

The Hitch-Hikers' Guide to Google

Charles Weir and James Noble

Version 30/04/2003 19:49-CAFAW

Abstract

The web is big. Really big. This is good because the Web now encompasses the entire history, culture, science, and achievements of the human race, not to mention JPEGs of every romantically eligible member of that race under the age of 35. This is bad because it is now too big to find anything. The 'Hitch-Hikers Guide to Google' is a collection of patterns that will be your guide through the Internet. By reading these patterns, you can make friends, improve your lifestyle, and experience excitement, adventure, and really wild things. Share and Enjoy!

Introduction

Virtually every fact, reference, quotation, play, film, song, concerto, or computer program component that you may ever want or need has already been written, produced, sung, composed, or hacked together from bits of other programs – and put on the web. This is good because it means you don't need to go to the bother of writing, producing, singing, composing, or hacking ever again. Rather, the crucial skill for the post-modern playwright, producer, DJ, composer, or programmer is to find and incorporate all this stuff that has already been done.

These patterns will help you gain this key skill, thus improving your lifestyle, job prospects, and personal happiness.

Form

The patterns movement is obsessively concerned about the form of the patterns that appear in its workshop, often to the detriment of the content of those patterns. These patterns carefully follow that principle. We have tried lots of different fonts, heading styles, dingbats, and even different spellings of "Theyrefoir", to arrive at this form. We hope you enjoy it!

Forces

Nobody forces you to use Google, but here are some things to consider as you take such a life-enhancing decision.

- *Precision* (technical term from information retrieval: fraction of pages found that are relevant). High precision means that the page you want will be at the top of the list of the pages returned.
- *Recall* (technical term from information retrieval: fraction of relevant pages that are found). High Recall means that the page you want will be somewhere in the list.
- *Interface Complexity* - how much you need to learn about the browser or website to use this pattern
- *Local configuration* – whether or not you need to change your computer, possibly crashing it, to use this pattern.
- *Browser dependent* – whether which browser you use makes a difference to this pattern

- *Navigational clicking* – whether you need to click around a lot to find what you're looking for.
- *Domain Knowledge* – whether you need to know about what you're looking for in order to be able to find it (see librarian's paradox below).
- *Effort Required* – how much effort you the user have to put in yourself to get the results you need.

Pattern: Google

The web is big. Really Big. You just won't believe how vastly, hugely, mind-boggling bit it is. I mean, you may think it's a long way down the street to the chemist, but that's just peanuts to the web.

The good news, gentle reader, is that the size of the web means that it now encompasses much of the sum total of human knowledge about every important subject ever - especially if that subject is pictures of Britney Spears or Pamela Anderson. In fact, a strange Californian sect believes that the web has now achieved sentience [1] and will soon begin creating brand new pictures of Britney Spears, Kylie Minogue, and Pamela Anderson purely for its own, silicon-based enjoyment.

Unfortunately, although the web does contain all knowledge from Aardvarks to Zwingli, and including the session times at your local movie cinema, finding this information is worse than finding the proverbial needle in the haystack, whelk in the supernova, or black cat painted black all over and hidden in a black room with no lights.

What, then can you do to find the information you need on the web? This is an example of what we'll call the librarians' paradox: if you know where information is, you can just find it, but then you don't need to look for it, do you? But if you don't know where it is, well, then, how do you know where to look? Apparently there are buildings in many places in the civilized world (and also in America) that not only hold large quantities of information, but also people who know where it all is (the librarians of the above mentioned paradox). Sitting in one of these buildings meditating on the futility of life, the universe, and everything, you could always get up and ask one of these librarians for help, whereupon they will tell you a) the book you want has been stolen; and b) please keep quiet. And in the earth-shatteringly unimaginably unlikely event that they do turn up the book they've been looking for, you find that the information is out of date.

So, the web is the last, best hope for all of humanity. But how can you find what you want on the web without really trying?

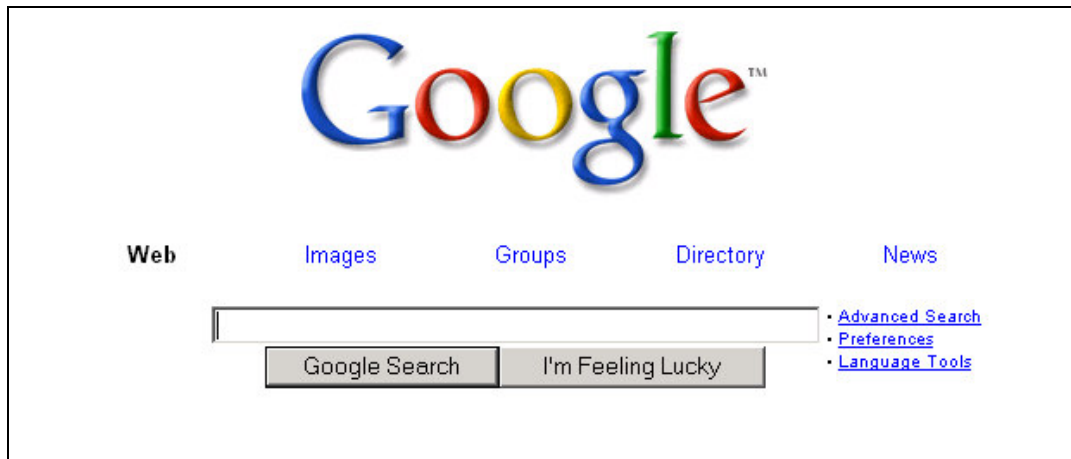
Therefore: *Use Google!*

Google is the world's best search engine. Google's technical specifications are amazing. Google has a huge database, a patented page-ranking algorithm (the right one: pages are ranked on how often other pages link in to them); fast access because it runs on lots of big computers, and a very simply user interface (one text box) and a cool logo (see table below).

Database Size	Huge
Ranking Algorithm	Patented
Search Time	Really Really Fast
Hardware	Lots of Big Computers
User Interface	One Text Entry Box
Coolness of Logo	Cool

Google's Technical Specifications

You don't need to worry about any of these technical details: the upshot is that if you type some text into Google's box, it will find what you want.



Consequences

The most important result of everybody using Google all the time (anybody who is anybody that is: anybody who is not using Google all the time is clearly not anybody) is the Google Effect (named after Professor Effect who first studied this phenomenon). Anybody who wants their web site to be found does everything that they can do to ensure that Google finds their web site. Thus, anybody wanting to find anything can always find it on Google. Furthermore, because Google can search the entire web in the time it would take you to go down the street to visit the chemist; Google can also search any other search engine or directory on the web. So, anyone who tries to set up a new search engine in competition with Google only makes Google stronger¹.

Google has good recall because it indexes the entire web. Google's Interface Complexity for simple searches is low (you just type text into the box and hit return). A simple search on Google requires no local configuration and is completely browser independent. On the other hand, Google's precision is not as good as its recall (Google can often find sites that are irrelevant to your query), and choosing a good query can require some domain knowledge.

Implementation

You may need some tricks to deal with the transitory and multi-cultural nature of the web:

- Use the cache if a likely looking page isn't there.
- Read it translated – into HTML from PDF, or into a language you understand.
- If the page doesn't seem to match your search, use your browser's 'search within page' facility to check whether the keywords are still there.

¹ Kind of like Microsoft. In fact, why hasn't Microsoft just bought Google yet?

Pattern: On The Button

ARTHUR: ... What happens if I press this button?

FORD: I wouldn't...

ARTHUR: Oh!

FORD: What happened?

ARTHUR: A sign lit up saying 'Please do not press this button again'.

Google is so cool that you'll want to use all the Google, all the time. But, even though searching with Google is quicker than a Texan president organizing an invasion, you'll need to go to the Google many times when you're trying to find out anything. You often have to load two pages each time you want to search for something: first the main Google page, and then the page that lists the search results. This seems slow: the main Google page is the same every time (except when it's Christmas), only the results page has anything really useful (that is, results). Also, how can you let your friends know that you are using Google, and not some other, non-Google search engine?

Therefore: *Decorate your browser with Google Specific features*

Typical of these are the so-called Google Buttons. These put a Google Button in the button bar. Pressing this button brings up a popup window with a text box in it. When you type a question into this text box, it is sent directly to Google, and the results page returned directly to your browser.

How you attach buttons to your browser depends on the kind of computer you have and which browser you run. Navigation to the "Google Tools" link from the main screen will let you choose to install whichever features you want.

Consequences

Installing Google's buttons is browser specific; they won't work on all browsers, and some browsers pretty much have them built in. Adding them requires a local configuration change to your browser, and this may crash it or your whole computer.

The interface to add features to your computer can be complex and require lots of navigation clicking to choose the right feature; on the other hand once you've fitted them you can get to Google more quickly and with less navigation clicking.

Examples

Some web browsers try so hard to be cool that they include Google buttons built in. For example Netscape has a "Search Google" feature on a pop-up menu when you type in the URL box. Opera is even cooler in that it allows you to type the search directly into the URL box by prepending it with the letter "g" (as in "g Google buttons"). Safari plumbs the depths of coolness with a special entry box that goes straight to searching on Google².

Of course there are other ways of making it easy to start a new Google search as required:

- You can make Google your browser home page. That way you will always be at home using Google.

² Of course, Mac OS X is by far the best operating system in the known universe [7].

- You can have Google as the first of your browser bookmarks. Charles uses the wonderful PowerMarks bookmark tool [8], sorted by how often he's accessed each; Google is always far and away the first.
- Or you can have a link to Google on your desktop.
- Or, if you're inspired to software development, you can even use Google as a web service from an application ...

Pattern: Nutrimatic

NARRATOR: One of the Sirius Cybernetic Corporation's creations is the Nutrimatic Drink Dispenser. One of which has just provided Arthur Dent with a plastic cup filled with a liquid that is almost, but not quite, entirely unlike tea. ...

NARRATOR: The way it works is very interesting. When the 'Drink' button is pressed, it makes an instant, but highly detailed examination of the subject's taste buds, a spectroscopic analysis of the subject's metabolism, and then sends tiny experimental signals down the neural pathways to the taste centres of the subject's brain to see what is likely to be well received.

However, no one knows quite why it does this because it then invariably delivers a cupful of liquid that is almost, but not quite, entirely unlike tea.

ARTHUR: I mean, what is the *point*?

So, you've selected Google, clicked on your Hot Button, selected a special Google window in your browser, or whatever. Unfortunately, you are now faced with a smallish, blank text area, inviting you to type in some words to search for (we'll call these words the *keywords*). All you have to do is type the words in to tell Google what to search for! But what words should you choose? If you pick the "right" words, Google will find just what you want, right? If you type the wrong ones, your boss will be able to look in your browser cache and see that all those pictures of Kylie Minogue were *not* really found by accident.

So, what should you type to Google?

Therefore: *Type keywords that are found only on the page you want —and nowhere else!*

This is the essence of using Google. You type some words into the box and press return. You look at the results, and then try again, typing more – or different – keywords into the box. Eventually you will find some pages that seem that they might be useful (or you get bored of just poking Google) and you can use the Honeybee pattern to explore those pages.

Consequences

You find the page you want, increasing precision and recall. However, you need good domain knowledge to choose the correct keywords.

Implementation

Choosing Keywords to Start

How do you choose the keywords to start with?

For a first cut – phrase the search as a question you might ask to a friend. Or even better – actually ask an (imaginary) friend the question, and type in the words you used. Make the question as specific as possible. For example, you could type in "presidents of the USA" to try to find list of all the presidents of the USA.

Alternatively, you could guess the *answer* to the question, and type that in. The catch is that this raises the librarian's paradox again: for this to work, you do need to *know* the answer you want in advance. Typing "Andrew Jackson X X" isn't particularly good, but "George Washington Bill Clinton" does provide good results.

If the information you want is related to a phrase, type in the whole phrase in quotes. For example, if you want to know how to solve an MS Wind-up error message, type in the error text in quotes. If you want the words to a song, type in a distinctive phrase from that song.

Narrowing the Search

What should you do when you get too many irrelevant pages? You'll need to specialise the search. And you can narrow the search in several ways:

- Use more specific keywords (e.g. "titanic" instead of "ship")
- Localise geography, by using your local Google Site and clicking the box to select pages only from your own country. This isn't very reliable as a search criterion since Google doesn't really know where all those '.com' domains are, but it can help a bit.
- Limit by language. Do you really need all those pages in languages you can't speak?
- Limit by time. You can use the advanced search engine facilities to specify only pages modified within the last few months.
- Try a search using keywords in a different language. For example "Zeitgeist" will give a more specific search than "spirit of the age".

In conjunction with the Honeybee pattern, you can look for pages that seem relevant – or simply interesting. Then you can scan those pages very carefully for more specific keywords, either to include in further searches or to exclude because they are used very commonly.

Example

For example, let's imagine you want to find out about Victoria University of Wellington in New Zealand. Let's see what happens if you type a range of different keywords

The table below shows a choice of searches, the number of results reported from each search, and (where it is not too far down the list) the position of the first web page from Victoria University of Wellington in the list.

Search	Number of Results	Position of Correct Result
University	984	527
Victoria	916	11
Wellington	899	4
Victoria University	745	3
Victoria Wellington	827	1
Wellington University	817	1
Victoria University Wellington	811	1

The problem is that there are lots of Universities, lots of Victorias, and quite a few Wellingtons in the world: there is even another Victoria University (in Melbourne).

See Also

The Honeybee pattern tells how you to move between Google searches and other pages to quickly find the page you want.

Pattern: Honeybee

Let's go back and consider the librarian's paradox again: how can you find out what you need to know if the only way you can find it is if you know what it is! It's a big problem! If you know what you want, then it's really easy to find it; if you don't know, then how can you possibly start?

And yet there's one particular thing about the Web that can help you. Like the human brain, it's a series of links from one concept to another. So whatever it is that you want to know or, you're sure to be able to find it by following concepts and ideas from one page to another. But how do you find such chains of concepts? How do you avoid getting lost in logical dead ends?

Therefore: *Use the Google main page as a base; go out and explore different ideas, and use what you learn to improve the your next Google search on the same topic.*

The process of exploring the Web using Google is very like the daily work of a honeybee. A honeybee starts from his home base, and goes off to find likely sites for pollen; if he finds any, he marks the direction and then will use that as the future direction to go for further pollen. If any direction proves unsatisfactory, then he simply writes that direction off for the time being. We can note two things in particular:

- 1) He remembers the direction that was most successful. Indeed some species of bee have complicated dances to tell other honeybees which way to go.
- 2) He's happy to follow any likely looking path during his exploration.

The practicalities of this approach using Google are very simple. Start with an initial exploration using some suitable keywords. Look at the search results, and follow the likely looking links. The resulting pages may be relevant; if so, study them to find more suitable keywords to narrow your search towards its target. If the pages are irrelevant, return to the cool and try another page, or another surge.

Consequences

Using the Honeybee pattern to get to your target means that you don't need to find the one perfect set of keywords that give the page you want.

You get a much more precise search, and will tend to recall much more of what you actually need. This pattern has no interface complexity, and no dependency on browser type or local configuration. However it does require significantly more user effort, navigational between pages, and requires the user to build up her own domain knowledge.

Implementation

The main tool for this is the browser 'Back' button. I personally regularly use the keyboard shortcut for this (the "Delete" key), as it tends to be quicker. Most browsers also provide a "Back History" (by pressing and holding down the "Back" button, by a small popup menu next to it, or by a separate menu elsewhere) that can take you back more than one page.

Another approach is use the "Open in new window" feature. Or better still to use Opera's "open new window in background". This feature allows you to open the links Google provides in the background, click on "Next", and then quickly cycle through all the resulting web pages.

As you get to one page, you'll often get more benefit by following the links from there onward, than by returning to Google. This is particularly true when you know the site where

your information is likely to be - unless the site is really complicated and Microsoft-ian. Particularly good sides are Portals and lists of links related to a particular subject. From these you can usually explore a new subject in some depth. If you're looking for a particular organisation, then Google, Yahoo!, or AltaVista provide category indexes which can be extremely helpful.

Of course, like the bee, it helps if you can bookmark the sites that work best for you. Web searches are really a process of stochastic optimisation: you try solutions at random and select ones that seem to be going in the right direction³. So you have to risk moving on from a likely page to find a better one. But you can mitigate the risk by keeping a bookmark of the page you're leaving.

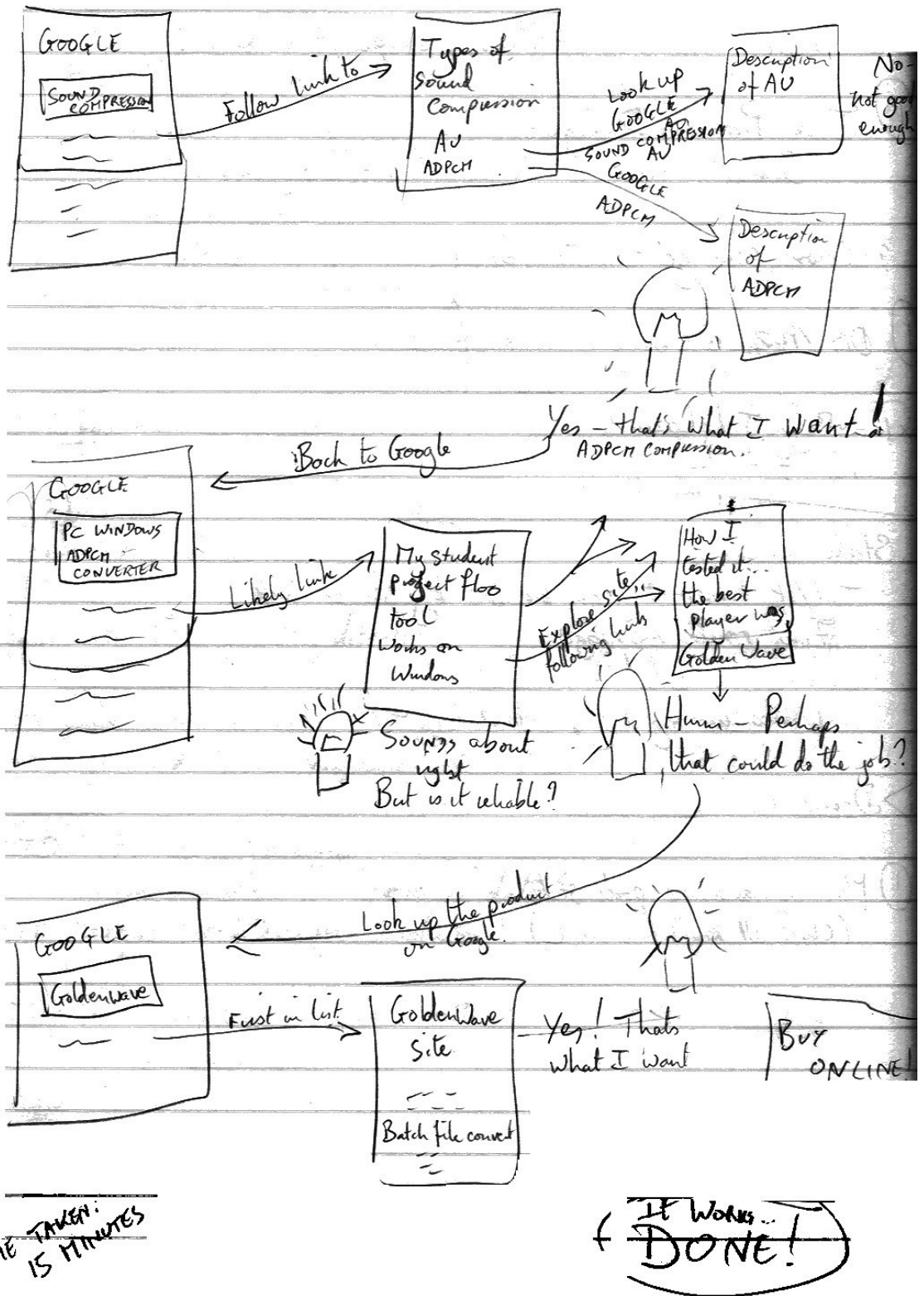
Example

The following illustration shows a recent search. We needed some way to compress a directory of WAV-format sound files to occupy less disk space (see the *Compression* pattern in 'Small Memory Software' [3]).

The picture shows the pages we found, and the process of iteration back to Google.

The whole search took less than 10 minutes. Add 5 minutes to buy and download the software package we'd found (Goldwave), and run it on the directory, and the problem was solved in 15 minutes!

³ Genetic Algorithms work in this way.



See also

Pick-n-Choose explains how you can evaluate each page you reach. Chef-O-Matic provides some guidelines how to pick keywords.

Pattern: Pick'n'Choose

So by following links and doing web searches, you can expect in due course to focus in on the page you're need most. Of course simply following links at random will get you there eventually, but rather in the manner of a series of monkeys trying to type Shakespeare's Macbeth. Yet unlike the monkeys, you have the ability to be selective about which links to follow up and which pages to take more seriously.

Therefore: *audition each page, and only take good pages seriously.*

Taste, or audition, each page that looks promising. The process should only take a moment or two for each page. You're looking for authoritative pages and pages that are likely to lead you into other useful pages.

Pages that are not authoritative may still be useful. Look for ideas from any page, information from reliable pages.

Consequences

You'll get more search Precision, since you're quickly rejecting unsatisfactory options. You'll also achieve a better recall, and may accumulate some domain knowledge just by the strange process of osmosis the brain is capable of. However it does require a lot of Navigation clicking, and a considerable amount of User Effort.

Implementation

What are you looking for?

Here are some suggestions for what you may be looking for:

Keys!	When you seen the kind of thing you're looking for, look at the words they use. Often they'll call it something different from you.
Experts!	Some pages contain links to other sites – you can treat these as recommendations for those sites; someone liked them enough to reference them. Or they might mention the name of an expert, or the best company in the area. If you've got that information then you'll know anything information on pages by those people is likely to be valuable.
Hubs!	Some pages – perhaps by students studying a field, or enthusiasts, or companies wanting to encourage traffic to their site – are lists of links to other sites. If the list is in the area you're interested in, you're in luck. Not only do you have a collection of suitable pages; you also have a validation for each.

Avoid pages that are very big; often these are simply indexes (and rarely lists of links). They contain all the words you want simply because they contain a huge number of words.

How can you evaluate a page?

Here are some suggestions to evaluate a page:

Is it from a domain expert? (Look at other pages to find out who the domain experts are.). Is it official? Is it from someone who's got a reputation to lose? E.g. For Taxation information, you might look first at government sites (because they're official): gov, gov.uk, etc. But more practical might be a site from a large accountancy firm. For health information, anything from prodigy.nhs.uk (the official UK health encyclopaedia), or the US National Institute of Health, is likely to be accurate.

Is the information recent? Many pages have dates that explain when it was last updated. More recent information is more likely to be correct than outdated information. A common warning for out-of-date information is a message ‘This page will be updated soon’ – particularly if it is dated several years ago.

Does it look professional? Is the page in ALL CAPS, containing lots of Blinking Text, or other stupid formatting, or written in H8K3R SP33K, D00DZ. Are there lots of spling miskates, bad formatting, or great big gifs (especially of Kylie Minogue or Britney Spears)? All of these indicate that perhaps this page should not be trusted.

Pages in a categorical index have usually been evaluated. If you hit a category page, you can look there also for other related pages.

If you’re looking for software, well, source code examples from Bell Labs are typically better than dummy applications knocked up by undergraduate students. Both reputations depend on it; but Bell has more to lose! If you want an application, then academic versions are typically not nearly so robust; often something you have to pay money for will be better – though free software that’s had the Gnu process of continual review and improvement by a large number of people is usually better still.

Pattern: Send in the Marines

As in the popular song “send in the marines/ there have to be marines / don’t bother, they’re here” [7]

So, armed with the preceding patterns, you are able to tackle any question or any subject in history, mathematics, geography, Divinity, reflexology or Googology. Although are you? Some topics seem to simply defy the web. Sometimes this is because putting the information together costs money, which doesn't fit well with the zero cost philosophy of the web. Others simply may be too specific or too old fashioned to be covered appropriately yet on the web. So what should you do?

Therefore: *Be prepared to use other knowledge resources where appropriate.*

Google doesn't replace every other knowledge source; it merely adds to existing sources. So be prepared to use those existing sources as well.

Consequences

You end up better educated, and certainly better informed. However many of the alternative sources are more expensive and, and most require a more user effort.

Implementation

Here are some suggestions what to do when Google fails you...

- Ask someone (particularly someone who knows).
- Use another way to get to the place you want. Charles keeps a list of URLs using the wonderful tool PowerMarks [8]. There are around 1000 by now, indexed by keyword. If, for example, he wants to know the times at his local cinema it's much quicker to follow one of those.
- Use one of the question services – some answer for money. Or Usenet or CIX. It takes a while, but is likely to get a more precise answer.
- For a purchase, go to a specific shopping site. These don't always index all their pages, so you can often find them more easily by searching the site specifically. You can buy books listing good shopping sites [6] – or, of course, search for them using Google!
- To learn about a new subject, or for well-tailored information – or indeed in any case where the information is worth money and lots of people will want it – the chances are that there's a book about it, and you'd do well to buy it. Use Google to find references to good books, or Amazon book reviews and rankings to determine which are best.
- Or if it's database style information, perhaps you need a specific database; addresses and telephone numbers come better from the directory lookup sites (most are free).
- Finding a person again is a more specialized search. Looking for “John Smith” is a daunting process, and the John Smith you want may not have his own web presence anyway. You might use some context: for example if you know John Smith worked for ZZZZ inc, then there's a good place to start. Better, depending on your needs, are friendsreunited.com for social kinds of search or telephone lookup if you know the location. We've never had much success using email directories like Bigfoot, but perhaps we just don't have the knack!

Examples

Most detective novels have examples of this kind of search. *The Name of the Rose* [5] is particularly good on libraries!

Pattern: Forty-Two

What do you get if you multiply six by nine?

Sometimes Google can't (or won't) find whatever it is you are looking for. Perhaps you search for a function to compute all the prime numbers and you get directed to Dijkstra's Prime Rib Bar, Peoria, IL. Perhaps you search for a suitable investment scheme and get directed to MAKE MONEY FAST by investing in Nigerian currency futures. Perhaps you search for instructions for keeping chimpanzees in captivity as pets and get directed to The Whitehouse. Perhaps you've tried all the patterns in this paper, climbed these city walls and those city walls, and still haven't found what you're looking for.

Therefore: 42

Google is the embodiment of the Mind of the Web. The spider sitting quietly waiting to pounce. The destroyer of worlds. Perhaps Google knows better than you do. Perhaps Google has answered the next question that you asked. Perhaps you need to eat, now, in Peoria, Illinois. Perhaps Google has answered the question you would have asked, had you known enough. Perhaps Google has answered the ultimate question, of life, the universe, and everything.

All life is linked. All knowledge is valuable. Be open to other questions. The answer to yesterday's question is an apparently unrelated search today.

Consequences

You achieve Enlightenment.

Acknowledgements

Lots of thanks to Pascal Constanza, our EuroPLOP shepherd for this paper. Quite a few thanks to Peter Sommerlad, our supervising PC member. Some grudging thanks to whoever wrote Google, so that we could write this paper. Not really many thanks at all to whoever designed the character style feature of Microsoft Word 2000.

And very many thanks indeed to Douglas Adams, who in some very small way is commemorated by this paper.

References

1. Santa Fe Center for the Study of Complexity. <http://www.santafe.edu/>
2. James' Post-modern paper
3. 'Small Memory Software', by James Noble and Charles Weir. Addison-Wesley 2000.
4. The Hitchhikers Guide to the Internet 25 August 1987 Ed Krol
<http://www.ifla.org/documents/internet/kroel.htm>, or (more readably)
<http://www.softlab.ntua.gr/miscellaneous/Internet/hitchhikers-guide/thgtti.html>
5. The Name of the Rose, by Umberto Eco
6. Virgin Guide to Internet Shopping.
7. Send in the Clowns, by Stephen Sondheim. E.g.
<http://lyricsplayground.com/alpha/songs/s/sendintheclowns.html>
8. Powermarks Home Page: <http://www.kaylon.com/power.html>
9. Douglas Adam's Guide to the Macintosh. Mac User "The Macintosh Resource" September 1987 page 144. <http://www-personal.umd.umich.edu/~nhughes/dna/stories/guidetothemac.html>
10. Pushing back: evaluating a new behaviour for the back and forward buttons in web browsers
Andy Cockburn, Bruce Mackenzie and Michael Jason smith. *International Journal of Human-Computer Studies*, Volume 57, Issue 5, November 2002, Pages 397-414