shown internet information as similar to non-internet information. We noticed how format applies to internet and non-internet information alike. Format is a fundamental aspect of all information. Format drives content (to a degree).

We drew our attention to the importance of the publisher to internet and non-internet information alike. Publisher identity is a fundamental aspect of information too. Publisher identity drives bias as well as content (to a lesser degree).

We also looked at quality, at prominence, at structure. All these non-internet phenomenon apply equally online.

Well, we finally reach a point where we must break away from bridging experience we learned elsewhere and start building our own reservoir of experience from our time online. Yet even now, we can copy since we want to build our experience in a very similar way to how we build experience with non-internet information.

I find science books by science writer Paul Davies less interesting than science books by Professor Stephen Hawking. In fact, I generally prefer books by scientists about science though I prefer books by journalists about conflicts in science.

In a similar way, industry insiders write business books that excel at revealing the inner workings of an industry. And I love them intensely because they answer certain questions candidly which outsiders are loath to mention; questions like what an industry will look like in the future.

I intensely dislike the ramblings of blogs – an online diary of sorts – just as I dislike mailing lists with a great deal of traffic unless they can be word searched. Today's popular computer magazines and news media seem to adore blogs but they have been very wrong before. They pumped push media and many of the minor search engines. They loved meta-search engines long after I thought little of them.

This kind of advice is just hearsay. Such personal feelings, ungrounded in any in-depth analysis or study, cannot be trusted. However, these mere suppositions and expectations form a body of experience that guides me around the internet. Oh, I abandon this experience quickly enough when an alternative suggestion arrives but I do listen to this experience when the internet is quiet.

You too will form this kind of experience. Dislike rambling blogs? We will not ask for them. Love insider books? Let us go look for them. Lost our way? Let us ask how to search before we proceed. If it works with cheap airplane tickets, perhaps it will work on our next search as well.

Make the most of this kind of experience. Actively build this experience from your own suspicions and expectations. Hang this experience on

the URL and the questions we ask. On the internet more than anywhere else, we persistently encounter apparently new information presented by apparently new sources. If we do not see this as an opportunity, if we do not see our previous experience with similar organizations and similar questions as relevant, then we will mistakenly assume we have little experience to draw upon. This only enhances the 'lost in space' experience we may feel on the internet.

However, it is with our attention that we turn this situation to our advantage. Yes, we have never encountered THIS publisher before nor have I asked THIS question but we have encountered similar publishers, similar formats, similar motivations. We have asked similar questions.

Mindful of the web address and mindful of our question, we build onto our reservoir of past experience. Even Sherlock Holmes would applaud.

Part Three:

FINESSE

Chapter Eight

_____

UTOPIA

The defiance of Toulouse was brief and foolish. 40,000 men at arms approached a city far too large to defend and far too sensible to choose suicide. "Suicide would have been simpler," grumbled Albert a week later as he struggled to house the many conceited knights and ennobled thugs within the best homes of Toulouse.

Peace had cost so much. Immense payments just negotiated would impoverish the town. The most prominent Cathars would burn at the stake. Yet Toulouse had at least evaded the fate of Carcassonne with their deadly six-week siege during a heat wave, or that of Béziers where twenty thousand residents were slain that first night of the war.

Foreigners now staggered about Toulouse in a drunken stupor. Locals, so supportive or at least permissive of the Cathar church just weeks before, now sat demoralized, afraid of their future and afraid for their children.

Albert feels the fear too. A young nun turns away the affections of a northern Duke. Rejected, the Duke declares her a heathen and tosses her in prison. "It's obvious," he says with twisted logic. "Only one consumed with a reverence for chastity, only a Cathar, would dare reject me."

This foolishness is not corrected. The Duke will not see reason. Albert considers intervening too but knows his own Catholic allegiance would be called into question. Where was Albert when the city fell? Unable to defend his loyalty, Albert stays silent. The next day the young girl takes her life as she waits for trial.

Even as Albert despairs, this same Duke writes to inform Albert he must supply soldiers to continue the crusade.

<center>*   *   *</center>

Any understanding of the internet must account for the near religious fervor it incites in those who believe internet information will set the world free. Some like myself, believe this technology will assist the world to become self-aware. It is our best tool to forge a truly post-modern world that invites a plurality of views and recognizes widely diverse perspectives. It catalyzes many welcome changes to our global society.

Utopia is always a personal story. It is also a story that will help us better understand the internet; to realize why information migrates to the internet and where it will lodge for us to find.

My story started with a dream that the internet would transform and empower our world. I saw in this technology a promise of information for all. In earlier days I saw and assisted the internet to re-energize the ideal of government transparency. I saw and assisted the internet to liberate information from the dusty shelves of hidden-away private libraries. I saw it liberate information from the cost of printing just as it liberates information from the shores of the developed world to permeate into countries more in need.

Looking ahead, I envisioned something even more profound, as utopian dreamers so often do. A vast reservoir of information. The world's prize jewel. An archival monument to human thought and endeavor. I prophesized the internet becoming the catalyst that clears away so much of the unsubstantiated, irrational and deceitful that goes for truth in our world. I was young and very much a dreamer.

The internet did not live up to my idealistic expectations. I suppose the internet's early gifts were so profound that later gifts seemed less monumental. Insufficiently impressive. Instead of seeing ever-greater glory, I began to see the limitations of this new technology. I see it now in less exultant terms.

Oh yes, the internet became all I foresaw. It kicks off many wonderful changes to our society. It will probably become the primary repository for human thought early next decade. Yet the internet is not all I dreamt since part of my dream was unspoken, still nebulous in my mind. My utopia included a truly breathtaking but unformed ideal of an information realm open to all and organized by merit. Yes, I dreamt of an open meritocracy.

The internet has not become a foundation built on merit and far from universal opportunity, great oceans of information that surround us are

effectively barred from participating. Let me state this pessimistic view most plainly:

> 1_ The internet is vastly oversupplied with information. Much too much is available and this causes difficulties for internet publishing and organization.

> 2_ The internet has a short supply of good real estate. Too few websites gather most of our attention.

> 3_ Selling internet information for money remains difficult. Until recently, it was largely a demonetized, moneyless zone.

> 4_ The internet intrinsically struggles with poor and under-funded vetting. Our virtuous freedom from censorship is also a serious defect: a lack of selection and filtering. We add our own but it is not done well.

All these flaws are real, significant and obnoxious. The empowering internet also entangles us. Interestingly, most of these flaws only came to light after the internet grew out of its infancy. In earlier days, the internet looked and behaved in more inspiring ways. Once upon a time, good content alone would earn fame. Today, content must marry prominence to reach an audience. Tomorrow, content without prominence will be irrelevant.

But I get ahead of myself. This voyage started with a mantra of information purity. It started with a dream.

THE UTOPIAN PUBLISHING MODEL

The early settlers of the internet believed strongly in this technology. Perhaps like early settlers in all new frontiers, we saw something worth fighting for. We saw an opportunity to rewrite the rules that bind us so as to communicate more freely and effectively. We would liberate information from closed systems, from occupational fiefdoms, indeed from physical reality itself. We saw great beauty in an Internet Informed society. We gloated how it levels archaic industries, improves the way we work and helps us make better decisions. We looked ahead and saw it reinvigorate life-long learning, civic attitude and government transparency. It would reduce some of the worst excesses of a capitalism based on incomplete

access to information. It would assist the underdeveloped world immensely.

At the heart of this utopian enthusiasm was a willingness to publish information onto the internet solely for the benefit of the internet and those online. Yes, in the name of peace, harmony and information liberation, we publish because we care. We publish as a sort of volunteerism.

The chant, "information wants to be free," is a facet of this view. This slogan, and the romantic dream at its heart, springs as much from a wish for something better than capitalism as a very real opportunity to develop a promising technology. We may phrase this motivation as giving back to the internet or giving to the internet because we love humanity but the essence is that we do not demand a reward for publishing.

Let us call this the Utopian Publishing Model: publishing motivated by a utopian dream.

Oh, many early publishers did expect a reward eventually. Perhaps banner advertising would come to pay twenty cents a visitor once advertisers saw our value. Perhaps the occasional small donations common to freeware would spread. Surely something would happen so that running a pivotal website would one day pay handsomely. Trends like rising traffic and a growing importance of prominence suggested a rosy future. Until then, perhaps the fame of being a pivotal internet publisher would gain us extra work and respect away from the internet. Besides, everyone at the time was getting more from the internet than they tossed in. Everyone was finding information they needed and all of it was free. Financial concerns were overlooked, postponed or ignored.

There are serious flaws with this approach to publishing. Volunteerism does not motivate the most talented of us. Trips to Tahiti and opera tickets to La Scala do a better job of moving the more skilled of us to action. Many of us are simply unable or unwilling to participate as volunteers. Of course, such people were not on the internet in the early days - not in the beginning. If they were, they published and reaped rewards in the shape of free information. They published and thought of the future.

Volunteering has other demands too. Firstly, volunteers must have no pressing need to make a living. Many of the early internet publishers were university students or professors. Secondly, the volunteer must value working for the good of our world – a noble but not universal inspiration. Thirdly, when volunteers need things like computers and telephone line

rental, they must pay themselves or find a benefactor. This was not much to ask at the start since we needed so little but as time progressed and as internet publishers had more and more promotional and design expenses, volunteers were among the least likely to pay. Do we really pay US$199 to be considered for a listing in the Yahoo Directory when we volunteer our time and experience as well? Later this price would rise to an annual fee of US$299. Volunteering time may be possible for many teenagers, senior citizens and not-yet-experienced experts. Volunteering money is another matter.

Lastly, volunteerism does not persist well over time. Eventually, most donors seek a return on their investment of time and energy. This often emerges slowly as a gradual desire to make a project self-funding and this transition often sounds the end of volunteered projects.

Time has passed and I have grown older. I have a daughter now and while not directly relevant, having a child did push me towards realism. I also had ringside seats to the development of this little piece of utopia. I have seen for myself how gradually this pristine dream of free information for all has shown itself as an illusion – a transient illusion.

At first the utopian dream flourished. Early activists recognized better information as an ointment for a wide range of social ills. An enriched public discussion on issues like the environment and poverty would help everyone. We could improve the world simply by creating an environment where everyone tossed in a little of their expertise while busily plundering the experience of those who contributed before them. Give a little, get a whole lot. The hackers ethic: "Information wants to be free," expressed this ideal too. The slightly illegal connotation was not too unpleasant. It was a war and we had our catchphrase.

Yes, it was war. It is hard to believe it these days but when commercial interests first came to the internet and started to demand a return on their investment, they were shunned and deeply disliked. Don't tease us with a chapter of your book, then ask for our credit card details. Give us the whole book or nothing at all! Don't plaster the walls of the internet with marketing slogans and advertisements. The internet is a sharing environment and those trying to change that were not welcome.

For a time, commercial interests were even banned from using the free internet backbone altogether, though this was not a victory in any sense. The delicate and pristine environment fed by volunteered information

came under threat and it was hard to defend cyberspace from greedy capitalists without manners.

This early internet was a transient phenomenon where giving was the norm. Its very structure enticed volunteerism. Its expectations fostered a belief that a little work now would deliver great changes and rewards later. It was not to last yet it had so much going for it.

Most commercial ventures invest deeply in promotion and marketing. Internet publishing in those early days had none of these expenses. It shorted out the cost of exchange by simply giving information away for free. As for promotion, a good document placed in the right spot would automatically attract attention according to its content value. Publish a great report or FAQ, then sit back and watch as colleagues link, the Scout Report mentions your name and in time, the Yahoo Directory lists your document. Information reaches its audience without further involving the publisher.

There were savings to be made too. Tremendous savings. I remember vividly how I helped a government agency with a report that cost twenty dollars to print a copy on paper but only fifty cents to deliver a copy online. That is a lot of savings. Experts everywhere began to build the world's pool of experience. Great excitement was in the air.

Furthermore, this liberating movement dared to ask questions only the internet could answer. Why can't a peer-reviewed journal publish an article sooner than six months from now? Why can't I read the complete Starr Report on President Clinton's sex life in full, now? Why can't I sidestep the influence of an Americano-centric media and read directly what Fidel Castro says about Elian Gonzales?

Such a beautiful utopia. Such an inspiring dream. Unfortunately this promised future was actually rather stupidly self-contradictory. We dreamed of an environment where information was free, where authors and publishers were entreated to participate and where the organization of information was adequate or better. We also wanted an arena little affected by the winds of promotion and marketing. Essentially, we wanted our slice of cake but not to pay for it.

Strangely, for a time, this is exactly what we got.

The utopian dream persists today and will always persist among those who have no pressing need to earn a living, who love the internet and who want to give their experience freely to the world. At its best, volunteers

form groups – associations really – where teams of volunteers work together towards a common goal. The operating system Linux is made in this way. The freeware movement in general works this way. Indeed, the very font of the words you are reading, Gentium by Victor Gaultney,[40] has a utopian motivation and a volunteer support group.

Many nexus points are motivated by a utopian dream. The Online Book Initiative brings ten thousand books, and Project Gutenberg seventeen thousand, to the internet for free by harnessing volunteer effort. The Wikipedia is a free multilingual encyclopedia based on volunteer content. It sits near or above the quality of commercial encyclopedias. Do not assume quality suffers under this publishing model.

However, this motivation does throw up a great deal of garbage and a great many projects that die for lack of persisting enthusiasm. For many of the early internet publishers, the utopian promise faded with time. It was only the initial transient promise after all; a promise of how it would look before the quantity of information outstripped our ability to digest and organize it. As the internet grows, it continually threatens volunteers with anonymity and the need for promotion. This is exactly what volunteers fear most – wasted effort and the need for additional effort in areas they have no interest. Volunteers giving information want to give information, not market and promote. Whenever promotion and marketing becomes a significant expense, utopia slips away. Part of the success of group efforts like the Wikipedia and the game Wesnoth is how they harness volunteer efforts while managing the marketing and promotion on behalf of all volunteers. Older and wiser, I too found it hard to postpone a move towards self-funding.

The utopian model usually sidesteps the supporting tasks of editorial work, graphic design and organization. Anything the author is unable to volunteer is set aside. Digital information may not require this assistance but all information benefits from editing, organization, vetting and promotion. The utopian approach is not rich in any of these.

Thus, we can spot internet material published the utopian way because it lacks many of the hallmarks of professional design and promotion. It looks not amateur but low budget. If wildly successful, it may positively reek of group input. Otherwise, utopian projects tend to be older, intermittent and focus on a single often highly informed individual.

This model is also the most nimble of the three – the first to supply information on a topic, the most outspoken and the least wary of bias and controversy. Many questions that do not demand rigorous quality control are ideally answered by a resource produced under this motivation. Beware though. Unless part of a group project, utopian publishers have little to curtail bias and exaggeration. We must review the author's experience carefully. Always look at utopian information in context since the author's experience largely determines quality.

THE COMMERCIAL PUBLISHING MODEL

Money please. Pay for this book at the cash register before you leave. With a heart of gold, the commercial publishing model asks, "What can we sell to make money? What return on our investment can we achieve?"

Marketing came to cyberspace and decided to stay. Like the vast army encamped around Toulouse inspired by Pope Innocent's offer of redemption and the promise of plunder, today an army of business minded publishers compete for our attention. The loser is the naïve Cathar faithful who plainly asks, "Can't we just be friends?" It is the utopian author-publisher who asks, "Can't we just share the internet?"

Of course, we cannot be friends. I don't want to share. There is far too little of the best real estate and I want it for myself. While the internet technically scales easily to hold all human thought, most publishers do not welcome the opportunity to publish in some anonymous corner, far from the hunting grounds of searchers like you and I. Arrayed against volunteers giving their expertise freely stand individuals and companies interested in developing a market in this new environment – perhaps a very lucrative market since promotion was at first so very simple and publishing so very inexpensive.

It turns out a serious flaw afflicts the commercial approach too. It does not work. Like a royal knight living in medieval Europe, we can save all the damsels in distress we desire. We just can't bill'em for it! Early internet users were very unwilling to part with money for information they felt should be free.

We can sell books online. We can sell motorcycles. We cannot sell just information. Time and time again with newspapers, newsletters, special reports, ebooks and commercial databases, no price could be found with

sufficient return on investment to impress a shareholder. Oh, speculation was rife – especially during the dot.com bubble – but real figures were universally disappointing. Direct sales of information was a bust.

We are discussing the sale of pure information, not the wider idea of commerce on the internet or of brokerage or advertising or similar services clothed as 'information exchange'. We will discuss those opportunities in a moment. This is about the sale of facts and advice and it did not work. Price competition was the Achilles heal since the internet was already populated with publishers willingly giving information for free. The internet's implied promise suggests a little more searching will find much the same information, much the same guidance, free of charge. This ran contrary to the hopes of business people who preferred to believe that a taste of good information would lead internet users to reach for their wallet.

A further challenge involved the rather sudden collapse of distance, forcing information into competition with equivalent information from comparable experts. This often triggered a race to zero as competitors dropped price to build audience.

The CIA World Factbook is a fine, respected and free publication. Comparable commercial compilations just could not compete. Should the CIA World Factbook ever seek to recoup their costs by charging readers, everyone would probably shift to the Library of Congress country profiles instead.

Books have a presumed value. Internet projects have a presumed equivalence – particularly for searchers not paying attention to the identity of the information they consume; to those who do not attend to context/format/source. A commercial publisher must convince skeptical customers both that the publisher provides value and provides something very different from what else resides online – a challenge indeed with the internet painted as a realm containing everything. Further complications included security concerns and the high cost of collecting money over the internet. In practice, only pre-existing customers looked favourably upon paying for internet information.

Of great significance, the databases of the commercial article world – those that populate database retailers like Dialog and LexisNexis – did not reach the internet in a convenient and economical way. Though poised and ready, they never managed to cross over and establish a popular

market where internet users could buy a single article for a near negligible sum. Even uncoupled from the database retailers and going it alone did not seem to work. Do you remember when the search engine Northern Light tried to bring a rather general, poor quality business article database to the internet? Northern Light is no longer in the search business and US$1.80 was too much to ask. Oh, free databases like the Library of Congress Online Catalog (LOCOC), US Patents Online and ERIC (the digital library of education related literature) became wild runaway successes. I still remember my joy upon discovering I no longer had to pay 50 cents per record plus connect time to review the card catalogue of the Library of Congress. Patent information once cost so much more. Now, basic details are free. Yet commercial databases could not follow us onto the internet. They continue today largely selling subscriptions through relationship marketing to shrinking audiences of libraries, universities and dedicated professional users.

This is no minor loss. Until recently, the commercial information world was the older, larger brother of the internet. In 2006, Google boasted more than twenty billion webpages and messages indexed. Yet I remember over a dozen years earlier searching Global Textline, a commercial database of over four billion news articles sourced from various international newspapers. One database, a fifth the size of Google! And this when the internet was in diapers! Global Textline was a fine database and one of the largest but there are tens of thousands of databases indexing or containing all manner of articles, books, reports and more, all tightly organized and searchable in a refined manner yet unable to mass-market effectively on the internet.

At one time, technology promised a viable internet-based market fed by a true micro-payment system using digital money. If small amounts of money could change hands for just a few cents, then royalty payments and purchases of five or fifty cents a page become a reality. With a digital wallet, a simple pop-up reads "5¢ please" and we click [YES]. With such a system in place, it was predicted information produced under a direct commercial model would flood online. Many models like the Theseus project[41] and even Xanadu[42] from forty years ago describes what per-page royalties looks like and how it reinvigorates internet publishing by funding authors directly.

We have none of this on the internet today. Digital money and micro-payments never seemed to arrive. What we have today is little different to using a credit card. Fees are not insignificant. Five-cent payments are impractical. And in its absence, we have a weak, anemic commercial environment lacking a proven business model and largely stalled, waiting for a sudden shift in the wind.

I do not mean to suggest commerce will never work on the internet. Quite the contrary, I am fairly certain it will … eventually. I merely stress that the direct sale of information has had a troubled past and will have persistent difficulties in the future until viable markets develop.

Yes, for some reason selling information requires a market. Anyone who ever tries to hold a garage sale on a quiet street understands this. A market is an abundance of shoppers, not speeding drivers thinking of dinner and determined to ignore us. Very few internet markets for information have developed.

Remember, the internet is not a market. It is a platform on which markets may develop. Unfortunately, at present we just don't link to, quote from and lead internet visitors to information they must pay to visit. Instead, we reference them in bibliographies we assume no one reads. If we do link to commercial content, most visitors would assume we get a kickback for securing a sale! A market in priced information should not look like this.

We see the same situation in real life. A single antique store in a small village does not make a market. A single antique store on a busy street does not make a market. Three antique stores beside each other with ample parking and word of mouth promotion and we have a market. Satisfied customers flushed with the glow of buying a bargain; we have a market.

Said another way, tap a need, satisfy a customer, we have a sale. Build the habit of buying, spread it around, we have a market. Information markets have not been quick to develop on the internet. They grow very large but not quickly.

Existing markets tend to focus on pre-existing clientele making use of pre-internet markets. Book buyers buy books through bookstores. Oh, the bookstore may move onto the internet but this step does not create a market. Book markets already existed. We came online for the bargain. Amazon's grand book database was a significant improvement. It moved

Amazon.com from rural to urban real estate. Amazon's abundant book reviews and user comments further enhanced this market. Amazon and Google's efforts at searching inside books would enhance it further as it comes to pass. These companies are opening book literature for view, consideration and mass participation. This is what an internet book market looks like: many informed customers buying, forming buying habits and passing on their satisfaction.

Part of the support for the book market is that books enjoy the presumption of depth and value, just like antiques. So does distance education from distinguished universities. Universities are busy creating a market to bring buyers and sellers together in a way that informs hungry prospective students and introduces them to talented educators. This market is not established. Not yet full of life. It will probably be very big business in a decade. I expect it will come to finance some of the world's most talented teachers to prepare some of the world's best educational content for delivery over the internet – something very different to what dominates the internet of today.

Priced articles do not have an effective internet-based market. In part, this is because articles are too expensive for most of us. How many five-dollar articles on planting roses does it take before the articles cost more than the roses? Information is not worth what it was a decade ago, though it is taking a bit of time for article prices to fall. This is one of the grand influences of the information revolution we will discuss shortly. In an oversupplied over-informed world crowded with experts, a six-hundred word article from an old Melbourne Age newspaper is probably worth less than A$2.20[43]; an old New York Times newspaper article is probably worth less than US$3.95.

Actually, articles are worth whatever we willingly pay but above a dollar it seems to interest few of us and enthuse so very few of us that it largely remains apart from the internet, available through but not on the internet. No market means occasional sales but little commercial reward for those involved. I am delighted to see the New York Times and their TimesSelect service offering past articles at US8¢ or less.[44] This is much more in line with earlier predictions of micro-payment environments and may give rise to a vibrant internet market further enhanced by internet publishers who link to specific archived articles. Will it offer value to the New York Times? Perhaps not but I am thankful they are trying.

We cannot market most commercial information on the internet. Oh clearly we can deliver it over the internet to pre-sold customers but we need a market to actually sell well. This is not wordplay. It means you and I can write a book, an e-book, an article, an online newspaper, a newsletter, a play, a musical score or a computer program and offer it for sale on the internet. Only by creating a market for this information will we sell well and only by creating a vibrant market will we sell enough to make a living wage. To create a market we must build interest, collect attention, affirm quality and in general, do this at a scale large enough to attract the prominence so necessary for internet ventures these days. Unfortunately, once we accept the need for a market, we must feed that market. The market and its attendant prominence becomes an asset. We should earn a return on assets.

Yes, we cannot sell a report without a market yet we cannot maintain a market without raising prices. Perhaps selling was always more involved than dumping text on a page.

This must change eventually. Theoretically, the most efficient way to sell the information found in a book is not by inviting us to read a page or two, then printing it on paper and posting it to us. There will eventually be a digital solution. Direct internet commerce of information will eventually become practical and then dominate. But this could take decades. Until that time, broadly speaking, everything that employs highly paid editors, authors and indexers has difficulty online. An established book author cannot abandon print and publish an extensive website instead. It simply does not pay well. An expensive article archive cannot join us online. It simply does not pay well. Not yet. Not through direct sales. Until recently, so little money exchanged hands for commercial information that we could talk of the internet as a demonetized zone; as a region without money.

ADVERTISING

Perhaps we are going about this all wrong. Books cost money but television programming is free to the consumer. The price of a magazine is negligible thanks to advertising revenue. Many information products only succeed through an advertising model. Could this work online?

Advertising works if we have advertisers. It can work. It does work. Most significantly, a market for attention has developed in programs like Google's AdSense that helps connect publishers to advertisers. However, let us be clear that an advertiser-funded approach only suits certain types of information. In particular:

1_ content quality is secondary to traffic,
2_ small publishers are disadvantaged,
3_ the pursuit of traffic draws us towards popular topics
4_ and advertising pays only for inexpensive content.

Let us review these each in turn. Firstly, in advertising, quality is a secondary concern. Oh, it is important but not to the degree of the utopian publisher asking visitors for donations or the commercial publisher asking for payment to read a page or two. Such publishers have only their quality and value to sell. Advertisers are interested first and foremost in bulk traffic and attention. We sell them not the contents of a page but the attention of the people who come to read.

Secondly, advertisers rarely work with small publishers. Part of this is simply the time it takes to monitor and audit each advertisement. Small to medium sized publishers also have higher costs to secure advertisements. They have no economy of scale in selling. Further, the internet has a technological hurdle to monitor and audit who visits a website and to divide traffic by region. We can surmount these challenges by working with an advertising aggregator like Google's AdSense/AdWords and the Yahoo! Publisher Network but such solutions have costs as well.

Thirdly, popular topics are more likely to attract enough advertising to reach financial health. Popular topics have larger audiences, gather more attention and appeal to a wider range of advertisers. This is not to say advertising works only for popular topics. Magazines attend to many niche markets as well. To state it precisely, advertising supports publishing in accordance with its perceived value to advertisers. This circular argument basically repeats that advertising works if we have advertisers. However, advertisers like prominent sites. They like traffic. They like emotive imagery and educated cashed-up local viewers. They do not offer much for the generic attention of visitors who whiz through one page on the way to another.

Lastly, there is a ceiling to what advertising supports. There is some very expensive information out there and always an opportunity to add more expense through editing, graphic art and talent. However, advertisers don't buy information. They buy attention – and that is worth less. In a real world example, consider the differences between magazines and journals. Magazines maximize advertiser value. Journals maximize content value. We rarely mistake one for the other. Advertising supports magazine content but not journal content. Advertising on the internet will support magazine-like content but not journal-like content.

But wait. Any site with enough willing advertisers is a success so why is this not simply a failure of publishers finding the right advertisers?

Some succeed. How many and how much success is both hard to determine and too soon to say. Certainly, market makers will be richly rewarded but how much of this will reward the information publisher is still uncertain. I certainly watch closely. The shape and flavour of the future internet pivots on this point.

Despite great uncertainty, let me voice my own expectations. The internet will continue to grow. It will grow much much more as we will shortly discuss. I know that in the past, wild enthusiasm for internet banner ads preceded a depressing crash in their value. I know publishers cannot simply add more advertisements to existing pages or break webpages into smaller and smaller pieces to raise revenue. We already see a limit in newspapers and magazines. I also know advertisers are not dumb. Eventually they will advertise only where they perceive value.

Personally, I think advertising is not a solution for everything. We have two models to consider. Firstly, the magazine publisher delivers topical articles, has trouble with focus and bias but attends to any niche market with enough advertisers. Secondly, the TV station delivers popular programs, has trouble with quality and presents only those shows with a high visitor to cost ratio. We do not get opera on our advertising-funded television channel. We do not get magazines on world poverty. Both models favour certain kinds of information.

Advertising is the most exciting roller coaster ride on the internet these days. It works. It does not work. It works once again. The mist will not clear for a long time. Certainly a vast crowd of experts and publishers intend to make it work; they willingly invest now in the hopes it soon will. We also see a landscape of failed projects; abandoned projects. Even

should advertising work, we will still see many failed relics plastered with advertising, life seeping out of them. We will also see a very big market for promotion services but that is another story.

Advertising plastered information has one final difficulty we must mention: a tension between supporting advertisers and telling the truth. This leads to a caution against controversy, a bias towards the wishes of advertisers and an over-willingness to believe commercial marketing copy. We see this enough in business magazines to recognize its influence.

SALES LITERATURE

We have mentioned direct sale of information and advertising supported information. The third and the most successful commercial approach simply involves publishing sales literature. Come. Read. In words of profuse enthusiasm, let me tell you why you simply must buy my latest three-in-one bug zapper.

Many a business strives to create that perfect visitor experience resplendent in graphic wit and convenience. In one of the great successes of the internet, sales literature perfectly motivates the publisher. Yes, most sales literature is hopelessly biased and reads like a travel brochure or worse – very fluffy and full of friendly atmosphere but without any balance, perspective and often without facts. However, sales literature can include items of value. As business use of the internet matures, I pray more businesses will see the wisdom in not treating visitors like brainless children. Will we see more refined publications that demonstrate rather than proclaim expertise? In this I am thinking of documents like the BankWest Quarterly Review of the Western Australian Economy – a high quality document full of economist commentary that deeply impressed me years ago. I also like public speakers who demonstrate their expertise instead of brag about it. Quality is possible though not common under this motivation.

Of course, sales literature first and foremost addresses something for sale. There is always an 'angle'. Content is always biased by the designs of a business marketing plan. Sales literature enjoys an uncommon amount of promotion on the internet too. Anything with a URL ending in **.com** and having no directory or filename – obviously a link to a business homepage

– strongly hints of sales literature. Internet advertisements almost always lead to sales literature.

The real clue to recognizing sales literature is our past experience. Almost all sales literature exists prior to or outside of the internet. Sales brochures become brochureware: websites that feel like a static brochure. Online annual reports gush at corporate successes just as printed annual reports have gushed for years. Internet catalogues mimic mail-order catalogues with just a dash of programming mixed in. The BankWest Quarterly Review had a previous incarnation as a newsletter/report sent to valued customers. All this sales literature is merely literature converted to the internet environment. Sales literature, you see, lives in a very mature publishing environment. Everything has been done before.

This is not to say sales literature cannot change. Companies seeding discussion groups with glowing reports of their products use a novel approach to marketing. It can be difficult to reveal these messages as sales literature though I see little of this yet and what I do see tends to be done badly. I wait with dismay for the inevitable onslaught of internet advertorials.

Keep in mind, none of the three commercial publishing models – direct sales, advertising supported and sales literature – produce the best quality information or focus on obscure topics. Each has subtle differences but none resound in quality without bias. For this, we need the third and final publishing model.

THE ACADEMIC PUBLISHING MODEL

Why would a talented researcher devote a whole weekend to carefully prepare a lengthy article only to email it to a distant journal in Canada never to see it again? The researcher receives no payment for the article. The researcher may even be obliged to remain silent for several months about the research involved.

Why write? Because the researcher gathers a reward from another organization not involved in this transaction. Put simply, a university, government agency, association or sponsor pays the researcher to publish.

Universities entice professors to publish. It is part of their job. Peer-review publishing is tied to career promotion and peer esteem. Publishing

also helps secure well-paid research projects. As a research director friend of mine explained in regards to the National Health & Medical Research Council (NHMRC), an Australian government research funding body, "Don't bother applying unless you have a couple pages of peer-reviewed articles to your name." A couple pages of citations, that is. Research funding goes to the very well published.

As a second example, why would the government agency, Family and Children's Services WA, upon compiling a detailed survey on family structure and habits in Western Australia, then publish the complete document on the internet for free?

Why? Because that is what they do! That government agency studies, then shares with constituents, peers and legislators the information important to doing their job. Internet publishing helps achieve this – and at less expense than other methods.

Third example: Why would an association of uranium mining and refining companies create a pivotal website highlighting the role of uranium in energy production, medicine, exports and security?

Why? Because the purpose of the association is to build awareness of its industry and promote the values of its association members. It achieves this through internet publishing and often with little expense using documents already close at hand.

That the university, government agency and association choose to publish on the internet in such a way that you and I can read is almost a complete afterthought. Neither the professor nor the government agency cares if you and I read their work. They publish for their peers, for their constituents and for their citizens. We may be none of these, yet given the international nature of the internet, we are of course welcome to drop by and read.

The hallmark of this publishing model is that the author has an outside organization that pays for publication. It is not volunteer work since academics and government employees have salaries. It is not a commercial exchange since we see negligible advertising and few requests for money. These are paid professionals doing their job and probably not even seeking our attention. Internet users are simply invited to look over their shoulders if they wish.

Library catalogues, pivotal internet databases like PubMed and ERIC, group projects like the Librarians Index to the Internet (LII) and internet

services like AskNow! virtual reference, for all these incredibly valuable resources, someone else pays the bill. Make no mistake, this slice of the pie generates the highest quality of information on the internet. All it requires is an organization willing to fund an internet project.

And it works like a dream. PubMed, the public database of current medical research, has had a tremendous influence on educating and empowering citizens with health information. It plays a significant role in reversing an earlier practice where doctors told patients about available treatments. In an earlier time, as the Medline database, it was expensive and had far less impact on public health.[45]

We shall call this the Academic Publishing Model though it equally applies to government and association publishing. This model persists well over time since it is tied to an organization instead of an individual. The quality expectations of the sponsoring organizations apply so we can expect better quality. In some settings, information will attract good peer input. We are not usually the intended audience so bias is often less significant to us and usually dealt with more explicitly. Academic and government circles generally have very little tolerance for bias.

Unfortunately, universities, government agencies and associations do not offer most of us a paid salary to write what we love. The rest of us poor saps simply find ourselves unable to participate on these terms.

We spot information produced under the academic publishing model first and foremost by the web address: **.gov**, **.edu** and perhaps **.org** or **.asn** (signifying associations in countries like Australia). This information also displays an unusual depth of understanding and a factual approach to writing. The work tends to use longer formats like research reports and detailed studies.

Associations usually inject more bias. They may have more of a commercial intent. They certainly have a political intent so we should always peek at the association membership. Occasionally, associations will not draw attention to their membership. More often, we may not notice the association publishing a page we visit. As discussed earlier, just notice the publisher and if an association, notice their purpose and bias.

Government agencies also have occasional troubles with political bias that may only come to light when we synthesize conclusions. Publishing under this motivation is not free of bias, just less severely affected.

There are limits to this motivation, especially around commercially relevant applications or information of marketing or strategic concern. We often find student discussion forums behind closed gateways and we may not be invited to peruse past discussion in prominent astronomical mailing lists. However, many universities think in the long term and welcome the appearance of fostering an active intellectual environment. Many government agencies foster open publishing to a fault. Openness depends on the organization involved.

THREE PUBLISHING MODELS

An author writes. How the author justifies their efforts determines, in a sense, some of the qualities of their information. As we have gradually progressed through this book, I have guided you to notice a halo of supportive detail surrounding all internet information. This halo includes nearby information like local context and endorsements. It includes how the information was prepared; its format. It includes the identity of the author and publisher, their bias and experience.

We use this halo of supportive detail in quality assessment. We use it to predict the kinds of information we want to find. We use it to anticipate the information we are about to encounter so as to discard information of little value, quickly. In Chapter Nine, we will revisit the elevated vista and this halo of surrounding information will help us ask better questions. It will tell us if we are listening to a balanced discussion.

This halo includes one more significant attribute. It includes the publisher's motivation. Motivation divides the internet roughly in three equal portions, each portion distinct from the others.

Seeing motivation is not difficult. The URL often suggests one or another. A simple look at a page will usually tell us directly. Commercial sources tend to scream advertising, polish and marketing. Academics and governments declare depth, credentials and institutional weight. Utopian projects proclaim enthusiasm, passion and perhaps group participation.

1_ No one pays (utopian approach)
2_ Someone pays (commercial approach)
    2a_ Consumer pays
    2b_ Advertisers pay

2c_ Business pays

3_ Someone else pays   (academic approach)

Each of these three models, these three motivations, populates the internet with different kinds of information. All jumbled together in a big pile, everything appears fairly messy. However, look carefully. Discern the differences. Motivation is one more element of the halo of supportive detail. In practice, as I search I ask myself which motivation I seek. I glance at an address or webpage and notice the motivation involved. I let this percolate and participate in my process of becoming Internet Informed.

Beyond its role in helping us search more effectively, publishing models open other doors as well. Motivation reveals a social aspect to internet publishing that often goes unnoticed. It reveals the many challenges and limitations confronting authors and publishers as they desire to make our acquaintance. Let us explore three revelations brought on by this enriched empathy for the publisher:

1_ publishing motivation drives internet history,

2_ systems of communication compete

3_ and internet development has been predestined by the nature of the technology that supports it.

A HISTORY

The early internet had few publishers and little information online. It was a time of computer techies with everything accomplished from the command line. I remember teaching this from the state library and oh, it was hard to convince attendees not steeped in a hacker's ethic that this internet would change the world. Few could understand why publishers would bother.

The web had not yet arrived. It was a time of pure utopian publishing. Everything was volunteered or paid from off the internet since the academic publishing model had not yet split from the utopian approach.

In May 1996, shortly after the web arrived, I undertook a rare census of all the significant resources in my home state. I repeated the census three months later.[46] These two surveys reveal little impact from commercial

publishing. It illuminates a quality landscape dominated by a few brilliant resources and little else.

This early internet was also relatively well organized. Well, there was not much of brilliance on each topic so a single volunteer could create and maintain a near definitive list of the best resources. This process began well before the web in the shape of research guides and FAQs. Research guides were grouped together in the Clearinghouse Project, later to become the Argus Clearinghouse Project, then just Argus. Usenet FAQs, back when all FAQs were long and detailed, were always found in one of the many mirror sites of the famous Internet FAQ Archives.

The other advantage of publishing in an undeveloped environment was that promotion was simple and effective. Write an FAQ, have it accepted into the FAQ archives and everyone interested would find it. Create an interesting significant resource, have it listed in the right guidebook, or later in the right Yahoo category, and again, everyone interested would notice.

Time passed. The guidebooks and FAQs, so very significant before the web arrived, gave ground to Yahoo and significant search engines like Webcrawler and the World Wide Web Worm. More resources were online now, though not so very many that an individual could not find all of the most important.

A very helpful way to find information at this time involved simply hunting for someone who had recently undertaken a comprehensive search. A web search like **accounting hot links** or a year later, **accounting further resources** would reveal these pages.

This was still a utopian time. The first publishers with commercial intent were chastised and ignored. Spam was not yet a problem and 'newbie' was the word du jour. America Online joined the internet and the flash flood of newbies received a poor reception from more experienced internet users.

Time passed. Early search engines based on frequency ranking were having difficulties with the increasing amount of information in their databases. They no longer worked so well. Page five results were as good as page one. The best search advice I had to offer was always place the plus (**+**) symbol before each search word. Always search in a precise manner. The Yahoo Directory stepped in with a little advertising and

quickly became synonymous with searching the internet. We could not reach for Yahoo fast enough.

Also at about this time, utopian publishing began to give ground to commercial publishing. A Yahoo Directory listing became very important for promotion. Popular internet marketing advice included asking Yahoo for a listing every six weeks until it acquiesced. Listing in Yahoo first slowed in response, then Yahoo asked for a fee for consideration. The fee was worth it.

Time passed. More and more information emerged on the internet. The early publishing war was won in that information utopian believers thought should be released online was now sure to reach the internet community without intervention. We sat back and celebrated as each day saw new digital libraries and extensive websites blossom.

Equal credit for that victory goes to simpler publishing tools. Earlier geometric growth in internet users also fed, about a year later, a rapid growth in internet users capable of publishing.

The next significant improvement was the injection of popularity in meta-search engines. Google's popularity ranking was better, so about a six months later I was advising everyone to ditch the meta-search engines for Google. This advice still applies today.

With this increased pace of publishing, Yahoo began to falter. Well, it certainly looked like their directory would not keep pace with the continued growth of the internet, though it held on far longer than I expected.

Also, with the growth in volume of information, web traffic began to get watered down. Where as before web publishers would see only growth, now many publishers saw monthly traffic rates level off. Compounding this was a rather dramatic fall in the value of banner advertising. Many a publisher who had earlier been content to volunteer their time expecting a payoff sometime in the future finally had to confront the fact that neither banner advertising nor unending growth in traffic would lead to a financial reward.

The struggles of the commercial publishing model coincided with the utopian struggles with promotion. A new website would require serious effort in marketing and promotion, so in theory, older websites had a goodwill value – though it was near impossible to capitalize on it. This did limit new publishers, though. Unless a publisher was first to offer content on a new topic, new ventures would always start ranked too low to attract

considerable attention through the search engines. For the first time, utopian publishers would have to claw their way up from relative anonymity, irrespective of the quality of their content. This new internet was not looking nearly as promising as before.

It turns out publishing involves more than putting an article online. A publisher must also be financier, editor, marketer and promoter. In early days, these tasks were simply ignored. Promotion has since grown more challenging and promises to be paramount in the future. In a sense, the internet moves towards standards more common to non-internet formats.

A first-time book author earns perhaps 6% royalties: that is A$3.60 for a book like this. The remaining 94% goes to publisher, printer, distributor and bookseller. Most of the remaining A$56 we would class as a financing, marketing and promotional expense. Internet publishing is incredibly streamlined in comparison but we are moving in this direction.

As the internet adopts a more mature structure more in line with other competitive publishing environments, one of the basic tenets of the internet utopia is lost. Internet publishing ceases to be a meritocracy. Oh, successful projects most certainly are rewarded for quality content but they achieve success through a combination of unique content, fortuitous promotion and skillful marketing. How strange to think it was ever otherwise.

A LIKELY FUTURE

Where does this journey lead? I have several personal expectations.

1_ Promotion will grow further in importance while the number of people capable of publishing will grow as well.

This claim should be obvious. Look around us today, not at the internet but at our everyday lives. We swim through a vast sea of information we either ignore or avoid. The internet still presents only a fraction of the information impacting us. As we digitize our lives over the next decade, of course content and competition will grow. As more of us become familiar with the internet, of course more people will become capable publishers themselves, adding more to the internet. This trend will continue until forces come into play to restrict or dissuade publishing.

2_ Internet search engines will index less of the internet.

Speaking of search engine coverage is always fairly contentious and I must sadly agree with anyone who claims I am quoting poor statistics. Unfortunately, there are few statistics worthy of our trust, yet a pervasive attitude exists that search engines like Yahoo and Google index almost everything on the internet.

As I have stated, I suspect Google indexes between 10% and 20% of the internet. No great study supports this claim but it troubles me that Google grew at 49% per annum between 2002 and 2003, then just 7% between 2003 and 2004. Then, with the Google's jump to eight billion records on October 2004, it grew 144% in 2004. In late 2005, Google jumped to somewhere beyond twenty billion, perhaps. This twenty billion is more questionable than the previous numbers. We do not know if this number has changed since.

If sustained, Google's leaping database size bodes well for coverage in the future. However, from 2002 to 2005, Google's database still grew a not-very-impressive 57% per year.

Estimates of the size of the internet are truly all over the place. Even at the turn of the century, size estimates of 10 billion to 300 billion, growing at 100% to 300% per year were suggested with a straight face. I discussed this issue further in articles at **SpireProject.com/art10.htm** and **art13.htm**. Keep in mind, growing databases must at least keep pace with the growth of the internet and race faster than the pace of the internet's growing population.

Will the internet be fully indexed? It depends on:

1_ what we call 'being on the internet',
2_ continuing the possible trend of growing database sizes,
3_ recognizing how very much information exists in our world
4_ and a business model that in all fairness, does not really improve with greater coverage.

Just on this last point, search engines already index the most prominent 10% to 20% of the internet. Blunt searches (and most internet users do little else) rarely reach beyond the most prominent portion of the internet so growing a database an extra ten billion records offers little search advantage unless we learn to be specific. Besides, do we really want old webpages, anonymous webpages and unrecognized webpages or do we

want just the portion of the internet blessed with traffic, including much of what is in constant development?

I do not think search engines will index the whole internet, though in all fairness, Yahoo and Google certainly index more than I expected.[47] As this chapter proceeds, the reason for my certainty will become clearer. So will an understanding that once we accept search engines as incomplete, the degree of incompleteness is not so very significant – though this is perhaps a conclusion to discuss on another occasion.

3_ Competition becomes fierce. Attention jumps in value.

Continued growth of information content and a persistence of prominence may lead to sudden shifts in the commercially value of prominence. It already has value but this value has been depressed because the commercial publishing model is still so fragile. If ever a commercial model proves successful, then prominent websites jump rather suddenly in value. In a short time, internet publishing would enter the realm of big business.

This is not to say small individual publishers cannot publish. They already do and often successfully. However, if prominence becomes suddenly valuable, small publishers on prominent sites would sit upon significant assets. They would not be 'small' anymore. Already many of the successful small internet projects cannot be duplicated from scratch since the starting costs to capture our attention and perhaps participation would bankrupt a small business. This future depends primarily on the success of internet advertising.

If, for some reason, commercial publishing is still struggling five years from now, the mere expectation of eventual commercial success may yet raise the value of attention. We are not always led by reality. Often the perception of reality is enough.

Should commercial models prove totally uninspiring, it will probably occur because of a change in the acting definition of prominence. We do not often appreciate how delicate the internet environment pivots on the acting definition of prominence. Redefine prominence in some manner that better captures merit and the playing field suffers an earthquake.

Probably, all of this will happen. Business will succeed, struggle and fail according to the topic and discipline. Business will populate the valuable

niche markets, abandon those that offer little reward and get stung by sudden swings in internet property values.

4_ The internet will continue to grow until it is truly enormous.

For this last prediction, I shall reach for the words of Reverend Thomas Malthus (1766 – 1834) and his ideas on population growth. Population, he suggested, will grow to consume all the available food thereby inviting recurrent famine. Only war, famine, disease and moral restraint would limit population growth.

Twisting his words to apply to the internet:

> Internet information will grow to consume all available attention, to the point where new publications court anonymity; to the point where no one cares if a new item of information is published.

Read this statement carefully. I do not predict a state of confusion or a collapse of pay-per-view advertising. Merely the fundamental nature of the internet is to grow as long as attention exists for a new page. If putting new information online is valuable, then more information comes online. After all, it costs so little to put information online, so many people can put information online and so much information surrounds us that could be placed online. The transformation of internet publishing into chiefly a struggle for attention is well underway.

It may seem a strange observation but the internet offers no limit to publishing except anonymity – anonymity and searcher boredom. In time, putting online per se will add no more value than printing a document on paper and stacking it by a street corner. At the same time, the promotional imperative on the internet rises ever upward towards a point where internet information is simply invisible without it. We are familiar with this in other publishing environments. A book without promotion is unknown and unread. Book content became subservient to promotion and marketing long ago.

Perhaps I am just uncomfortable with the idea of an internet that becomes one continuous advertisement. It occurs to me our internet will come to look very different to the original meritocracy we started with. On its current path, I wonder if I will like the internet in five years time.

An enormous internet offers special difficulties for those hopeful of selling advertising. We can shatter webpages into little pieces and plaster

on numerous advertisements – already evident with online newspapers and article sites like **WIRED.com**. We can theoretically have four or five times more advertisements than pages on the internet. How will this abundance affect the commercial advertising model?

An enormous internet offers even more difficulties for searching. How do we find information not indexed, not searchable and nearly invisible on the internet? This is actually a very revealing question we will address in Chapter Nine.

The missing puzzle piece to this discussion rests in understanding the more mature communications systems we currently use instead of the internet. For a second look at a different aspect of this evolution, read the cover article I wrote for an issue of the trade magazine ONLINE titled: Evolution of Internet Research: Shifting Allegiances.[48] The text for this article resides at **Spire Project.com/art19.htm**

SYSTEMS OF COMMUNICATION

The three publishing models find themselves frequently in conflict with one another. From the utopian perspective, commercial publishers piggyback on a system of communication established on donated information. How dare they corrupt this pristine giving environment!

From a commercial perspective, the internet misappropriated a great deal of enthusiasm due to a transient veneer of meritocracy. Yes, the internet provides a freedom to publish – a freedom to blog. Serious public communication has always been (and should always be?) the domain of big business. How audacious of the utopianists to challenge this! And while on this topic, would they please stop giving information away for free. It ruins the market.

The academic model is more aloof of this disagreement but earns enmity from both sides since scholars are paid to publish what utopianists must volunteer and commercial publishers must purchase.

This squabble, however serious, pales when compared to the conflicts between 'systems of communication'. Look most widely for a moment. The internet, whatever a publisher's motivation, exists as just one chain of communication. The chain looks like this:

Author → Internet → Reader

Information flows from the author/publisher, through the internet to us as readers assisted by only a few organizations: our ISP (internet service provider), perhaps a search engine and anyone who offers us guidance along the way. Let us call this a system of communication.

Another system exists based on journals. An academic writes. A journal peer reviews, edits then publishes. A database indexes these articles. A library purchases both the journal and database so we can search, retrieve and read these articles.

Author → Journal → Database → Library → Reader

Another system encompasses print news media. A distant journalist submits a news article to a newswire. A local newspaper picks it up and prints it. For a dollar, we buy a copy from a news agency.

Journalist → Newswire → Local Paper → News Agency → Consumer

Books, magazines, TV, public speakers and a few others, each system encompasses a collection of participants, a style of presentation which we earlier described as format and perhaps a few variations in how the information reaches the consumer.

Collectively, these vast systems infuse our world with a varied and multi-coloured panorama of communication. This is our environment – our information ecosystem. From an elevated vista we see a world not of objects but of vectors, of friction, of ideals and especially of competition – competition between messages, between motivations, between systems. The journal article competes with the book and public speaker both for the author's time and the reader's attention. This multi-system world is our home.

Is the internet really just a collection of webpages? There is more involved. The internet captures our imagination and toys with our emotions. It chaperones our messages and in doing so, showers blessings on some businesses but not others; on internet service providers (ISPs) but not booksellers; on search engines but not database retailers. The internet competes with the other unruly curs, then tries to convince us of its inevitable supremacy. The internet promises the future and I dare say, sells itself well.

As usual, history offers a slightly different perspective. Yes, a change is in the wind but perhaps this wind blows like a gusty breeze, not a wind-

storm. Let us pause to consider the strengths and history of other systems of communication.

Historically, a scientist like Charles Darwin wishing to share his theory of evolution with his peers had few options other than writing a book or presenting it in person.

This changed in the 1900s with the advent of peer-reviewed journals. Journals, with their desirable mix of high quality and depth, retain a specific focus but demand far less effort than publishing a book. However, journals do not serve as a good way to organize and retrieve information. Without the bibliographical databases that emerged after World War II, a scholar would essentially read everything relevant to their field, then rely on their own personal library and filing cabinet (or association library as these developed) to piece together a puzzle. As the volume of information and research grows, this approach fails.

I enjoyed an elegant demonstration of the power of bibliographical databases about four years ago while attending a lecture on Egyptology. Dr Rene van Walsem of the University of Leiden, Netherlands, described how he and his team excavated an unknown tomb near the Egyptian Step Pyramid of Saqqara.[49] During the excavation they uncovered a beautiful intact statue of a loving couple named Meryneith and Aniuia, their names recorded on the base of the statue. On the day of discovery, with great excitement, Dr Van Walsem consulted the Annual Egyptological Bibliography (AEB) database to learn of any other work by Egyptologists that referred to Meryneith. With this search he confirmed Meryneith as a historical figure worthy of a tomb in the Saqqara city of the dead. Meryneith's tomb had been found.

The Annual Egyptological Bibliography indexes scholarly work stretching back to the year 1947. It is a key resource in the field of Egyptology. And it assisted Dr Van Walsem to find information pivotal to his research on the day of discovery! Early last century, an Egyptologist would personally have to page through the many prominent journals to find information on Meryneith. In such a situation, research becomes a proactive study of all information about Saqqara. The alternative involved publishing an article about Meryneith in the hope other scholars with relevant details would read and initiate contact – both steps labourious and time consuming.

Journals display great power when indexed by databases – a power built on peer vetting, ease of use, a fair speed and avoiding the need to read everything that might be relevant.

As an aside, I find it delicious that the internet revolution was preceded by the bibliographical revolution and the journal revolution. From a certain perspective, changes today look more like evolution than revolution.

SYSTEMS OF REIMBURSEMENT

Unfortunately, communication by journal article is actually a very expensive process. The author receives a fair wage if working under an academic publishing model. The journal publisher earns subscription rates. The public library is publicly funded to purchase and offer patrons these journals. The database owner earns money too, yet readers pay only a small fraction of this exchange. Notice how everyone is paid for their help and a great deal of effort is made to refine, improve and organize this information. The true brilliance of the journal system of communication is not only in how well it assists researchers to work together. Journals also attract hidden subsidies very efficiently. For instance, without the assistance of a library, the expense of a journal and database subscription would make this information prohibitive.

Books are expensive too. Bibliographical databases charge the publisher. Book reviews are advertising funded but most costs eventually land on the consumer. Authors are often very badly paid. I find it a little sad that the person who typesets a manuscript and the person who crafts the index often earns as much as the author. They certainly earn a better wage. Books, however, are very efficient at attracting advantages for their authors; advantages in respect, fame and the illusion of payment. An expert without a book is not an accomplished expert. By illusion, I mean the J.K.Rowling Syndrome: the rather pervasive and unfounded hope a book will become a wild financial success.

Magazines are very efficient at attracting advertising to subsidize content. Magazine content also benefits from library involvement and database indexing. For example, the Australian Interlibrary Loan and Document Delivery Benchmarking Study[50] quantified the average 2001 cost for requesting a document at A$32.10 (A$18.98 for public libraries). Costs charged to patrons are significantly below this. In response, some

universities simply ban undergraduate students from requesting articles not held on-campus.

Yes, our community certainly subsidizes our efforts to be informed – except on the internet. The internet excels at speed and access but not yet at vetting procedures or attracting subsidies as with journals, nor bestowing fame and expert status as with books, nor attracting advertising as with magazines, nor in personal contact and inspiration as with public speaking. The internet is rather clumsy at all these things. One possible future simply sees each of these systems working to their strengths.

Will the internet push aside these other communication systems? I do not foresee internet advertisements becoming more captivating than the full colour glossy advertisements of magazines. I do not foresee internet information becoming more personally inspiring than public speaking. The internet can duplicate most of the vetting and organization of a journal but until such vetting, editing and organization is funded, this is theoretical, not real. Journals are created in a moneyed environment. I have not heard of many moneyed internet journals, only journals we can read online.

Will the internet eventually attract substantial subsidies? Clearly, yes. Libraries already purchase databases for the use of internet patrons. Free internet databases like the Library of Congress Online Catalog (LOCOC), the Catalog of US Government Publications (CGP) and ERIC (Education Resources Information Center) excel brilliantly. Most internet projects, however, still succeed or fail on their own and attract little or no grant funding. Until recently, even academics earned little recognition for internet activities though certain internet journals now do count towards peer recognition and career advancement.

Subsidies will shift in the future from non-internet environments to the internet. If we want to promote public health do we advertise on TV or build a definitive health website? If we wish to promote economic development, do we print a directory of regional suppliers or create a free state-wide suppliers database on the internet. Answers to such questions will gradually tilt towards the internet in time. The most persuasive reason I see for this is that the current lack of subsidies and public funding of internet projects does not emerge from the internet technology itself. In the words of Dale Spender in "The Last of the Print Proficient":

"… when we want to look more closely at what is happening today, it is helpful to distinguish between those patterns of behaviour which are associated with the process of change, from the more specific changes that are the product of the medium itself."[51]

I would suggest this famine of financial support stems from the process of change, not some facet of internet technology. In a few years, we shall see much more funding of internet projects for our collective benefit.

Unfortunately, I see three downsides to this noble use of internet technology. Firstly, little funding will probably go to existing utopian or commercial projects where it would do the most short-term good but would be raise issues of favouritism. We have no difficulty in assisting commercial journals or buying pricey database services but the public purse will probably not support the Wikipedia or some commercial health site. Projects undertaken by academic, government or non-governmental organizations (NGOs) will attract the bulk of public funding.

The second downside involves our apparent march towards the liberation of talent from the organization – something due to occur gradually over the next two decades. When associations, businesses and especially universities and government agencies have a much better claim to the funding that becomes available, the pace of this kind of liberation slows.

This liberation of talent from the organization was greatly boosted by the development of the web. Individuals could stand and compete successfully with organizations in delivering outcomes on the internet. Indeed, in earlier days, individual talent had the advantage since swift movement and a lack of institutional inertia usually won the day. However, this dynamic is now reversed. The organization has a better claim to external funding that makes the academic publishing model possible.

Do organizations deserve this advantage? Yes and no. Organizations are better equipped to manage diverse perspectives and viewpoints. They are better at bridging academic input into practice and in providing a persistent resource. They have a history and experience at the best of times. However, existing experience suggests these advantages are at least partially mitigated by the group management and peer participation found in the better utopian projects. Large groups of people on the internet can bond together and forge a meaningful resource even without

external funding! Mailing lists, discussion boards and webbed resources manage everything an organization does without having to ask for payment. Imagine what could be accomplished with funding.

The unfunded internet projects of today will gradually give way to the funded projects of the future. Many of the earliest internet activists naively felt this transition would benefit them; that this transition would be well underway by now. Projects begun in the early days of the internet could cash in as the foundation for much larger projects much as the small tech company gets swallowed by a large corporation. This can still happen but probably will not.

The third downside to public funding is how little will be directed towards organizing the internet. With the significant exceptions of librarians offering internet assistance and the free commercial-quality databases, I do not see many opportunities to expand search assistance. Other systems of communication extensively assist and subsidize how readers find information of value. The internet largely leaves these services unpaid and unaccomplished. To put this crudely, we can organize the web much better than Yahoo and Google have but we would have to pay for it. The grand achievement of Yahoo and Google is that they offer so very much for the very little we pay in attention.

We can highlight this view of the internet by focusing more closely on the internet transaction. Look superficially at the internet and we seem to see information flows directly from author to reader. The internet shorts out the exchange of information like a fork in a toaster. Zap. The author publishes to their website. We find it, then print it on our printer for two cents a page.

$$\text{Author} \rightarrow \text{Reader}$$

Unfortunately, this idealized view hides the costs even more subtly than in our earlier example. Firstly, we pay the professor to do the research. This cost remains. Or perhaps more accurately, this cost remains unpaid since publishing on the web by professors is unlikely to be rewarded financially. It is probably unrecognized as worthy of tenure. Next, the vetting and organization of information as previously performed by the journal and commercial-quality database has been dropped. The library, with its great mandate to deliver assistance and provide tools to assist researchers is also withdrawn. Next, the internet's infrastructure

(the computer and internet connection) is assumed to be omnipresent and free: a view certainly not applying in the developing world but a fair assumption for everyone else. Lastly, the organizing of the internet is undertaken by someone who may perhaps have volunteered their time or may be paid. If paid, then most likely they are paid by advertising or by the publisher directly – and we have already discussed the limitations of these approaches.

Thus the naïve picture of just an author, a researcher and two cents a page dissolves into something far less enticing: an author, a researcher, two cents a page and none of the assistance of journal editors, librarians, no royalty payments and only unpaid or indirectly paid assistance with organization.

This communication is more like the telephone than publishing. Books, magazines, journals and news all fund promotion, editing, vetting and organization. Well, news does not fund the organizing of old news but you get the idea. The internet does none of this. Magazines, journals and the occasional book attracts significant subsidies for these tasks. The internet does not.

FOREORDAINED

Notice the tragic beauty to this layered perception of the internet's recent past and near future. So much of this story directly emerges from the nature of the internet itself. The commercial difficulties. The growing role of prominence. The three distinct publishing models. Everything. Perhaps we were even destined to see too much of a glorious meritocracy in this technology; to celebrate too much the delights of empowered publishing, only to see many of these advances dissipate with time.

Any system with competition for awareness, room for abundant over-supply and a gradual evolution leads exactly to what we have seen.

1_ We start with exuberant enthusiasm and idealism.

2_ Organization then grows difficult. Promotion grows more important.

3_ Organization then lags far behind the quantity of information on the internet.

4_ Eventually, we reclaim some of this disappearing informa-
tion. Years in the future, social priorities and technical
abilities may allow us to organize far greater quantities of
information, though always in dynamic dis-equilibrium
with the volume of information published.

What I am suggesting is that the internet world has developed along a predictable path from utopian idealism to a more mature realm concerned with promotion and marketing. The technology that supports the internet held within it not just the promise of the internet but also our reaction to it, our initial utopian enthusiasm.

As an aside, I wonder if this situation occurs every time we shift mediums. A few years from now, as audio then video information joins us on the internet in abundance, will we again have a rare situation where quality begets attention. Then, as the need for promotion reasserts itself, we find once again that quality content becomes insufficient and our meritocracy slips away?

I am a little fatalistic about the quantity of information we will have on the internet. There is simply no attrition of information, no disposal of old newspapers. Digital information just piles up in disused corners of the internet. Furthermore, we are squishing together the vast ideas of a whole humanity into one space, one cyberspace. Of course our cup will overflow.

As a simple and early demonstration of this, consider the experience of the Amsterdam Digital City as described in The Internet Galaxy by Manuel Castells:

> "Data from a log-file analysis over time showed that the ten
> most visited websites accounted for 85% of all hits, while 75%
> of the sites were not visited at all."[52]

Such gross analysis does not do justice to internet communication but these numbers are troubling. Such an early project with such unbalanced attention. Does this reflect our quote by Voltaire? "The earth is covered with people not worth talking to." Perhaps it merely reflects our habit of always reaching for the popular and prominent, disregarding almost everything else.

Today I have a different vision of the future. The internet is organized in three distinct layers.

1_ The first contains a highly structured layer of prominent, popular sites we can easily find. These sites enjoy the opportunity of advertising revenue and serve as directories and nexus points for ...

2_ A second layer of valuable but disorganized information, articles, reports, thoughtful essays and expert commentary, that vie for the attention delivered by the first layer and sit upon ...

3_ A third layer of unorganized information without prominence, much of it unindexed by global search engines or nearly anonymous and therefore difficult to find.

In some topics and disciplines, merit will decide where a document resides – merit and merit-by-proxy. For other topics, promotion will decide where a document resides – promotion and source identity. As a merit-by-proxy system develops finesse, we will more willingly allow ourselves to be led to specific information on the internet. Today we would be mad to ignore quality concerns but in the future, quality will be less an issue if we stay on the top level.

Search tools like the global search engines will index only the first and second layers. They will show us only the first layer unless we craft a specific query. The third layer, like the 75% of the Amsterdam Digital City, will have negligible traffic and require talented search skills to excavate.

Thus we will easily find a directory of libraries, a recipe for blueberry pancakes and the websites to the most prominent law firms in our state. We will not find that recipe for Brazil Nut Cake – the one that does not use wheat flour. Nor will we find most of the messages posted by some stranger we are investigating. Below the prominent sites, below the valuable documents, will sit a vast collection of both disorganized and undigested information. So much information will exist as to make adding to this third layer of no value. Thomas Malthus and his population studies will return to haunt us.

We admire the summit, respect the high mountain meadows and generally overlook the foundations of the mountain. We shake our head and accept that much of a mountain will always remain hidden from view. After all, this picture resembles other information systems....

When Camas magazine (**camasmagazine.com**) won its second prestigious journalism award for ongoing coverage of a city council's financial mishap with a media/real-estate firm, you and I living far away have little opportunity and little expectation of ever encountering this exposé. Even if it is critical to a search we embark upon, we can only trust serendipity to bring it to our attention. A breakthrough in rural electricity production from powdered straw burnt in a modified diesel engine; the latest interpretation of Voyager Spacecraft data and what it says about the current confusion surrounding gravity; these documents and others exist but routinely do not find their audience. They are lost to us like the mountain beneath the forest.

This image, then, becomes the jewel of humanity. Born of competition and surplus, born of volunteers and subsidies, we create a grand smorgasbord of resources beyond par, at the same time as filling the garbage bin in the kitchen with food scraps.

A MISUNDERSTOOD IMPACT

Internet publishing broke time and space. It ripped away the cost of printing information – printing but not publishing. In a sense, it shredded printing costs so suddenly that it created a parallel system of communication – a fiscally barren but potent means to communicate. We have been clawing back ever since making this environment work in ways similar to other systems of communication. We are in the midst of shifting subsidies, improving vetting chains and drawing into the internet the best features of journals, newspapers, books and articles. In this process, many an internet dream becomes road kill.

What is often missed in all the cacophony of chatter and hype is that the internet revolution is about a drop in the value of information.

> "Everyone said: 'The Internet is great because it removes the friction from commerce.' And that is good if you're a customer or an economist looking at business productivity. But it's often the friction in commerce that gives companies their profits."[53]
>
> Nick Carr – Harvard Business Review

Information, whether considered advice by regional experts, pockets of information sequestered in commercial databases, breaking news of

murderous events or commentary about astronomy, travel or Brazil nut cake, information value has collapsed. At the same time as we celebrate how information is so critical to our lives, it becomes more abundant and easier to acquire.

In such an environment, how strange to think businesses would not have difficulties and utopian publishers would not court anonymity.

PURSUIT

The end of the Cathar conflict came quietly. A few minor conflicts persisted for a time and were brutally suppressed. Albert tried his best not to be involved but time and time again, he was dragged into supporting some siege that always ended with burning heretics and economic ruin. The dream of a civilized permissive culture in southern France was smashed beyond repair.

Albert's own personal quest had been flawed from the start, he reasoned. Albert chose to be a knight to preserve the ideals of justice. In the end, he was part of its destruction. This bothered him. Perhaps he had lost his battle years earlier. Events took control in a way he could not redirect once begun. In earlier days, he felt content to move forward; to support peace; to advance in career and fame. By the time he realized his path had gone astray, Albert could no longer see a future he liked.

Yet it was not Albert's mistake. His world simply found the emergence of a liberated, permissive society as too dangerous; too confronting. Southern France had arguably become the most cosmopolitan, renaissance region of Europe but as it advanced, as it began to shake off the staid ideas of the age, its citizens naively mishandled their conflict with Rome.

*   *   *

We once sought information by looking for something relevant AND within reach. Answers crafted by famous but distant authors or archived in splendid but far away venues did not interest us. Such work simply took too long to arrive. Better we gather local resources and leave international opinions to people who live 'over there'.

Oh, a steady stream of the finer items journeyed across the great divides to lodge in nearby libraries and local newspapers. A book here. A magazine there. For the dedicated connoisseur, a copy of Bambini, air freighted in from Italy. If not, let us wait for a local magazine to convey to us the latest trends in international baby couture.

Then Disneyland built their 'It's a Small World' pavilion and began brainwashing all their visitors with a cacophony of repetition. Small wonder that just two decades later, the world is indeed smaller. Change after change smashes holes in the walls separating our world. In a shift more monumental than any I can think of, information now flows everywhere. Even language and computer ability seem set to crumble against the liberating effects of an open information architecture.

There may be other reasons for this change besides Disneyland but change is certainly in motion. Little will separate us from what is out there.

What is out there? Mostly, more of what is nearby. Very few of us create something truly unique and most of us produce something decidedly common. Every region has its experts in tax law, experts in software coding and experts in liberation theology. Only today do these experts rub shoulders. Why not buy Bambini direct from Italy? I already buy the occasional New York Times at a local news agency and read Le Monde in English through the internet.

This abundance of common information, this internationalization of our information world, has a stunning drawback. How do we find information that is not abundant, not found on a thousand websites and not already organized for our consumption by a publisher with prominence? How do we find information not published by an organization we already know?

We seek it. We pursue it. Strangle this unwieldy beast we call the internet and squeeze until it gives us what we want.

Let me pose two questions:

1_ How do we photograph a natural car crash – unarranged, on the street, in public, as it happens?

2_ How do we find the cheapest airplane ticket to Paris?

Internet searching is built on a simple premise that when we ask the right question, we get the answer we seek. Ask the wrong question and we

spend a great deal of time hunting for hints on how to proceed. Look back after a difficult search and success so often pivots on one single inspiration. At one point, we ask the right question and everything proceeds smoothly from there.

This hunting for the right question is not sloppy or unfortunate. It is the essence of our pursuit. To photograph a car crash we must go to where cars crash. Set up a camera near a nasty intersection and wait. Cars can crash anywhere but we want to wait where they crash most often. Let us stand at the first turn of a grand prix and hope for tragedy.

To find the cheapest flight to Paris, we must learn how to find the cheapest flight to Paris. Seek guidance on how to find cheap airplane tickets. The advice I found suggests checking all three primary sources of discount fares: the online travel discounters, travel agents and the ethnic travel agent. Knowing this, our search proceeds smoothly.

It often helps to view our search as more than a search for information. We also search for a question. Our quest pivots on where best to look or how to look as with the two questions I just posed. We may seek one thing but find it by first visiting somewhere else on the way.

We will take this notion in four directions:

1_ Identify where the awareness of a resource will flow – the 'footpath' others have followed to reach it. These places will also lead us where we want to go.

2_ Seek the 'page-next-door' that introduces a project but does not include the text of the document itself.

3_ Hunt through 'the elevated vista' for clues that tell us we are approaching where the information we seek congregates.

4_ Seek search assistance from others who know how to find what we seek.

We will end this chapter with the startling conclusion that the many techniques and tactics we have covered in this book unite to deliver a comprehensive and definitive search – thus answering a question I posed at the beginning of this book.

FOOTPATHS

Footpath is a simple concept. If we publish, how will people find us? Discard, for the moment, the naive hope that people will toss a few choice words at a search engine and find us among the millions of potential sites. Let us focus just on the needs of publishers who do not swim in excess prominence. Focus only on the non-random ways our audience finds us. If visitors usually find us by searching for our name, how do they know our name?

Do they read of us in of a newspaper article? Through a prominent report we publish? Do they arrive at our website by way of a significant directory? A reference site? A database entry?

How others find our site is our footpath. These footpaths guide visitors our way. They direct visitors to the information we provide. This footpath is valuable to us too – it is an aspect of our prominence. Like the owners of a small hotel, we place a sign on a nearby intersection. We pay for a billboard on the approach to town. We place leaflets in the town's tourist centre, we host an annual jazz festival and we support the local junior football team. Each pointer tells strangers we exist. All these pointers establish a footpath to usher visitors our way.

Now view this as a searcher instead of publisher. The publisher we seek is trying to capture our attention. They are trying to reach us. They have placed information in places they think we will visit; in places where others like ourselves have found them in the past.

The hotel owner has a financial incentive to build an extensive footpath. This is marketing and promotion, a task most businesses take very seriously. On the internet, most footpaths are surprisingly small. Many internet publishers still spend little effort on promotion. We know utopian publishers are poorly positioned to deliver effective promotion. Promotion is not what they come online to do. Publishers using the academic publishing model often embark on only the barest promotion aimed only at peers and constituents. Even businesses tend to focus their internet efforts on search engine promotion. This leaves few footpaths for us to follow.

As searchers, however, if we can find this footpath, we have found our destination. Need to stay the night in this small village up ahead? We do not need to find a hotel ourselves. We only need to find their billboard,

the town tourist centre or the village kid who plays in the local football team.

In every other medium, there is great clarity in how readers find an author. To find a book, we search a bookstore, a library or a book database. Perhaps we read a book review. For a book publisher to reach their audience, they simply build onto these established footpaths. Contact bookstores. Sell to libraries. List our books in book databases and encourage positive book reviews.

On the internet, we do not have such clarity. There is a real possibility that readers seeking information and writers seeking an audience will never meet. Such a strange predicament. Authors expect the challenges of competition but this is worse. Readers may get lost on their way. They may not arrive not because they read something else but because they could not find anything worth reading.

As we search, we do not need to find that obscure scientist perfecting a diesel engine that runs on powdered straw. We do not need to find the obscure sociologist studying the distribution of drug profits among street gang members. We do not need to find the obscure doctor sewing plastic retinas on the unseeing eyes of Himalayan poor. We need only find our way to the article in the local farmer's magazine, to the book review in the Sunday insert and to the commercial database entry that will lead us to each of these obscure individuals. This may still be difficult but at least we have more than just the obscure individual in our sights.

So who would report on a scientist working with powdered straw, the sociologist studying drug gangs and the doctor working with plastic retinas? Where would such specialists build their footpath? Equally, what would these introductions say? What terms, what identifiable features, would they have? These are the questions I am urging you to ask.

To reveal the footpath of an existing page, ask for a list of links or endorsements. My **SpireProject.com**, for instance, benefits greatly from my Information Research FAQ, from a popular article I wrote on country profiles, from directory listings in both Yahoo Directory and the Open Directory Project (ODP) and from links found on many library websites. Looking at my traffic statistics, I also see I get many a visitor by way of **CEOExpress.com** and I get the occasional short flood when a university professor directs students my way. This is my footpath.

Once we study a few footpaths, we can better place ourselves in the role of a publisher. Empathize with a publisher. Where would they promote the information we seek? If they promote only on their website, then we must find them there. If they promote through discussion lists, then we have that avenue to explore. If the publisher is skilled enough to earn a listing in the Yahoo Directory, we can find them there. We can aim to find their billboard, their footpath. We can aim for those places where they promote their existence.

Keep in mind that for a great deal of information on the internet, we may not be able to search directly for the information we want. It may be simply unindexed. Even if indexed, it still may be easier to locate through pages on the footpath – pages that often have better prominence or have better reason to catch our attention. It may be easier to find the billboard than the hotel.

VETTING

Beyond where the publisher tries to reach us, we may still find other people lending the publisher a hand with promotion. A familiar task of all internet users is leading others to information we consider valuable. This involves vetting. To vet is to select one item of information among alternatives as particularly worthy of our attention. A link or endorsement is a successful vetting.

Chains of vetting serve to bring the better, more popular information to our attention. As expressed earlier, the journalist reports only on the most captivating stories. The newspaper editor selects only the most significant stories for the front page. We tell our friend only the most interesting stories we read in a newspaper. A story travels only as far as these chains of vetting push it along.

If the information we seek is topical, valuable or significant, we can usually expect that news of that item will travel far; the footpath will extend well beyond the source. If information is of limited value and significance, it will not reach far. It was the initial internet miracle that the footpath of significant resources extended so very far beyond where the publisher personally promoted. We shared and vetted and informed others so very often. The strength of this vetting is not so evident today.

However, vetting is not a mirror of prominence. Different groups of people will discuss and vet a resource very differently. To use vetting trails in our search, we must find a group that values the information we seek. We seek where such information would flow.

We can describe vetting trails another way too. If we want information, find someone who knows of the information. Find someone who already encountered it and will help raise it to our attention.

Certainly, discussion groups serve a wonderful purpose in resource discovery. Years ago I asked for guidance from BusLib-l, a discussion list for business librarians. In the two days after my initial email, I received advice from five people describing six places to access a database I sought.

Vetting trails may also direct us towards information instead of people. The BusLib discussion list archive may assist us. A local directory may guide us to the resources we are looking for. As our example, let me describe my search for US seminar venues.

One of the more difficult tasks in preparing to deliver a seminar in a foreign country is selecting a venue. I seek sites that have some synergy with searching; inspiring sites likely to put the attending researchers at ease. In Australia, I often choose facilities associated with a State Library. In the US, library meeting space is often unavailable. Hotels obviously want to help but I consider them neither inspiring nor inexpensive.

Finding suitable venues was initially very tedious and unsuccessful. I tried a range of concepts and words like **"seminar facilities" Denver** or **auditoriums -hotel san francisco**. Such searches alerted me to the largest facilities but not to the venues I hoped for.

My first success was to realize that while 'auditorium' is a popular word, the phrase preferred by event organizers is 'meeting rooms'. With this phrase, I could usually find a few meeting rooms in a given city to choose from, though I was still missing many of the exotic venues. I was still being forced through my ignorance into choosing less impressive sites.

My next success was to realize that many of the interesting meeting rooms I was finding were within museums. Obviously there are lists and directories of museums so with this realization came an effective way to find some suitable venues.

With this second success, I looked for other groups maintaining lists of meeting rooms. Yes, in several states the government agencies involved in

disability legislation created master lists of auditoriums and took note of wheelchair access. I could use these master lists to find obscure venues.

Twice rewarded, I sought the final piece to this puzzle within the footpaths of several excellent venues I had already found. Unfortunately, this mostly led to individual meeting rooms. It led to few lists of venues. What directories did exist almost always came from businesses catering to the needs of professional event organizers. These businesses index large auditoriums and sporting arenas. Requests for small meeting rooms lead invariably to hotel meeting rooms, to rooms with prominence and a promotional budget.

This was the difficulty, of course. I sought the unpromoted meeting room in the historical house run by a society with little experience and no budget for internet promotion. I sought the small museum lecture theatre amidst aging American Indian artifacts. Many of these meeting rooms have little or no internet presence.

The final piece to this puzzle was that a local visitor and convention bureau would often maintain a master list of meeting rooms in their city. These lists were not often prominently displayed on their websites and were not well indexed by search engines – hence my difficulty finding them. Lists often had only the first or first two pages indexed. Many of these directories were published as databases and sat beyond the reach of many search engines.

I am never lost for an inspiring venue now. I know where to look. I know how to look. A complex, challenging search resolved itself into a simple search for the city's primary tourist website, usually the local Visitor and Convention Bureau website. When I cannot find a directory, I phone to ask where it is.

This complex and ultimately comprehensive search demonstrates several techniques we have discussed including:

- the use of a thesaurus revealed the key phrase used by event organizers,
- author-publisher identity led me to lists of museums, many with meeting rooms,
- anticipating information helped me select the most likely sites for information about meeting rooms

· and reviewing footpaths to known sites revealed that the visitor and convention bureau often has a meeting room directory.

I no longer find auditoriums by tossing words at a search engine. Notice also how this search demonstrates a common quality to quests with depth and complexity. They unfold gradually. They develop as I notice the players, reveal key terms I need and refine what exactly I want. My search constantly flips between searching for possible venues and searching for how venues are organized.

THE PAGE NEXT DOOR

A discussion of footpath and vetting must also focus on one special case: finding information from a page located elsewhere within a website. Traditionally, the page at the top of a directory introduces the resources within. This page includes anything from a short synopsis to a single phrase where it links to the page we want. This page may also provide information on a similar, related or complementary topic. This neighboring page, this 'page next door', can be a target for us just as we might aim for other pages on the footpath. Find it and we have found the document we want.

Neighboring pages will not have the same keywords as the document we seek but they will have the same author and publisher. A neighboring page may also have more prominence. Perhaps we can search for the neighboring page with more precision too. Under some circumstances, a neighboring page is easier to find.

A government department devoted to economic development publishes a report on the development of a rural region that interests us. This same agency also publishes and actively promotes a page that lists all the rural regions in the state and links to the reports on each region's development. We can search for the report that interests us, which may or may not be indexed by a search engine, or we can search for the more prominent page that lists the reports for each region. Thus, we can search for the economic development of another region and use it to find the document we want. Alternatively, search for pages that mention two regions – triangulate a likely resource.

Very early in my career, I searched for the rate of depression in the adult population of Singapore. Notice this is a challenging search. The answer will not be prominent. It may be hidden. It may be unindexed. It may not be on the internet. Indeed, an initial precise search suggested it was not to be found.

How to proceed? A search for **depression singapore** reveals a great many students lamenting how depressed they feel living in Singapore. A search for **depression medical singapore** reveals a confusing mess of different webpages discussing depression.

Did you notice we are discussing Singapore? We have a geographical dimension to this search. I reached for a regional search engine via **SearchEngineColossus.com** and the search became a fraction clearer. Instead of results that refer to Singapore, results come from Singapore.

After a little more aimless searching, each search a mass of diverse leads with little value, I decide simply to chart each of the organizations in Singapore that have anything to do with depression. Singapore is not so very large, I reason, so the page I seek should stand out clearly from the mess of leads I am looking at.

The first page is a website that describes what depression is. I hack this web address – I reach for the index page in the directory above – and see this is a small psychology clinic. I reason that they are unlikely to have the information I seek and I would not value a figure they tell me anyway, so I move on.

The next link reveals a page on depression. I hack this address and realize it comes from a university student health page. We look for car crashes where they happen most. A student health page is not the most likely place to find statistics about Singaporean health.

As the search proceeds past the first dozen likely sites, I begin to think the information I seek may not be online. I imagine reasons why the government of Singapore would not want to reveal how many of their citizens feel depressed. I begin to consider giving up on my search.

The fourteenth reference leads to a page linking to several documents on depression. Hack. The directory discusses mental health. Hack. Health in Singapore. Hack. This website belongs to the Singapore Medical Association and there, halfway down the front page, rests a search box for their online Singaporean Medical Journal. A search there for **singapore depression** reveals a short snippet summarizing a triennial survey on

mental health in Singapore that mentions 13% of Singaporeans are classed as depressed.[54] Success! I also have the name of the survey should I need further details.

Once again, this search was not simple. I did not find my answer by tossing words at a search engine. We could almost say I was lucky except that I did use several search techniques to my advantage:

- · I reached for a regional search engine.
- · I anticipated the organization I required.
- · I hacked the web address to reveal the publisher and purpose.
- · I recognized the page next door.

I found the answer not as a document that measures depression in Singapore but as a reference to such a document. I approached this reference through other pages that merely discussed depression and mental health. I approached through pages I recognized as potential neighbors. I was looking for evidence I was nearby. This is exactly what I found.

Perhaps I would have succeeded earlier if I had sought a **mental health survey** – words found in the title of the Singaporean mental health survey. Unfortunately, this kind of guessing game usually fails us. There are so many possible titles for such a document. Our searching cannot depend on accidentally stumbling upon the right search words. Instead, I sought a regional organization involved in understanding depression. I recognized the Singaporean Medical Journal as a likely resource – as a publisher I both respected and felt should have the answer.

Such subtleties as the footpath, vetting chains and 'the page next door' are easy to overlook while we search. Yet after we search, when we look back after a successful complex search, we usually find we succeeded not because we selected the right keywords or accidentally looked on the website of the right publisher. We succeeded because we found a page that leads us by the hand to the information we wanted.

This happens rather often once we shift from surfing or searching with precision to higher-order search techniques like feedback, structure and identity.

For many documents on the internet, we may not be able to approach a page directly. Perhaps the page we seek is not indexed, as in the last example. Perhaps it is not distinct in some way that we can find with a finely crafted precise search. Finding a footpath or the page next door

may be our best option. Searching in this manner requires we take a step back and search without precision. We search with a regional resource; with triangulation; with the help of a directory. We search with something to narrow our field of interest other than precise keywords. We then marry this to a finely tuned expectation of what kind of publisher and resource we want to find.

This can be tricky to pull off but remember, for the unindexed page, it may be our only option. If we think about the types of authors, organizations and motivations likely to present the information we seek, finding the hidden page may not be so hard. We already have a clear idea who we want to listen to and what their information will look like.

THE ELEVATED VISTA

Every search establishes a dialogue with the internet discussing the information landscape. Clues drop on us all the time, though we usually fail to notice. How much information exists on our topic? Are the searches clean or muddy? Are the perspectives balanced or polarized? Does prominence favour one side or another? Only by listening to the qualities of groups of information can we decide.

This is the elevated vista – a view of the expanse of information before us. Just as each individual web address tells us something of the content involved, so collections of information offer clues as to how balanced, complete and clearly they cover a topic.

The clean or muddy search is a particularly helpful distinction. Say we ask our question of a search engine, then peruse their recommendations. If these recommendations include many results by authors and publishers that should know the answer we seek – if these sites seem to be 'on topic' – then we have a clean search. Clean is good. It means we will probably find our answer shortly and trust the answer we find.

Muddy is bad. We search but the recommendations are 'off topic'. They offer advice from organizations unlikely to know the answer or organizations we would not trust if they did.

A muddy search is an invitation to search again with more precision. We may find the one or two possible organizations within a muddy list and proceed from there. One of the better matches may show us the keywords we need to generate a cleaner search but that is perhaps the

best we can hope for. A muddy search may simply reflect a muddy topic – a topic muddied by the use of generic terms with more popular meanings. Irrespective, a muddy search tells us it is time to think of our search, to look for alternative ways to search. Set aside our search for an answer. Time to search for our question instead.

We usually waste our time when we scan muddy lists for possibilities. Already we have given up on finding excellent information and now hope to bump into something passable. On the other hand, as we approach where excellent information congregates, we should begin to notice more people and organizations with relevant experience and knowledge. We should begin to notice fewer spectators, commentators and bystanders. The crowd thins, so to speak. We see this in the environment. I am not discussing a single organization having the experience we need. We may see that too. I am referring to how the precision of our search becomes more evident as we come closer to our destination. Information on a particular topic tends to group together in similar structures, formats and authors. It tends to group around a few key nexus points, around discussion groups, around a certain publisher motivation, around something.

This is actually a more general phenomenon evident even back in the days before the web arrived. Information clumps. It groups together firstly, because internet users link similar information together. Secondly, because information built in isolation tends to be forgotten while information produced near similar information tends to be discussed, mutually supported and reinforced with further publishing, recognition and promotion. New content tends to emerge in successful formats and enters discussion near existing successful resources.

In the old days, information would clump around one, perhaps two, internet tools. Copyright law clumped around a mailing list and an FAQ. Today, information may still clump around just a few formats, a few contexts, a few source-types but this clumping is often less evident. Shakespeare seems to clump as databases of Shakespearean quotations just as virus information clumps around several 'ask an expert' discussion forums. Yes, a random stranger may offer advice on how to remove some pesky virus from our system but we will also find several strangers with advice reaching out to us from the same kind of terrain; from the same type of environment. Seeing similar environments in our search is a welcome sign. In a way, it is like finding five antique stores in a street.

Grouping in this way, such herd behaviour suggests competition, cooperation, turnover and a track record of success less common to antique stores that stand alone.

Competition and cooperation brings information together. It reinforces what works and isolates what does not. Many of the better sources, many of the most informed sources, will share format, publisher motivation, context, source, something. They will share similar traits. Think of this as another internet influence akin to prominence, just more subtle. Success breeds competition and cooperation. Competition and cooperation breeds a subtle grouping of information.

This subtle grouping has another, far more significant implication. It guides us to a definitive, comprehensive search. First, however, a note on assistance.

SEEKING SEARCH ASSISTANCE

Can we jump away from the search entirely and merely seek guidance? Yes. We have several options to consider, some so very efficient simply because they require so little effort from us.

1_ Ask our peers or our peer network.

Send a short message to colleagues or a network of peers like an association mailing list. Ask for help seeking information they probably know where to find.

2_ Ask an expert or an expert network.

Send a short message to a person we identify as an expert in the topic that interests us. Look for perhaps an author of a book, a respected article or a prominent website. Email them directly with our question.

Expert networks often work better simple because you ask several people at once. However, if we decide to email an expert network like a relevant mailing list, do remember three points. Firstly, demonstrate we have searched prior to approaching the mailing list. Show what we have found. Secondly, be specific. Choose a clear title and explain exactly what you seek. I usually add a "Q:" to my titles to make it clear I am asking a question.

Thirdly, unless inappropriate, summarize the advice and post it back to the group with a thank you note.

3_ Approach a librarian for assistance.

In recent years, teams of librarians sequestered in public libraries have established a presence on the internet to assist us to find our way. Here in Australia, the AskNow! (**AskNow.gov.au**) service connects us to a reference librarian via webchat to discuss our troubled search. This is great for three reasons. Firstly, these librarians are skilled in searching the internet. Secondly, they have immediate access to resources we often lack, like a business thesaurus and a selection of government directories. Thirdly, just having another skilled searcher think about a troubled search often reveals a solution. I often use this service to great effect.

Keep in mind, reference librarians also inhabit public libraries. They have internet access on their computers and usually count internet assistance as just another form of reference assistance. Visit a public library with your work in hand and ask for help.

4_ Approach a librarian discussion group.

This last option must be approached with caution since we can easily consume a great deal of public time and effort. Mailing lists for librarians like GovDoc (government document librarians) and BusLib (business librarians list) offer us the rare opportunity to ask precise questions of many librarians at once. Do not use this lightly but for very specific questions, this can give spectacular results.

COMPREHENSIVE AND DEFINITIVE

We now reach the one topic that unifies all we have been discussing. How do we pursue all the information on a given topic?

Once embarked on a comprehensive or definitive search, we encounter some very peculiar, very challenging circumstances. By definition, the comprehensive search requires we find information that is not well promoted. We must find information perhaps not directly indexed and certainly information not prominent enough to swim easily into our nets.

There are deep fish, quiet fish and shy fish in our oceans too. Up in the night sky, some stars twinkle only dimly.

It is standard these days to believe everything exists on the internet. We must merely find it. I hope our discussion on publisher motivation has revealed how information may not be available on the internet simply because no one has yet found a way to bring it there. Much commercial information cannot survive commercially on the internet. So much volunteered information goes unnoticed and unrewarded that volunteers must question if their efforts are worthwhile.

Our related discussion on the size of the internet and the vast numbers of matches global search engines generate should equally convince us we may easily miss information present on the internet. Yes, the internet is full of indexed but unreachable information. With so much absent or hidden, we need a way to decide when we have found the best, the most or all of what we seek.

Obviously, we cannot draw this conclusion from any prominence-based search. Such a search reveals only that portion of the answer blessed with attention. We only listen to the loudest voices. Similarly, we cannot claim a definitive or comprehensive search based on precision. The precise search only reveals that portion of the internet answering to a particular term. We assume some sort of tacit agreement on what terms apply that rarely exists. We also search only source documents, not the footpath or the page-next-door. Besides, many topics are not organized in this way.

A true comprehensive or definitive search only emerges by grasping the structures, orders or manners of organization important to a given topic. Comprehensive and definitive answers only emerge once we know how the information we seek is organized.

Let me backtrack briefly and discuss with you several situations where a prominence-based search will not work and a precision-based search may lead to difficulties. These are situations where we need to reach most deeply into our varied arsenal of weapons.

1_ The Comprehensive.
'All' means those with prominence as well as those without. We can find prominent listings easily enough. How will we find those that lack a loud voice? Precision may help but what if we cannot find a way to limit our search effectively.

2_ The Controversial.

Contentious issues are handled equally badly when we only listen to the loudest voices. Those who do not have prominence are unable to reach our attention and are drowned out by louder voices – by those with more prominence. With wind farms, there appears to be agreement among government, wind industry and environmental groups that wind farms are great. However, two different, minority positions will not be heard if we only stay among the very vocal. A blunt search will easily reveal the consensus opinion but may miss other sides to a controversy. A precise search may help but how will we listen to a side of a controversy we do not know exists?

3_ The Critical.

When we have a critical situation, when serious money is at stake, a blunt search is just too amateur. We cannot afford to have a sloppy search and a blunt search based on prominence is just too likely to accidentally miss something important. A precise search will likely miss the unindexed, mislabeled and foreign. We need more certainty than a prominence-based or precision-based search provides.

On three further occasions, we may, or may not be very poorly assisted by prominence and precision. We must usually embark some distance along such a search before we learn we require a different approach.

4_ The Detailed.

Very detailed, very specific information, information for which we want only a certain kind of answer, can be badly handled when we listen to the loudest, most prominent responses. Does orange juice cause cancer? We want a medical answer for this question but a blunt search takes me to newspaper articles instead.

5_ The Difficult.

Difficult searches may be just, well, too difficult. When a blunt search does not work, doing another blunt search usually will not improve matters. When a precise search fails, perhaps the page we seek is not called "A List of Auditoriums in Sydney". Perhaps it was published only last week.

6_ The Definitive.

There are 113,000 webpages by various David Novaks scattered all over the world.[55] It may be too much to ask that a blunt search or even a precise search will lead us to the right one.

Comprehensive, controversial and critical. Detailed, difficult and definitive. As I stated in Chapter One, as we improve our search skills, we see less and less reason to use a blunt search. Yes, a blunt search will always answer certain questions with surprising ease but will make a mess of other questions with equal ease. Furthermore, as we become more of a connoisseur of information, it becomes more and more difficult to craft the precise search that leads to exactly what we seek.

Let me summarize. Prominence and precision is not enough for a comprehensive, controversial or critical search because information can be organized in a range of other ways as we have covered. The same may hold true for detailed, difficult and definitive searches. Only by first identifying how the information is organized, then chasing it down through that structure, order or manner of organization, will we find our way.

Find how information is organized, and we find everything.

We can claim comprehensive if we do this because anyone who has made an effort to stitch their work into the realm of information will first stitch their work into one of the structures, orders or manners of organization we aim to identify. Furthermore, if a resource is good, other internet users may drag it into place even if the publisher does not.

The particular structure, order or manner of organization important for a given topic can be any we have covered. On rare occasions, all our answers will indeed share a keyword or phrase. Rarer still, they may share prominence (as in a list of likely presidential hopefuls). More often, several hidden threads of order and organization exist and we must reveal each one through attentive searching.

Yes, there are difficulties with this stance. The quiet, unpromoted page of an unknown expert may not be stitched into one of the structures or elements of order we reveal as important for a given topic. We may well miss the eccentric neighbor who rents his mansion as a meeting room (but has not told the town council or the visitor and convention bureau, is not a museum and is not listed by the state agency attending to disabilities – the three structures I identified as important for finding meeting rooms).

However, our claim of comprehensive is certainly far truer than restricting our attention to those who use a certain term or those who succeed at internet promotion.

The other part of the answer rests with continually executing muddy searches. Where we find a detailed look at internet search techniques, we usually find several. Not only is it unusual to find a completely original idea but people talk about original resources, particularly those hard to find. Other people will have sought the same information we seek. They too looked, perhaps longer, perhaps then stitching a resource into its rightful place on one of the existing structures important to a topic.

As we search, as we seek the structure, order and organization of our topic, we consult the elevated vista. How muddy our search and how publishers justify their involvement will guide us in this. The elevated vista also allows us to double-check our claim. When all the significant sources we uncover are encompassed by the structures, orders and manners of organization we conclude as significant, then indeed we have a comprehensive and definitive search.

In theory, we may always miss a fine item of critical information that sits alone on the internet. In practice, we see abundant clues that we are coming close to the end.

As a final step, we may wish to invite others to tell us of new resources. Just publish our list somewhere prominent and invite comment.

Before I proceed further, there are some subtle differences worth mentioning when we tackle a controversial search. We especially need the elevated vista for controversial questions. Quiet and poorly published perspectives may not emerge easily. Only by shifting our view to the collection of information, not as a mass of facts but as a mess of perspectives, will we see where we want to go. Try to perceive 'dark areas', under-represented perspectives and bias.

As our example, let us contrast two searches. The first, a search for the best search engine. Should we be using Google? Yahoo? AlltheWeb? The latest tool by Microsoft? We can search for some advice but we really want advice from talented researchers (not just the average internet user) and we could probably focus on librarians for this. Thinking further, we will want to read the advice of published experts, some discussion pieces and perhaps some advice columns associated with library associations as well as library research guides – not the guides aimed at patrons but those

aimed at peer librarians. Ok, I have a clear idea of what we want, and an even clearer idea of what it will look like. I think we will recognize the answer easily.

Contrast search engine selection with a search on the usefulness of nutritional supplements. Even a cursory search will tell us: here is a topic deeply biased by the selling of nutritional supplements. Money is involved and it affects the presentation of information. Add many a researcher explaining complex bio-chemical pathways with scientific certainty and perhaps we can understand why we at first read an abundance of articles that generally avoid telling anyone they don't need anything. The elevated vista is unbalanced. As I search, I see a strong bias pulling me towards buying nutritional supplements.

Ok, we have established a bias. How will be counter it? Can we find a perspective unaffected by an urge to sell? A researcher? Most research would superficially support the need for nutritional supplements. Maybe we could locate at large study comparing the health of people given supplements against those withheld....

Ok, time to change our approach. I think I will try to find someone I respect who tells me I do not need nutritional supplements. I will start with a search for **against nutritional supplements**.

I have looked at the elevated vista and found a likely perspective that I must investigate. That I can not anticipate this perspective only means I must be careful in uncovering it.

Eventually, of course, I must decide if I take nutritional supplements or not. By then, I hope to have found significant arguments for and against taking them. I also hope I have not missed a third or even fourth perspective not immediately obvious to me. Perhaps eating liver once a month would provide most of the vitamin I need. Perhaps a couple visits to a nutritionist will lead me to eat better food and see to my nutritional needs that way. I do not know if either is true but if so, then we have found alternative perspectives not otherwise easy to reveal. Remember the farmers who like wind farms, just not next door? These alternative perspectives may land in our lap while we search or we may seek them out from hints and suspicions but they can be hard to find. Varying the format, context and source will often help reveal these alternative views. I think I really should read some advice from a nutritionist. (Is there one

that does not like supplements?) I would also like to find some sort of health department report or United Nations study if possible.

In the back of my mind, I may also decide the whole industry of selling nutritional supplements is so commercially powerful, and the alternative perspective so unfunded, that any conclusion I reach is certain to be biased. I may need to retreat and say I will always be ill-informed on this topic.

Just remember, a controversial search is a hunt for alternative perspectives since we want to listen to more than the loudest voices.

*   *   *

I will soon refresh an article I once wrote on country profiles. To make it comprehensive and definitive, I will:

· reveal the significant structure, order and organization for the information I seek,

· confirm all the significant resources I find are stitched into these structures, orders and elements of organization

· and establish a way for new significant resources to reveal themselves to me.

With this in mind, I will

· scan various prominent directories and lists for new country profiles,
· triangulate further lists by searching for two of the most popular country profiles,
· identify the various organizations involved in publishing country profiles,
· reach out to comparable organizations through directories and government hierarchies to seek further country profiles,
· post a copy of my list to the GovDoc and BusLib librarian mailing lists seeking comment
· and publish my article with a invitation for visitors to suggest further profiles.

As I search, I will pay particular attention to the elevated vista, looking for pockets of information I expect to find but miss. I will seek until I am

satisfied I left no gaps unexplored. When I search for travel advisories, I will approach the websites of government departments in countries likely to have travel advisories. For country profiles by organizations like CARE and the Red Cross, I will seek a definitive directory of such organizations then look through each organization listed for evidence they maintain a country profile. In this way, I allow notions of who would publish to guide my search.

When I am done, I will look again at the mass of detail I have collected and determine if I uncovered any country profiles from unexpected sources, perhaps from organizations not part of a structure I identified as significant. If all are stitched into places I have looked, then I will sigh in relief and claim a comprehensive search.

Comprehensive and definitive is possible but it requires the best of search skills.

As an aside, I am amused that after our long journey of discovery into how internet information is organized, after over 280 pages of discussion, description and observation, here I sit writing that many of us almost had it right. A simple searcher searching among prominent sites and searching for webpages with certain keywords will tentatively state they have searched the internet and found what it has to offer. They know 'comprehensive' is too much to ask and they hope too much information did not slip through their fingers.

If only the simple searcher knew a few more ways information can be organized besides by keyword and by prominence.

Let us break for a moment and look at comprehensive in a very blunt, less theoretical way. Before we conclude we have a comprehensive search, we must:

> 1_ Search a really big database in a very precise way.
> Google indexes, or is aware of, more than 20 billion webpages, a number sure to grow in time. This is a lot of information. If we ask our question in a very specific way, we may find our answer just by chance. With search engine punctuation, refine our search until we have just a few matches, then hope. This approach works best with clean, not muddy, searches.

2_ Reveal how the information we seek is organized.

What terms, techniques, tactics and structures best lead to the information we seek? We can do this by following the foot-paths of the first few suitable resources we find, by specifically searching on how to find this information (How do we find cheap airplane tickets?) or by working it out ourselves.

3_ Build a list of resources that interest us.

Use the revealed structure, order or organization to find more other resources of interest to us.

4_ Consider if what we have found is unbalanced in some way.

Does the elevated vista suggest we are not looking at all the resources of interest? Perhaps a discipline or industry is not represented among the sources we find. Perhaps we have not heard from a type of publisher we thought would have more to say. Perhaps the missing voices use a different term or do not have the prominence to reach us.

5_ Ask a real person for further advice.

Beyond the landscape of information, the internet is also a community of real people in discussion. At this level, we may even move beyond internet resources to reach information not yet published but found in the minds of those participating in this wondrous creation.

At the end of Chapter One, I posed a simple question: "How can we find information not indexed by a global search engine?" I suggested that until we can answer this question, we have not truly touched the heart of searching.

The answer is in our pursuit. On the next page is a graphical representation of the answer. We capture some key – geography, format, structure, organization, a specific phrase, an author or publisher identity, the 'page next door', something! We then turn this key and unlock our question. Unindexed or buried internet material will surface when we use a tool, structure or technique that takes us there.

These keys empower us to unlock a topic and craft a truly comprehensive search.

Finding Unindexed or Hidden Information

global,
regional or
topic-specific
search engines →

internet structures
like
government hierarchy

related publishers

target
website

link companions

website
search

page
next
door

page
next
door

★ the
page
we
seek

similar
page

Chapter Ten

_____
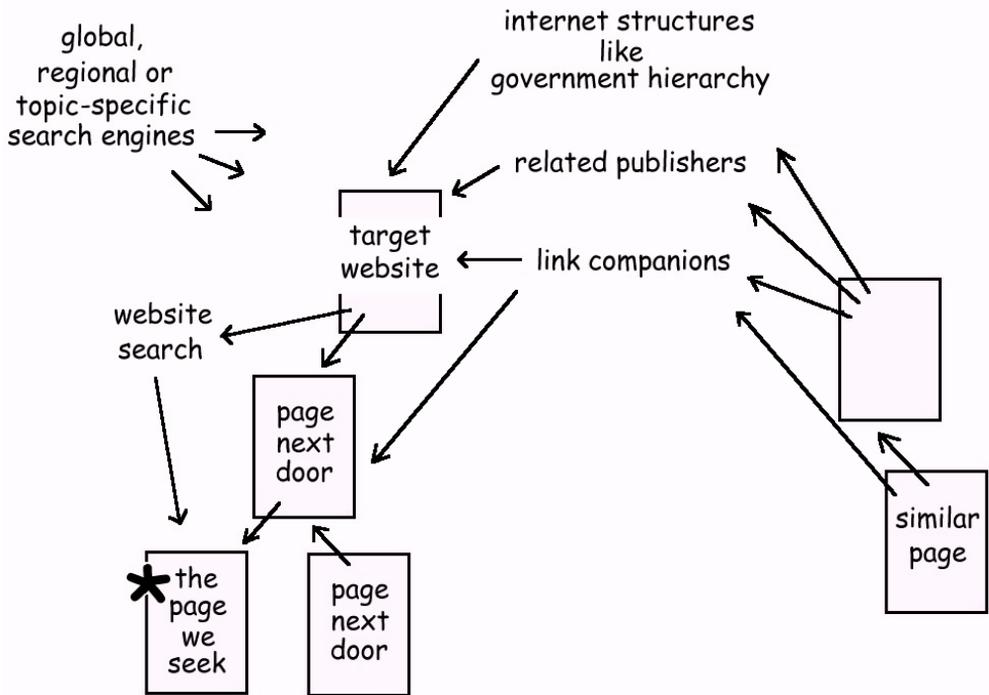
CHOREOGRAPHY

I nformation is an object. We sort and select information by the words within. With Boolean, proximity and field searching, or simply by chasing specific words, we shake the internet and hope the facts we seek surface in the moving sand.

Information also has fame. Frame our question so the answer has prominence and our trusted global search engine, searched in a blunt manner, will surely direct us forward.

Information has history; it has context, format and source. It stands beside related information. It is prepared in one of only a few standard formats. It has an author and publisher whose identity influences the information. This halo of supportive detail both identifies information as unique and enriches our appreciation of information. We use this halo both to reveal quality and to anticipate our destination.

Information has organization too, indeed multiple overlapping layers of organization. From carefully vetted and peer-reviewed criteria to professionally managed categorical schemes to near random web-spidering, information is indexed, catalogued and presented to standards. All our search tools offer their judgement based on selection criteria and bias. With finesse, we can use this to draw out better information.

Consider also structure. It creeps in from publishers, from government hierarchies, from geography, nexus points and directories. These and further structures internal to a website offer pathways for us to find certain information more easily.

Now view information as more than an object. As a message. Each message plays a bit role in the grand cacophony of communication surrounding us. As a competitive realm, publishing must be justified. Not

everything is available. That most available, most visible, must justify its existence most persuasively. In this way, motivation flavours content, presentation and location. Search with something specific in mind and abundant clues will tell us as we approach.

Lastly, more than facts, more than history, more than organization and message too, consider also the elevated vista – the holistic view of the information environment. Gaps in information, missing perspectives, unbalanced and biased discussions all become visible if we look at collections of information; if we look from a wider perspective.

Objects, prominence, history, organization, message and elevated vista form a sliding scale of perspectives we all too easily overlook. Popular culture portrays the internet as a great chaos of webpages tamed only with links and brute force searching. We lack the opportunity for finesse.

Let us not encumber ourselves with such a view. Instead, let us revisit our metaphor of the internet galaxy. The stars above are more than pinpoints of light. More than solitary objects. Spun into patterns by forces unseen, our galaxy is awash with structure, order and organization. Much of this structure and order hides from view but remains all the same.

Our internet is deeply organized. Position encoded in the humble address. Organization embedded in directories, nexus points and link companions. Patterns embedded in context, format and source. Structures spring from the way information is published; from the purpose it attempts to achieve. The internet weaves each item of information into a host of patterns, structures and more.

This is why searching demands of us such clarity of thought. Confused, our search joins us in confusion. Fail to see bias and our search becomes unknowingly biased. Ask too general a question and we forge too general an answer. Imagine the internet as chaos, overlook its structure, order and organization, and we will surely notice only chaos.

Here is a list of the many search techniques, tools and concepts we have discussed in this book. Keep these in mind as we search and we will not so easily drift into confusion.

SUMMARY SHEET – SEARCH TECHNIQUES

--- Precision ---
1_ Punctuate our search queries
    a. Plus, minus, OR and quotes
2_ URL field search
3_ Link field search

--- Prominence ---
4_ Prominence-based searching
    a. Prominence as an asset of commercial value
    b. Prominence as a quality we may desire
5_ Importance
6_ When surfing works

--- Quality ---
Q1: Internal clues
Q2: Author and publisher identity
Q3: Local context
Q4: Endorsements
    a. Three types of endorsements

--- Identity ---
11_ Context
    a. Link companions
12_ Format
    a. How do we want the information prepared?
13_ Source
    a. Who would author and publish what we seek?

--- Haste ---
14_ Bookmarklets
15_ Embedded forms
16_ Juggling windows
17_ Shortcut keys
18_ Non-linear research styles

A SEARCH AS A DANCE

What do we search for? Our question is the flashlight we use to illuminate our answer. Questions lend our mind focus. Ask and we shall follow. With many search techniques available to us, we have options. We can move through the internet quickly now, in several directions with a certainty born of familiarity. We know what we look at, what we will shortly see and what we hope to find.

To this list of techniques, I wish to add another list; a list of principles on how best to use the information present on the internet; a list on the choreography of our search.

    1_ Search.
        a. All challenges are needs for information
        b. Synthesis is faster than innovation

Let us start this next list simply by urging ourselves to search. We live surrounded by a vast galaxy of information. Most of what we do is not original. Even original work benefits from experience borrowed from other disciplines and industries.

Any challenge we encounter is a request to search. Any project that stalls is a chance to see how others succeeded. Any desire to improve is a time to ask how others improved similar projects. It is egotistical of us to say we know best; we are the experts; we will find our own way forward. This reticence to compare notes with the rest of humanity is only a choice, nothing more. If we would rather solve a challenge in isolation, we can, though it is generally faster and easier to synthesize a solution. Help from the world's reservoir of experience stands waiting.

    2_   Ask the right questions.
        a. Our questions is our flashlight

There is a nimbleness to asking questions that reach for exactly what we want. Make our assumptions explicit. Evolve our search step by step. Practice. The best questions lead to the best information. Ill-considered questions lead us astray.

    3_ Search for something excellent.
        a. Raise our expectations
    4_ Search for something that exists.

Too often we retrieve poor information only because we do not ask for great information. Embrace whatever the internet tosses our way and we have lost an opportunity for excellence.

Instead, ask the tough questions. Seek the latest opinions, the best statistics, the least-biased truths. Gather multiple truths from alternative viewpoints. With skill, we generally find what we demand of the internet.

Of course, the internet is just one part of the information realm. Not all information congregates here. Sometimes we must reach for a library, for a commercial database, for professional advice. Search for something that exists, something someone has chaperoned to the internet, and we will enjoy greater success.

5_ Gather experience.
    a. Where to look
    b. How to look
    c. Who publishes what, where
6_ Practice attentiveness.
    a. See what is before us
    b. Watch our questions

Part of search excellence is simply experience mixed with attention to where we are, what we are doing and what we are asking. Excellence comes from having done something before and noticing the similarities. Yes, we may never have searched for this key piece of information but we have surely searched for something similar or undertaken a search that felt a similar way.

7_ Recognize bias.
    a. Search tool bias
    b. Perspective bias
8_ Know what we consume.
    a. Habitually investigate quality
    b. Prefer prominent resources with longevity
    c. Information overload as imprecision

Excellence also comes from our attention to the information itself. Let its rich flavour reach us. Who wrote it and with what perspective? Where did we find it and why? The halo of supportive detail tells us this and more. Attend also to the influence of search tool bias and of bias in general; even bias affecting a whole topic.

9_ Understand the internet revolution.
   a. A drop in the price of information
   b. A refining of what constitutes original
   c. A transient, demonetized environment
10_The internet has a history and a future.
   a. Stages in the internet's development
   b. Three levels to its organization
   c. Unending growth

And of the internet itself? Do we understand what is happening to our lives? Do we recognize this massive tidal wave of information for what it is? Plan for a future where information has less value, not more, and truly original work is rare indeed. Too many intelligent people in the world have looked at too many of the challenges that beset our world and the people in it. Too many have something worthwhile to say. The future belongs to those who can marshal diverse opinions and experience. It belongs to those who can synthesize a fine conclusion, then act.

11_   Enjoy the hunt.

Finally, no longer quite so isolated from the ideas of others, no longer quite so distant from the information world surrounding us, we encounter new ideas frequently now. We seek new ideas simply because we can.

1_ Search.
   a. All challenges are needs for information
   b. Synthesis is faster than innovation
2_ Ask the right questions.
   a. Our questions is our flashlight
3_ Search for something excellent.
   a. Raise our expectations
4_ Search for something that exists.
5_ Gather experience.
   a. Where to look
   b. How to look
   c. Who publishes what, where?
6_ Practice attentiveness.
   c. See what is before us
   d. Watch our questions
7_ Recognize bias.
   a. Search tool bias
   b. Perspective bias
8_ Know what we consume.
   a. Habitually investigate quality
   b. Prefer prominent resources with longevity
   c. Information overload as imprecision
9_ Understand the internet revolution.
   a. A drop in the price of information
   b. A refining of what constitutes original
   c. The demonetized environment
10_ The internet has a history and future.
   a. Stages in the internet's development
   b. Three levels to its organization
   c. Unending growth
11_ Enjoy the hunt.

INTERNET SKILLS AS DECISION MAKING SKILLS

What follows a good search? Let me place searching in context. New ideas are not the same as challenges surmounted. As part of the Leonardo Institute here in Melbourne, I teamed with Geoff Kelly, a public relations/strategic-positioning expert, and with Adrian Farrell, a competitive intelligence expert. For my part in this trio, I explore just what to do with better information.

Decisions are complex affairs, dependent not only on the information at hand but also the opportunities open to us and what we must achieve to make a decision successful.

I was most intrigued to learn that in competitive intelligence environments, a researcher may spend more time analyzing information than collecting it. A complete biography created in a skillful search may then be analyzed for behavior patterns. Through in-depth role playing, this biography may reveal how the CEO of a rival company will act in a given situation. An internet search represents only a small slice of such a research project.

Successful decisions, furthermore, depend not just on our information but on the careful reading of our opportunities. Given our resources and the people we must work with, a decision may be very ill-advised because of factors that have little to do with information and everything to do with presenting an achievable vision or achieving a solution.

In a sense, politics is resplendent in taking the best available advice and then doing only the achievable. This is a far cry from learning the truth, then deciding our course of action based on this alone. The whole notion of what is 'doable' is critical here. So is the rival notion of whether a decision takes an organization in a direction that empowers the organization for future opportunities.

For example, we may well decide based on a search that a wonderful opportunity exists to save a great deal of money by implementing a given technology. However, it requires more technical expertise than our business has, we would be unable to scale the savings through our organization and it would conflict with other technology projects already underway. Our internet search shows savings. Our considered decision: Don't pursue it.

Decisions are complex activities. Some of this complexity also involves us as decision makers. We are not separate from the decision process. We

delay a decision for personal reasons. We hunt for proof of our precon-
ceived conclusions. We search well but decide badly. We decide well but
fail to learn from past mistakes. Making a decision is itself a skill, not
unlike searching the internet.

I am particularly drawn to several of the skills that help our decisions
become successes. Skills in leadership and effective communication.
Presenting a decision in the interests of those who must participate.
Presenting a decision as a story, as a tool, as a significant difference, as a
worthwhile goal.

We can also learn to make decisions more swiftly, to move through the
decision making process faster. Not rushing decisions, just deciding more
smoothly, without delay. Such speed can become a significant business
advantage.

Now given this additional layer of activity sitting above our search,
should we not consider this as we search? In the same way as I earlier
described how the needs of synthesis can guide our search for balanced
information, so too can the needs of decision making guide our search for
effective information. I find this both obvious and neglected. Yes, we
search with our needs constantly in mind. However, do we let our search
affect what we decide we need? Do we establish a dialogue between what
we think we need and what we are finding?

All too often, decisions are made after the facts are all collected – a
rather out of date approach given an internet that makes collecting "all
the facts" a rather laughable statement. A search could unfold with
synthesis and decision making in mind. We can attend to the transparency
and diversity of materials (synthesis) just as we attend to how appropriate
the emerging conclusion would be for our purpose and audience (decision
making).

I do not mean we toss a backward glance at whether the information
we are finding is suitable for our client. I mean we ask if the decision our
research is leading us towards is a decision our client should make. If we
assume some of the role of decision-maker, we will waste less time collect-
ing information without value. We can establish a better dialogue between
what we are finding and what information is ultimately found useful.
Decisions will not just be based on how the research was initially framed
but will evolve as we execute our search.

ATTITUDE

Let me end this chapter by conveying something of my attitude towards information. We do not need to become the cynic as we learn search skills.

I did.

Something in knowing the quality and background of information fed my cynicism. I look at a television commercial and see a marketer spouting legally grey words that have no real meaning. "Begins to work immediately." "Part of a balanced diet." What rubbish! I do not think breakfast cereal is good for me because I believe the World Health Organization and not the image of an athlete on a cereal box.

Similarly, we do not need to look at information and see questions.

I do.

I look at a politicized event and ask for the alternative perspective. I look at a picture of the Andromeda Galaxy and wonder which telescope took the picture. Is it real or artistically enhanced? Something in the practice of asking questions reinforced my habit of asking questions. I encounter a challenge and questions quickly rise to mind. How do the French handle Islamic fundamentalism? They have dealt with it far longer than the US.

Something in the internet also changed the way I hoard information. I dump information quickly now. Need something again? I will search again. I will find something more up-to-date, something more inspiring. I save little from previous searches except perhaps the knowledge of how the information was organized. I rarely notice the specific tools that organize a topic; I rarely record those authors and publishers I encounter along the way.

All these changes stem from learning how to find information. They stem from being Internet Informed. Will you change too?

EPILOGUE

The mid-morning mist crawled quietly up the hill towards the castle where the aging Albert sat gazing from a second story window. Dirty grey-white sheep grazing in the valley below merged with the mist. He knew they were there. Even without seeing them, Albert knew he need only grab his walking stick and hobble towards the fields of his youth. The sheep would be there, waiting for him.

Albert was the castle's master now. A small band of soldiers looked to him for leadership. Charged with offering peace and security to the small town below, he prepared for a day sometime in the future when his small castle would stand alone in keeping chaos at bay.

"Work at peace, prepare for war," he said softly to himself as the day began to unfold.

With a shout from below, his youngest son ran along the castle wall thrusting up at him a freshly caught fish and a beaming smile that warmed Albert's heart. With a short grunt and a last look at the horizon, he stood. "Time to work," Albert politely reminded himself. "Time to make my experience count."

*　　　*　　　*

The internet revolution washes over us. Our place in this revolution – as passive recipients of ever-growing quantities of pre-selected and prepared information or as members of a talented digital elite, gathering and synthesizing exactly what we need – our place depends on how firmly we grasp ideas like those expressed in this book. Are we information connoisseurs or inattentive consumers? Do we search and explore artistically or simply guess, then select from the prominent morsels paraded our way?

We decide. What we cannot decide is our participation. Like so much driftwood, we have no choice. After 40 days and 40 nights of rain, our fate is sealed. Our lives must forever contend with this sea of information.

I introduced the story of the young Frenchman, Albert, to help us grasp the nuances and subtleties to information. There is a tendency to believe peace is achieved with the edge of a sword, with the point of a pike. Albert found this is not so. There is a tendency to consider internet information as something special, as something unique. There is a tendency to consider everything internet-related as truly original and post-post-modern. This is not so, yet conditioned in this way, we set aside everything we know and instead, whisper a few words to a plastic box, then pray.

With such simple gestures, we certainly unearth a vast amount of information – some of it with real value. Yet with such reckless simplicity, we also fret amidst information overload. We frustrate over getting our search words right. We pay too much attention to every new advance and search engine improvement in a desperate and vain hope it will at last provide what we should never have asked in the first place.

Search engines are wondrous. Place them in perspective and we see how truly wondrous they are – magical even – but never solitary; never our only avenue to explore.

There was another story within this book, a story of how the internet has grown and developed into what we have today; what it will be tomorrow. All Albert wanted was to contribute to the sense of justice and security in his community; to contribute to the renaissance emerging in southern France. Yet in the end, a very different sense of security emerged, one that addressed the security needs of the Roman Catholic Church and their wish to keep Christianity safe from heresy.

In a similar way, the internet's youthful enthusiasm and utopian dreaming wished only to contribute to a happier, more lively sense of community; to an alternative to capitalism. Yet in the end, a very different sense of community will emerge – a future that includes the participation of business and academia; a future both enriched and encumbered by the fruits of other systems of communication. What will emerge will be very different to the dream inspired by the internet's arrival.

Will this revolution of ours succeed in creating a new renaissance? Just as Albert's Southern France emerged as something very different to the

flowering romanticism and religious tolerance evident for a time before the Abyssinian Crusade, our internet revolution will create something wondrous too – wondrous but different.

Of the highest significance, this revolution triggers a tremendous fall in the cost of information. It triggers the sweeping aside of social structures and business ventures based upon pricey information. It liberates and surrounds us in an abundance of information. Only ignorance can blunt this aspect of the revolution now. A community of brochure-munching prominence-preferring information consumers does not receive, nor perhaps deserve, many of the rich rewards this technology offers. Will only the digital elite be Internet Informed?

This concern now motivates my work as I strive to understand and convey the art of searching the internet. The story of our internet empow-ered society may yet be very different if we collectively understand and grasp the art of searching the internet.

Except that searching is not art. It is too simple to be art. Speak of it instead as 'artistic attentiveness'. Like a passion for painting, do we gaze at a canvass and see paint strokes, palate and emotion? Or just a pretty meadow? Do we gaze at the internet and see qualities, structures and avenues to explore? Or just a pretty webpage?

Like art appreciation, information appreciation is for everyone. The skills are simply too simple to think we cannot attain them. Appreciate how information is produced, arranged and found. Appreciate that what we know about finding information away from the internet applies online as well. Fine searching follows quickly enough from here.

Yes, it helps to be artistic. Polish those skills at crafting specific searches. Practice asking illuminating questions. Take the skills and tactics in this book to the pinnacle of speed and precision. Just don't, for a moment, imagine the skills and tactics shared here are not for you.

Let us consider some of what we have learned:

- Ask specific questions or ask for something prominent.
- Reveal quality by investigating the halo of supportive detail.
- As we search, anticipate our destination.

So simple! Our journey starts only with a degree of interest; with an attentiveness to the great internet canvass. Little else separates the digital elite from other mere mortals.

Stars hang overhead. Even with so splendid a night-time view, we don't often see stars. We let our attention drift earthward. We merely acknowledge night arrives. Yet, within the stars, the moon and swirls of cosmic debris, the very process of creation is made visible. We view the universe being born. Nebulae melt away. Star clusters dance. Hot young stars blast holes in the carpet of cosmic dust. So much happens above. Do we notice? Do we appreciate the view?

Information surrounds us too. Yet even with so splendid a panorama, we don't often see. We let our attention dwell on the convenient. We leap erratically from fact to fact. We merely acknowledge the environment is there.

Be the connoisseur of information instead. Move swiftly. Anticipate destinations. Let our search skills change the way we relate to information – the way we hold, hoard and value information. Frame our challenges in terms of questions we can answer, then answer these questions to overcome our challenges. Informed by the internet in this way, lead on to greater achievements. So much worth achieving awaits us.

The midnight sky stretches overhead. Thousands … millions … billions of stars twinkle and shimmer….

SUPPORT

(curriculum assistance and future publications)

GLOSSARY

I use many novel terms and phrases in this book; some not usually associated with the internet. This simple glossary may prove helpful.

Bias

Bias is a preference, a prejudice for certain types, perspectives or qualities. All search tools display a bias in how they select and rank information. This bias can assist us or misdirect us but should always be acknowledged and understood. Bias is also an issue in quality assessment.

Bookmarklets

A bookmark is a link kept on a web browser for quick access. A bookmarklet, (notice the suffix 'let') instead of linking, causes an action to occur. Click a bookmarklet and run a javascript. One of the most famous bookmarklets is the Google Browser Button. Bookmarklets can help us move more swiftly through the internet or quickly retrieve halo information at a single click. Bookmarklets are described in Chapter Five.

Boolean Logic

George Boole first established the use of AND, NOT and OR and their role in set theory. Together, these three tactics are called Boolean. On the internet, however, it helps to treat Boolean as three separate tactics and to use Plus (**+**) for **AND** and minus (**-**) for **NOT**. Boolean, brackets, proximity and field searches combine to deliver very precise searches. In this book, we collectively refer to them as search engine punctuation.

Context

Context refers to the place information is found. If we find some information in a library or bookstore, this tells us something of the information. It tells us it passed their vetting and it conforms to their selection bias. This in turn suggests

a format and quality. On the internet, this context still exists and indeed proves very helpful in quality assessment and anticipating resources. We find context with endorsements and with the URL field search.

Deep URL Interpretation

If we look beyond how **.gov** means government and **nike.com** means Nike shoes we will see the address suggests much more – from publishing model to format to publisher identity to depth of focus. The process of teasing out meaning and suggestions of meaning from a web address is Deep URL Interpretation.

Elevated Vista

We lift our eyes to the elevated vista when a single item of information seen on a webpage is viewed as just one statement of a larger discussion. This holistic view of the internet may suggest an issue is contentious, may tell us we listen to only some of the significant voices and may indicate what we are missing. We need this help to answer comprehensive and definitive questions. We see the elevated vista in the number of matches returned by a specific search and by considering a collection of resources and perspectives.

Embedded Forms

A form is a bit of complex HTML that we usually see as a textbox and button on a webpage. The Google search box is an example of a form. We can move forms from one page to another in a way described in Chapter Five. Moving, embedding forms within descriptive text and making alterations to forms can help us search more swiftly and effectively.

Endorsement

Internet links do not stand alone in stringing internet information together. Webpages may also mention other webpages by name or include an address. These three are all endorsements or referrals. Do note that internet endorsements may be positive or negative. We may focus on the fame, attention and presumed significance found in the number of endorsements or we may focus on the content and source of key endorsements. Endorsements feature prominently in quality assessment. They also help us to move horizontally through the internet to comparable resources. (See link companions.)

Feedback Research

The process of searching often unfolds as a search progresses. Questions gradually become more specific and we may only uncover the critical keywords or phrases after visiting and reading through several resources. Feedback research involves us 'feeding back' what we learn as we search into making a better search. In an unrewarding search, for instance, by focusing carefully on

one or two resources we like, perhaps we can determine what we like, then refine our search to reveal more resources with those qualities.

Field Search

A field is a collection of facts about items of information. Publisher, length, format and publication date are all fields. Often fields can be searched, as with the author/title field of a library catalogue. On the internet, prominent fields include title, URL, link, filetype and index date.

Footpath

Once published, knowledge of information moves outwards, away from the information first to neighboring resources, then gradually further. The path that publishers create to reach those looking for them is their footpath. It is this path visitors may tread to reach them. Prominent sites have extensive footpaths. Search engines alone usually make very poor footpaths.

Format

Information is prepared in only a small range of formats like the book, article, memo and press release. It is important to remember format describes the logical manner of preparation, not its physical form or its manner of presentation, its filetype. Web is not a format. Indeed, all web material is first prepared in a format then presented as a webpage. Internet information is organized by format. Format also has a role in quality assessment and in anticipating information content.

Forms

A form is a set of text boxes, radio buttons and clickable images that allow us to interact with a database. The form is created in Hypertext Markup Language (HTML) with just seven HTML tags. Forms can be moved, altered and adapted to our needs as described in Chapter Five and **SpireProject.com/art20.htm**.

Hits versus Visits

Hits record the number of images or text files retrieved. It corresponds to the number of lines entered into a log file that records every request of a web server. Thus, downloading one page may trigger twenty hits (one in HTML, a supporting javascript page and 18 separate images). If we add more images to a webpage, the number of hits goes up. Visitor counts, on the other hand, approximate the number of real people visiting a website, usually by using unique DNS entries to track one visitor travelling through a website. Hits, in comparison, measures web server activity; how many images are found on a page, how many pages the information is spread across and how many times the pages are requested. Do not

be impressed when we read some event triggered a million hits. Visits are the only traffic metric of value. Visits, unlike hits, measure prominence.

Identity

Information has an identity that includes the identity of the author and publisher, the format the information was originally prepared in and the context of its publishing – its neighbors. This identity is much like a fingerprint or a short biography. In this book, I urge you to observe and acknowledge the identity of the information we encounter. Elements of identity may be revealed in the URL, from a quick glance at the document or from a precise search for local context, endorsements or source information. When searching, we can reverse this image. We search and scan for resources that feature the identities we desire.

Importance

Important information is factual, trustworthy and valuable. Such information is superior to the alternatives. We judge importance based on the needs of our question. It usually includes depth, trust, currency and other hallmarks of quality but not always. Because of this, importance differs from prominence. Importance cannot be computed by a computer algorithm that does not know our search question and criteria.

Information Venue

Some websites are the internet's equivalent to a specialist library. These venues are characterized by a collection of information resources on a specific topic as selected by someone with relevant experience. They have a certain look, can be found using the link field search and have a role in revealing comparable information. Information venues are often directories but the better information venues are rarely the global directories.

Library Science

Computer science deals well with information as a fact but library science deals much better with information as a statement or opinion. Library science encompasses quality issues like bias, authority and support. It includes search techniques like Boolean, target searching and feedback research. It also encompasses such tasks as collection development, virtual reference and other topics that impact librarians. The discipline continues to develop but has a firm history partly grounded in the commercial information world and commercial database research.

Links

A link refers to the pointer going from one webpage to another internet resource. Links that point to the page we are on, inbound links, have an important role in

revealing comparable information. Links on the page we are reading, outbound links, are used in surfing and can be found by looking on a webpage or using footnote software.
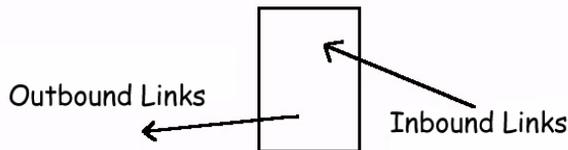
Link Companion

Start with a webpage. Find an inbound link, preferably from an information venue. Now look at the other resources our information venue links to in addition to our initial page. These link companions are considered similar or comparable by the author of the information venue.
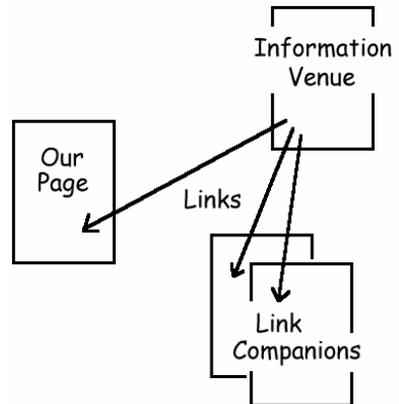
Link Search

We can search for links by using the standard **link:address** on a global search engine. This tells us just one form of endorsement; it lists other webpages that link to the page we specify. We use the link search to judge prominence and to find comparable information.

Nexus Points

Some sites succeed in gathering together a good collections of information about a specific topic, then assume the role of informing visitors of new resources as they emerge. Such 'nexus points' seek and report on further resources often through moderated discussion. Nexus points may be websites, newsletters or mailing lists. They share much in common with portals and information venues. Unlike directories, however, nexus points do not focus on being definitive or complete. **SearchEngineWatch.com** and the newsletter ResearchBuzz, for example, are fine nexus points for information on internet search engines. **SpireProject.com** is not a nexus point since it focuses more on education than resource discovery. Nexus points are a common source of structure on the internet, and occasionally a desired destination we will seek with purpose.

Non-linear Research Style

Juggling windows is a tactic that improves the speed of gathering information by running several copies of our web browser at the same time and jumping between them. There are several keyboard shortcuts to assist with this. If we move in

different directions with different copies of our browser (or different tabs when we use a tabbed web browser) we can also change the way we explore from a linear page-by-page manner to a non-linear, amoeboid approach, opening and closing windows frequently as we search. Non-linear research styles are very effective and more closely match the way our minds work.

ODP – Open Directory Project

The Open Directory Project (ODP) is a very significant directory of similar dimensions as the Yahoo! Directory. Both were pivotal points of reference during the directory period of internet's history. Yahoo is a commercial effort that charges a significant US$299 annual fee[56] for consideration and likely listing, (though many if not most Yahoo listings were awarded rather than purchased or were purchased at an earlier time when lower fees applied). The Open Directory Project harnesses a vast network of volunteer editors instead. Most editors specialize in their area of focus. In this way, the ODP resembles the Wikipedia. The ODP emerged from Netscape and their **Mozilla.org**. Indeed it was once known as the DMOZ directory though it has a history reaching before their involvement. Today's Google Directory is a version of the ODP. Indeed, so many copies of the ODP appear all over the internet that a listing in this directory, like a listing in the Yahoo Directory, is a significant boost to a site's prominence.

(The) Page Next Door

When searching for a particular page, we can search either for the document we want to read or a page nearby, perhaps the page that introduces that document or any page residing nearby. This is the 'page next door'. A neighboring page may briefly describe the document to website visitors. A page next door may have greater prominence or may address related topics. Such pages may have other concepts we can search for but will not have many of the technical terms found in our eventual destination. Searching for the page next door is a significant way to find unindexed resources or documents lacking in prominence.

PageRank

Google measures webpage prominence using a standard metric that they share with us through the Google Toolbar. It is just one element in search engine ranking but a most significant and visible element. The exact algorithm Google employs to calculate PageRank is not publicly known and changes periodically, though much is known from Google's behavior and from the original Stanford University dissertation that led to Google's creation.

Precision

By using search engine punctuation like Boolean, proximity and field searching, as well as adding or subtracting additional words and concepts, we can craft a precise search where we get about as many results as we are willing to consider. We should certainly have fewer than a thousand matches and hopefully, more than ten. A precise search tells us something of the elevated vista; the amount of information available on a topic. Precise searches also offer better answers for certain types of questions. We discuss precision in Chapter One.

Prominence

Prominence is a description of fame, awareness and traffic. Prominent sites have more links, more significant links and more traffic than comparable but less prominent sites. They appear high on the search engine results pages. The interplay between prominence and importance helps us understand internet promotion and search engine bias. We discuss prominence in Chapter Two.

Publishing Models

We can reduce internet publishing to three basic motivations as represented by the utopian, commercial and academic publishing models. Each have different reasons to publish, different ways to fund publishing and different success. Internet history unfolds as a conflict between these three models. Publisher motivation, or publishing model, is also part of the halo of supportive detail we use to recognize and anticipate information.

Quality

Quality describes the value of an item of information without describing what we specifically need. It usually includes a range of desirable qualities like reputation, depth and currency. If a statement of quality is not grounded in our question, however, quality becomes a crude estimate, not a great measure of value or importance.

Quality Assessment

To reveal the many qualities of information, we ask, investigate and interrogate information. In this book, I suggest Q4 quality assessment based on asking four specific questions of information, addressing content, source, context and endorsements.

Recommendations

When a search tool presents information that is not comprehensive and represents just a few resources selected from many, we are working with recommendations. Directories obviously recommend certain resources because they select information to fit their criteria – usually persistence and significance. What is not

often recognized is that search engines used in a blunt manner (not a specific manner) also offer recommendations. Indeed, search engines and directories tend to offer similar assistance.

Search Engine Punctuation

This includes the following tactics: Boolean's + - and OR, the use of brackets if available and field searches, particularly title, URL and link. By using punctuation, we build specific searches or we restrict a search in some way. For example, we can ask for only Australian websites by including **inurl:.au** in our search.

Search Tool Bias

Search tools select information for our attention. They select both the information they index and how they rank this information. All tools, especially the global search engines and directories, skew our attention towards certain types of resources. They display a bias. To make refined use of these tools, we need to understand this bias and recognize when this bias works in our favour.

Shortcut Keys

Most software has certain set key combinations that make often-repeated tasks easy to execute. Control+C, for instance, copies the highlighted text in almost all Microsoft Windows-based programs. Practiced use of such key combinations makes for faster and smoother use of a computer.

Sociology – Internet Sociology

Sociology is the study of society; a kind of group psychology. It deals with how we influence society and how societies influence us. In relation to the internet, sociology considers information as a message from one person to another in a competitive social realm. Cyberspace is more than information. It is also the habits, standards and institutions that guide internet communication. This perspective helps us understand publisher motivation and reveals another dimension to the internet's history and future.

Source

Source simply refers to the author and publisher. Too often internet information is presented or read without concern for who creates the work and who chaperones the work through the publication process. The experience and bias of both author and publisher are significant not only to quality assessment but also to searching swiftly and in search strategy. If we can predict the kinds of authors and publishers who produce the information we want, we can search accordingly.

Synthesis

As an alternative to invention, synthesis involves the merging of information from separate sources into a conclusion. We raid existing experience, then forge or apply it to something new. The strength of synthesis depends on diversity, agreement and transparency as we discuss in Chapter Three.

Tags (HTML tags)

HTML, Hypertext Markup Language, is used to change a normal text file into a web page ready for display in a web browser. HTML includes a great many tags: simple words framed on both sides with the < > symbols. For instance, the title of a webpage is found between the tags: <title> and </title>. A link is created most often with the and tags. An image usually appears as <img src="image">. We can see the HTML file by opening a webpage as a text file. On Microsoft's Internet Explorer, select View→Source.

Tags (Field search punctuation)

To request a field search, we can select from the advanced search page of our preferred global search engine or we can use the normal search box of a search engine and simply precede our search word or phrase with a special term – sometimes called a tag. The title tag (usually **intitle:**), the URL tag (usually **inurl:**) and the link search tag (**link:**) are the three most important field search terms or tags. This can be confusing, since this 'title tag' (either **intitle:** or **title:**) in particular shares the same name as the HTML title tag (<title>).

Tags (as in a Thesaurus)

Tags can also mean a descriptive word attached to an image, webpage or website and used to organize information. Such tags work in much the same way as a thesaurus and are common to tools like Del.icio.us and Flickr – services that let internet visitors add descriptive tags to webpages and pictures. Meta-tags, a specific HTML tag, do a similar function but must be added by the publisher.

Thesaurus

In creating a commercial quality database, a definitive list of standard terms may be created to help organize the information. All information on juvenile diabetes, for instance, will use that term instead of childhood diabetes or early-onset diabetes. The list of all the preferred terms is called a thesaurus. In searching, we may reach for a thesaurus to uncover the preferred terms and concepts so as to search with greater precision and execute a comprehensive search. Commercial databases make great use of a thesaurus, unlike the internet.

Title Search

The title corresponds to the few words that appear in the upper left corner of a web browser. Only occasionally do titles describe the contents of a page. With the title field search, we can specify a word or phrase must or must not appear in the title. Generally, a title search is a fairly crude tool, too random to be a good specific search. To request a title field search, we usually type: **intitle:word**.

URL Field Search

This is the most significant internet field. The URL field search allows us to demand a portion of a web address appears or does not appear in our search results. We can use this to find local context, to restrict a search to a given website or to search for a word or concept in the address. To request a URL field search, we usually type: **inurl:portion_of a_web_address**.

URL / Web address

The Uniform Resource Locator, or URL, provides a unique address to information on the internet and also defines the tool used to access that information. A web address, for instance, starts with **http://**. There are other tools besides the web but given its voracious appetite, most information has migrated to the web. Thus, the difference in meaning between URL and web address is largely an artifact of history.

URL Interpretation

The URL or web address holds much meaning. It may indicate a date, format and likely publisher. It may suggest a quality and type of author. With practice, as we search we will read the URL, decipher many of its qualities, then decide if we wish to visit.

Wikipedia

The Wikipedia is firstly, a multilingual free encyclopedia at **wikipedia.org**. Part of the Wikimedia Commons, it is at the forefront of the open source information movement where knowledge is held in common. Anyone can add information to the Wikipedia, which in practice means recent additions are of uncertain factual value, peers and visitors tend to correct and improve information with time and popular culture is covered in great depth. As an encyclopedia, I find the Wikipedia near or above the quality of commercial encyclopedias and improving.

NOTES

[1] 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.

[2] Subscribe to her newsletter via ResearchBuzz.com. Tara Calishain is prominent as co-author of Google Hacks (O'Reilly Media 2005)

[3] Information Today Inc Periodicals [www.infotoday.com/periodicals.shtml]

[4] Online Currents [www.onlinecurrents.com.au]

[5] End Of Size Wars? Google Says Most Comprehensive But Drops Home Page Count (SearchEngineWatch 27 Sept 2005) [searchenginewatch.com/ searchday/ article.php/3551586] Retrieved Sept 2006

[6] Tara Calishain and Rael Dornfest, Google Hacks, 2nd Ed (O'Reilly Media 2005)

[7] John Battelle, Google Announces New Index Size, Shifts Focus from Counting, *John Battelle's Searchblog* 26 September 2005 [battellemedia.com/archives /001889.php] Retrieved April 2007

[8] David Novak, Plagiarism Therapy [SpireProject.com/art23.htm]

[9] Country Guardian, The Case Against Wind'farms' May 2000 [www.countryguardian.net/case.htm] Retrieved June 2005

[10] British Wind Energy Association, BWEA corrects some misconceptions in The Case against Wind Farms [www.britishwindenergy.co.uk/you/cgcase.html] Retrieved June 2005

[11] This number compares mention of www.loc.gov to www.bl.uk. 5.8 million matches from a search for "www.loc.gov". 0.68 million matches for a search for "www.bl.uk". Retrieved using the Yahoo! search engine May 2006

[12] In regards to La Mer, to be precise, advertising copy states "Invented by a NASA scientist to treat burns on the face and eyes" and "Creme de la Mer was created by Dr. Max Huber, an aerospace physicist with NASA." We do not have certainty that Dr Max Huber worked as an employee of NASA when creating La Mer.

[13] John Mueller, The Iraq Syndrome, *Foreign Affairs* Vol 84 No6 (Nov/Dec 2005) p 45,46. John Mueller is a professor of political science at Ohio State University and the author of War, Presidents, and Public Opinion; Policy and Opinion in the Gulf War and The Remnants of War.

[14] Noam Chomsky, Writers and Intellectual Responsibility [sound recording]: an address by Noam Chomsky at the NSW Writers Centre 23rd January 1995, (Gil Scrine Films 1995)

[15] Danny Schechter, Weapons of Mass Deception [film] 2004. The Chicago Reader describes this movie as "A comprehensive and devastating critique of the TV news networks' complacency and complicity in the war on Iraq ... brilliantly argued and scrupulously documented" as reported on Welcome to WMD the Film [www.wmdthefilm.com]

[16] Daryl Gates, Frontline: L.A.P.D. Blues (Public Broadcasting Service (PBS)) February 27,2001. [www.pbs.org/wgbh/pages/frontline/shows/lapd/interviews/gates.html]

[17] Judge Jones of the US District Court for the Middle District of Pennsylvania ruled on December 2005 against a curriculum change made by the Dover Area School District. It rules that the inclusion of a four paragraph statement about Intelligent Design (ID) in the classroom was an unconstitutional breach of the separation of church and state. Kitzmiller Vs Dover Area School District 25 Dec 2005 [www.pamd.uscourts.gov/kitzmiller/kitzmiller_342.pdf] p 136,137

[18] Scientists, Teachers, Clergy Hail Court Ruling, *Inside Science News Service* (American Institute of Physics (AIP)) [www.aip.org/isns/reports/2005/023.html] Retrieved 30 Mar 2007

[19] Radio Free Europe/Radio Liberty, About RFE/RL [www.rferl.org/about/organization/radiostation.asp] Retrieved 30 Mar 2007

[20] Robert Nemiroff & Jerry Bonnell, About Astronomy Picture of the Day [antwrp.gsfc.nasa.gov/apod/lib/about_apod.html] Retrieved Dec 2006

[21] Robert Nemiroff, Re: Dear Robert. Question regarding NASA role in APOD, Private email to David Novak, 20 Oct 2005

[22] Marcus Aurelius, Meditations (Maxwell Staniforth Trans) (Robin Waterfield Abgd) (Penguin 60s edition, Penguin 1995) p 24. Of course, an internet version of Meditations exists [www.gutenberg.org/etext/2680] though with remarkably different text.

[23] The New Powder Keg in The Middle East, *Nida'ul Islam* 15 [www.islam.org.au /articles/15/LADIN.HTM] Retrieved 2000. Though no longer online and no longer held as before in the Internet Archive, you can find a copy by searching the web for the title: "The New Powder Keg in The Middle East".

[24] Christian Missionaries in the Muslim World - Manufacturing Kufr, *Nida'ul Islam* 20 [Islam.org.au] Though no longer online at Islam.org.au, copies can be found on internet by searching for the title.

[25] The continuing saga unfolds through pages of WikiNews [en.wikinews.org/ wiki/Special:Search?search=wikipedia+class+action] though has recently toned down. For example, Wikipedia class action site vanishes, backers revealed (WikiNews) en.wikinews.org/wiki/Wikipedia_class_action_site_ vanishes%2C _backers _revealed] Retrieved 1 Apr 2007

[26] Searched on AlltheWeb, April 23, 2006. These numbers change radically over time.

[27] Noam Chomsky, The Prosperous Few and the Restless Many (Odonian Press 1994) p 40

[28] The Washington Institute for Near East Policy (Columbia International Affairs Online (CIAO)) [www.ciaonet.org/pbei/sites/winep_policy2006.html]

[29] Neoconservatism (Wikipedia) [en.wikipedia.org/wiki/Neoconservatism] Retrieved Mar 2007

[30] Joel Beinin, US: the pro-Sharon thinktank, *LeMonde Diplomatique: English Edition* July 2003 [mondediplo.com/2003/07/06beinin]

[31] See Reference: Education: Instructional Technology: Evaluation (Open Directory Project) [dmoz.org/Reference/Education/Instructional_Technology/Evaluation /Web_Site_Evaluation/] for alternative approaches to quality assessment.

[32] "Executive Directors expressed concerns over the misreporting of fiscal data to the IMF between 1997 and 1999. ... Pakistan's budget deficit for 1997/98 was revised upward by 2 percent of GDP to 7.5 percent of GDP, and the deficit for 1998/99 was raised by 1.4 percent of GDP to 5.9 percent of GDP." News Brief: IMF Executive Board Reviews Pakistan Misreporting, Remedial Steps *International Monetary Fund News Brief 00/23*, April 28, 2000 [www.imf.org/external/np/sec/nb/2000/nb0023.htm]

[33] David Novak, Manipulating Forms to Improve Research, *ONLINE*, July/August 2003 (Information Today) This article also appears with minor corrections at SpireProject.com/art20.htm

[34] Sunny Worel, Automatic MEDLINE Searching: Integrating Medical Information into Web pages, *EContent*, 22, Aug/Sept 1999, p38-42

[35] Gary Price, ResourceShelf [resourceshelf.com] as well as Gary Price and Chris Sherman, The Invisible Web: Uncovering Information Sources Search Engines Can't See (Cyberage Books 2001)

[36] J.R Calladine, K.J Park 2, K Thompson & C.V Wernham, Review of Urban Gulls and their Management in Scotland (Scottish Executives 18 May 2006) [www.scotland.gov.uk/Publications/2006/05/18113519/0]

[37] www.romagiubileo.it/en/index_homepage.html

[38] Alain De Botton, How Proust Can Change Your Life (Picador 1997) p 48-49

[39] Jello is the US name for a gelatin-based desert. It does not nail to a wall.

[40] Gentium is a non-commercial script released under the SIL Open Font License. Started by Victor Gaultney with more recent assistance by Annie Olsen and others. See scripts.sil.org/gentium

[41] The Theseus Project was among other things, a royalty payment system that attracted seed funding from the University of Western Australia (UWA) in 1995.

[42] Gary Wolf, The Curse of Xanadu, *Wired* Jun 1995 [www.wired.com/wired/archive /3.06/xanadu.html]

[43] The Age, News Store [newsstore.theage.com.au/apps/help/qa.ac?sy=age] Retrieved 13 Oct 2006

[44] The New York Times Company, TimesSelect [nytimes.com/products/ timesselect/overview.html] Retrieved 11 Feb 2006

[45] David Novak, Commercial Information, Publicly Funded [SpireProject.com/art01.htm] This webpage describes Medline's dramatic transition from a commercial to a popular public-supported database.

[46] David Novak, Census of Regionally Important Web Documents [SpireProject.com/past/census.htm]

[47] In 2002 I estimated the internet size at 15 to 20 billion webpages and estimated coverage would fall below 5% by now. David Novak, 10 Billion pages or more. [SpireProject.com/art13.htm]

[48] David Novak, Evolution of Internet Research: Shifting Allegiances, *ONLINE* Jan/Feb 2003 (Information Today). A text-only version of this article resides at SpireProject.com/art19.htm

[49] The Tomb of Meryneith/re [www.let.leidenuniv.nl/saqqara/Excavation/Tombs/Meryneith/Meryneith.htm]

[50] The Australian Interlibrary Loan and Document Delivery Benchmarking Study was completed in Oct 2001 and while dated, does quantify the library's support for Interlibrary loans and document delivery. National Resource Sharing Working Group, Interlibrary Loan and Document Delivery Benchmarking Study (Oct 2001) [www.nla.gov.au/initiatives/nrswg/benchmarking.html]

[51] Dale Spender, "The Last of the Print Proficient" in A History of Information Science 1945-1985, Dorothy Lilley & Ronald Trice ed. (Academic Press 1989) p169

[52] Manuel Castells, The Internet Galaxy (Oxford University Press 2001) p 151

[53] Nicholas G Carr, IT Doesn't Matter, *Harvard Business Review* (May 2003) as it appears in The Tech Advantage is Overrated, *Businessweek* August 25th 2003

[54] The original webpage is no longer online but the survey is called the Singapore Mental Health Survey. A more recent survey is described on sma.org.sg/smj/3906/articles/3906a4.html

[55] Google search for "David Novak" August 2007

[56] Yahoo!, Yahoo! Directory Submit Cost [https://ecom.yahoo.com/dir/reference/cost]

INDEX

.htm v/s .html, 204
.pdf, 27, 137
.ppt, 27
+/- symbols, 18–20

A

AdSense/AdWords, 236
advanced search page, 24
agreement in synthesis, 98
allinurl field search, 25
AlltheWeb, 27, 114
Amazon.com, 233
amoeboid searching, 174
Amsterdam Digital City, 258
Annual Egyptological Bibliography, *252*
art appreciation metaphor, 303
Ask.com, 60
AskNow! virtual reference, 135, 241, 277
associations, 183–84
Astronomy Pictures of the Day (APOD), 94, 114, 190
attention. *See* Chapter Seven
attitude, 299
Aung San Suu Kyi, 108

B

B4 bookmarklet, 159
BankWest, 238
Bernadette Robinson, 55
Beyonce Knowles, 109
bias, 66, 309
    reveal, 75–76
biography
    finding, 93
    providing, 75
blunt search, 59

book format, 140–41, 231, 235
bookmarklets, **160**, 309
Boolean, 18–22, **18**, 309
Borders, 72
British Wind Energy Association (BWEA), 58, 99
Bruce Pannier, 96
Bruno the cleaver metaphor, 87

C

Camas magazine, 260
census of internet material, 243
cheapest flight to Paris, 265
chef metaphor, 30
choreography. *See* Chapter Ten
Christ in Piss, 145
CIA World Factbook, 32
commercial-quality databases, 20, 23, 187–88, 231
competition, 154, 251
comprehensive, 63, 100, **277**–**85**, 278, 280
computer science, 38, 39
connoisseur of information, 7, 97, 118, 304
context, 5–6, 128–29, 309
    local, 101–8
context bookmarklet, 158
context/format/source, 127, 149
controversial search, 279, 283
copyright of this book, 333
country code, 201
Country Guardian, 58
country profiles, 32–35
credentials, 148
critical search, 279, 281
currency, 82–84
cynicism, 119, 299. *See* attitude

D

Dale Spender, 254
Danny Schechter, 94
Danny Sulivan, 44
Dean Gates, 64
deceit, 76
decision making skills, 297–98
deep URL interpretation, 199–203, 210, 310
definitive. *See* comprehensive
definitive search, 280
depression example, 272
detailed search, 279
Dewey decimal number, 22
difficult search, 279
directories, *184–86*
diversity in synthesis, 98
domain, 201
Dr Rene van Walsem, *252*
driftwood metaphor, 70, 301

E

Egyptian Step Pyramid of Saqqara, *252*
Electronic Frontiers Foundation (EFF), 131
elevated vista, 6, 35, 251, 274–76, 282, 290, 310
embedded forms, 162, 310
endorsements, **116**, 310
    absence of, 111
    bookmarklet, 159
    recourse to negative, 113
    three types of, 110
English-language newspapers, 182
enjoy the hunt, 295
ERIC (Education Resources Information Center), 22
Eugenie Scott, 90, 93
Evolution of Internet Research article, 250. *See* SpireProject.com/art19.htm
expert networks, 276

F

fame. *See* prominence
Family and Children's Services WA, 240
fax format, 139
feedback, 42, 212, **215–16**, 310
field search, **22–28**, 311
filename, 204
filetype, 28, 137

footnote script, 164
footpath, 266–68, 273, 311
Foreign Affairs, 77
format, **137–40**, 311
forms, **160–62**, 311
frustration, 35, 36, 40
ftp format, 138
fuzzy AND, 19

G

Gary Price, 164
geographical markers, 59
geography, 181–83
good company, 71
Google Groups, 45
Google Hacks, 45
Google Toolbar, 52, **173**
government hierarchies, 179–80
Goya, Francisco de, 145
Greenpeace, 110

H

hackers ethic, 227
hacking the web address, 107, 208, 272
halo of supportive detail, 127, 242
Hamurrabi's Code of Laws, 145
hard refresh, 165
haste. *See* Chapter Five
hidden subsidies, 253
history, of the internet, 5, 243–46
hits v/s visits, 311
holistic approach to information, 41, 155, 153
holocaust denial, 89
HOSTS file, 166
HTML, Hypertext Markup Language, 160
HUDusers group, 187
Human Rights Watch, 35

I

identity, 312. *See* Chapter Four
importance, 60–61, 312
imprecise, 216–17
in-bound links, 26
information
    as a message, 153, 289
    as an object, 289