

2 BECOMING A SEARCHER

2.2 SEARCHING THE WEB

GOAL

Participants will develop more effective and efficient approaches to searching for information on the World Wide Web using a popular search engine (Google) and directories.

OBJECTIVES

Participants will:

- Understand the how the most popular Web search engine, Google, ranks results and its principal search features
- Be able to enter sound search statements and have strategies to refine these to increase the relevance of results
- Develop awareness of an alternative approach to finding relevant websites through Web directories and when to use this in preference to free text searching.

NEEDS ASSESSMENT

Trainers will ask potential participants PRIOR to attendance:

- Do you regularly use Google to search for information on the Web?
- Does Google provide reasonable coverage of NZ websites?
- How would you search for a book title on the Web?
- Is it necessary to put an 'and' between 2 or more words you are searching for on Google?
- How would you bring a NZ focus to a Google Web Search?
- Can you name a NZ directory of websites that you have used?

If the trainer is confident in the potential participants' demonstrated ability and understanding of these questions, then they may not need to complete the total session - they could commence with the Web directory exercise, then move on to s2.3 *Searching Databases*.

2.2 TRAINING CONTENT: SEARCHING THE WEB

DURATION OF SESSION: 55 MINUTES

Content	Duration	Key Points	Practical												
Session Objectives	3 min	Context (relation to earlier sessions) Core strategies and features of Web (Google) Use of directory approach													
A Scenario (The 'context')	2 min	Start with an information need: A Y12 student has chosen to research the topic "the effects of violent video or computer games on young people" This is major piece of work and she is required to find and cite a range of reputable sources - 'quality' Web stuff, journal articles etc. - and see how the issue is treated in news media. She has done a bit of searching using Google but become confused by the diversity of material and turned to the library for help.	<p>[Analysis of stated need]:</p> <p>Topic analysis is not used in this scenario. In another it may be useful to elicit possibilities from participants eg. <i>Ask participants to suggest useful terms:</i></p> <table border="1"> <tr> <td>young</td> <td>violent/violence</td> <td>games</td> </tr> <tr> <td>youth</td> <td>aggressive</td> <td>computer games</td> </tr> <tr> <td>teen</td> <td></td> <td>video games</td> </tr> <tr> <td>kids</td> <td></td> <td></td> </tr> </table> <p>Recognise that there are many ways of expressing topic but decide most likely initial keyword search could be: young+violent+games (as 'games' often shorthand for video etc. in the context)</p>	young	violent/violence	games	youth	aggressive	computer games	teen		video games	kids		
young	violent/violence	games													
youth	aggressive	computer games													
teen		video games													
kids															
Google Features	3 min	Google's base features: <ul style="list-style-type: none"> Words searched on in bold Ranking by relevance, with sites that are linked to by many others and assessed as 'quality', given higher weighting A loose search may turn up a few good results but typically will be very mixed. 	Start with the students basic Web search on Google: <i>young violent games</i> Ask participants: Lets examine the search - How many hits are there? How does G. indicate which terms are being retrieved for each result? [in bold] How useful are the p.1 results? What sort of sites are showing up? Why does stuff like the <i>NY Times</i> get near the top?												
Putting Terms Together	4 min	Google always inserts 'AND' by default between any 2 or more concepts/terms - this requires all terms to be present in any page.	At this point the student takes a closer look and volunteers "I put a '+' in front of the violent and games to make sure both are included" Ask: How would you respond? Prompt: What does the 'and' actually mean in retrieval of												

		<p>[If asked] The '+' only required if want to include a specific 'stop' word that is normally not searched on eg. 'how'. It is not the same as a general 'Boolean' 'AND'.</p>	<p>sources?...look at bold words in the hits? [Has to include <u>all</u> search words in page not just any]</p>
Phrase Searching	3 min	<p>Phrase marking (using double quotes) is useful to add precision where is not already a common order of words.</p> <p>If the words are commonly expressed in this order then it wont change much – but very useful for searching for titles (of books, names, articles etc)</p>	<p>Ask: How can we make it a bit more specific? How are the terms 'violent and games' commonly expressed when brought together? [By doing a phrase search where this is likely way of expressing concept] eg. <i>young " violent games"</i></p> <p>Ask if Improvements? [Not much except in order]...why not?</p>
Alternative Expressions of a Concept	6 min	<p>Google has several approaches to providing for variable ways of expressing terms:</p> <ul style="list-style-type: none"> ▪ Use an 'OR' (in <u>caps</u>) to provide for an either/or alternative ▪ Use the similar words operator (~) to cover a concept with a number of related possibilities <p>The similar words operator has its limitations as not entirely predictable BUT is very useful where you want to quickly and roughly cover a range of possibilities in a category/concept where there is a ready range of equivalents e.g. <i>wind ~power, ~car safety...etc</i></p>	<p>Student asks "what if reports talk about kids, youth or teens instead of young - might be missing some good stuff?" Ask: what would you do to include eg. 'youth' ? [Could run different searches - not very efficient - or combine them somehow? Elicit use an 'or'...but how do we put that into Google?... [with caps]: <i>young OR youth " violent games"</i></p> <p>[If some use lower case 'or' then note no. of hits and 'grey' message under search box] Note: small increase in hits, some new stuff on p.1</p> <p>Ask: But what about 'kids' etc? Do we just keep adding terms with OR? [clunky!] There is an alternative! [Elicit/prompt the 'similar words' function, the tilde ~]. Try: <i>~youth "violent games"</i> and note what words searched on (in bold) on p.1-2? Ask what is happening...? [synonyms, related words brought in] ...But also what's missing? (eg. <i>teen</i>).</p> <p>Ask: so how useful is this ~ [fuzzy search, 'squiggle factor'] operator?...in what contexts could it be really useful?</p>

A NZ Focus	4 min	<p>A NZ focus can be brought by:</p> <ul style="list-style-type: none"> Using the radio button 'pages from new Zealand'...or Adding 'zealand' as a search term. <p>Which is preferred <u>depends</u> on the context, the requirements of the searcher.</p>	<p>The student says "there's some useful material here but can't we get more NZ sources?" Ask: How can we limit for this? [Elicit options: radio NZ button or add 'zealand'] Ask: which is the better option? [Neither - it always <u>depends</u> on what's required! We may need to do both]</p> <p>Compare [if time]: ~youth "violent games" + NZ radio button with ~youth "violent games" zealand</p> <p>Note: distinct results...some overlap but quite but different rankings.</p>
Definitions [Optional: include if time]	3 min	<p>The 'define' operator is a useful Google feature for providing technical or specialist definitions of terms in context, but is not country specific.</p> <p>Express it as - define:term (no space)</p>	<p>At this point the student starts think a bit more deeply about the topic and asks "Can you get me a definition of 'violence?" Can we?</p> <p>[Elicit 'define' feature] Try: define:violence</p> <p>Ask what is being retrieved here? How useful is it as a function? Are there any NZ definitions? [No, even with NZ button on]</p>
'Fill in the blanks'	4 min	<p><i>Quick answers to factual type queries can be generated by entering the hoped for answer (<u>not</u> the question) in quotes minus the missing bit. This known as 'fill in the blanks' approach.</i></p>	<p>Our student is impressed and is starting to want to fill gaps in her understanding. She asks "how long have video games been around - when were they first marketed?" Could you suggest a strategy to hone in very precise material or facts?</p> <p>[Prompt: "fill in the blanks type statements"]. ie. consider how the statement you are trying to find might be expressed and search on this as phrase <u>minus</u> the bits you don't know or aren't sure about. Try: "video games were first introduced"</p> <p>When would it be appropriate to apply this strategy? [When satisfied by a few, very select indicative references, or want to quickly establish a factual point]</p>
Limiting to different types of Web sources	3 min	<p>'Tabs' at the top of the search screen enable limiting of a search to a particular form of material ie. searching just a small, selective 'slice' of the Web such as illustrations (<i>Images</i>), online news media (<i>News</i>) or <i>Books</i> (excerpts from)</p>	<p>Student says, "Great, getting there - but can you also find me a few examples of how the issue is handled in recent mass media?" How can we provide this on the spot?</p> <p>[Elicit: click on <u>News</u> 'tab' top of screen] Try: young "violent games" + click <u>News</u></p> <p>Ask: What is it searching? [Newspapers, TV, agencies...]</p>

			What is it doing here? [Searching just a 'slice' of the Web resource – in this case news media only]
Limiting to selected periods or publication date	3 min	A recent development are options to filter existing results using the <u>Show options...</u> feature. This enables the slicing or sorting of results by time or form criteria.	But what if, the students says, "I want to view the results by date and compare earlier perspectives with current?" Prompt for 'options'? Click on <u>Show options...</u> then ask what might provide such a view? [<u>Timeline</u>]. Ask: how could we compare viewpoints in 1998 vs. 2008? [select for years from graph at top] Ask: what else could we do with the options? How do we return to the original search?
Recap	4 min	Consolidation [key step!] Tick off: Combining terms, alternative expressions, NZ focus, defining terms, limiting to specific sources....	Ask: based on this scenario what can we conclude about useful features and strategies for searching and in particular refining searches on Google? [Handout: 1 page Google reference checklist]
Web directories to browse for Web resources	15 min	As in a library some needs are may best met by browsing areas of interest rather than keyword searching. Key Directory points: <ul style="list-style-type: none"> ▪ may provide some assurance of quality or at least representative selection ▪ can save you (or a customer) from having to wade though pages of a search to select sites ▪ only a small number of total websites, may not be very current or not cover NZ sites well (excel. TPWD) ▪ need to be familiar enough with a subject to select the right options as you drill down ▪ for information on very specific topics a Google Web search is a better strategy, but if need just a few good sites on a 	1. Participants asked to go to Te Puna Web Directory (link from NLNZ site) and find a listing of websites 'organisations that might work with youth at risk'. [If required guide them to <i>Subject > Community and Social Studies > Children & young People....</i>]. Note how can move up/down in subject levels by 'breadcrumb' trail top left. Then ask them to try and do a quick Google search on the same topic (NZ button on) and compare results. (NB. TPWD link included in results). Ask which approach is the more comprehensive?...efficient? Ask if they think TPWD includes all sites on topic? And who/how selected? [note link to selection criteria if time] 2. Participants asked to go Google Directory (from <i>more > even more >... tab/link</i> above search box) and find listings of suitable sites on the 'Middle Ages for younger school kids (Y8-9)'. [If required guide them to drill down through <i>School Time</i> category ...> <i>Social Studies...</i> + note breadcrumb

		<p>broad topic, to familiarize yourself with a topic, or find ideas for query terms then directories can be a useful option.</p>	<p>navigation]. Note useful site annotations. Ask for comment on how could be used in school/public libraries? [if relevant]</p> <p>Additional search: Participants directed to find a listing of NZ online newspapers [TPWD > Subject > News, Media... > News] Asked when this type of searching might be useful?</p>
Recap	3 min	<p>Review Google features using the 'Checklist'. Note that while useful to trial various options, best to focus on those most likely to use in their work environment.</p>	<p>Handout <i>Google Search Strategies Checklist</i></p>

2.2 SEARCHING THE WEB

SEARCH ENGINES / GOOGLE

Search 'engines' are a combination of software programs that index, search for and deliver Web pages. They create their indexes by employing Web 'spiders' or 'robots' ('bots') that regularly move around the Web from link to link sending the data back to base. Another program then scans key bits of the Web page and builds a vast index of key words and other elements of the page. A third program, a query processor, then steps in to search the indexes and deliver pages that match up to the words of your query entered in the search 'interface' (screen).

Given the huge size of the Web, the 'bots' take a while to crawl around it so they are set to check on sites and pages periodically, not every day. The time between these visits varies with search engine nature of the page so that a new page or website will not immediately register – if it turns up at all. If a page is not linked to by anything it may not be picked up and, as different search engines have different 'rules' for gathering and storing pages, no one engine will cover all the Web and all have some unique content. Although much debated, Google claims to have the most pages indexed (10 billion+?) but even then probably covers only a minority of Web pages. And, as we have seen in *s.1*, there are a whole host of resources accessed via the Internet that Google and other search engines don't reach (the so called 'invisible' Web).

What perhaps is more important than sheer size of their index is the way that search engines work out which pages are the most relevant and so are pushed to the top of the pile in any search. Most engines work in part by counting the frequency of your key words on a page and where they are located – if in a heading they may be ranked higher than further down the page. Google considers as well how many pages link to any page and the 'quality' of the pages that link to it. This quite complex calculation (all done in few microseconds) improves the chances of useful stuff getting onto the first page of results. Look at the type of material that comes to the top in any 'serious' search and you will see that generally well regarded or well used news or educational sites are frequently served up. Another reason for using Google is that it appears to give quite wide and up-to-date coverage of NZ sites, certainly better than some of the smaller, NZ based search engines. And making sure that you use *google.co.nz* (rather than *google.com*) will favour NZ sites. Whatever engine you prefer, none will cover all the Web so if one fails it may be worth trying another mainstream engine such as *Ask*, *Yahoo*, or emerging ones such as *Hakia* or *Quintura*, which have alternative approaches to searching or presentation of results. There are also engines that specialize in searching certain categories of material that may be used by special or academic libraries to complement the all-purpose ones ('vertical' or niche engines).

CREATING A SUCCESSFUL WEB SEARCH STRATEGY

There is no 'ideal' strategy - it is always depends on what you need to find and how you want use the information. Its important to give some thought to that before you start the search as will affect how you sort the good from the poor results. There are however a number of features of search engines that it's useful to get familiar with to help you achieve sound strategies and a higher success rate for your searches.

Taking a basic search string of a few words and modifying them step by step in Google can be an instructive way to learn a bit more about how it (or any search engine) behaves and the sort of strategies and features we can use to sharpen up our searches.

Making the Most of Google

Assume we want to do a preliminary (not the only) search on the Web to see what turns up on the topic (as described by a student) of: "the effects of violent video and computer games on young people" . For each search below note the number of hits (top right), the words in bold in the results (indicates those that Google has searched on) and the relevance of items that turn up on the first page.

1. Start with just a simple 3 word (**young violent games**) search: over 270,000 hits, some relevant stuff with *NY Times* article near the top. Those who are sharp eyed will note each of our terms turn up (are in bold) in each hit. Google in fact interprets our statement as 'find the words: young AND violent AND games' in any page. Some other search systems require us to actually insert the 'AND' ; Google does it for us, assuming you want to find pages that include all your words, not just one or the other.
2. Given that 'violent games' are commonly referred to in that exact word order, we can perhaps sharpen things a bit by searching on this concept (the thing or idea) as a phrase - put the words in double quotes: **young "violent games"**. This reduces the hits to about 87,000 but doesn't alter the top ranking items much given this is the way it is usually expressed in the most of the items. With a less common phrase or title it may though be a useful refining strategy. Note also that it is now in effect only searching for two terms – young AND violent games. The quotes tell Google to regard the phrase as one term only.
3. Thinking more on our search we realize there are various other words that could be used to talk about the concepts eg. 'Youth, adolescent or kids' may be used instead of 'young'. We could provide for at least one of these by using an 'OR' between the options: **young OR youth "violent games"** making sure the 'OR' is in uppercase (if not Google ignores it). Note that as you might expect this expands the number of hits a bit.
4. It still doesn't allow us to pick up on other possibilities though. We could keep adding other words with 'ORs' but this would quite a clunky way to search. Google thankfully has another neat little feature that lets us search for alternative expressions of a concept - the 'similar words' operator, the tilde (~). So if we took 'youth' as our core concept we could rerun a more expansive search as: **~youth "violent games"**. Look carefully at the bolded words and you will see 'young, kids, child' etc are being picked up on as well as 'youth'. Note though that 'teen' variants don't show, so it pays not to assume too much as to what might be covered. This little 'squiggle factor' can be useful though when you have concepts with many alternative expressions such as 'car'. And by the way if you're worried about also covering all the minor variations in words (eg. child/children/children's) then you don't have to - Google has a useful 'stemming' feature that automatically includes such small differences. Great for singular vs. plural variations. Try changing 'youth' to **'teenage'** in the above search and you will note 'teen' turning up also.
5. Now that we've seen what's available world wide, it would be good to limit the focus to NZ material only. There are two means to do this in Google – add 'new zealand' (or even just 'zealand' as this is unique enough) or click the 'radio button' **pages from New Zealand** under the search box. These are quite distinct options - try it and see with our search. Adding **'zealand'** means we depend on the word turning up in the web page or document. Turning on the radio button will limit the search to items with a **'.nz'** domain but of course excludes any NZ relevant information in overseas sites of with say a **.com** or **.org** address. Which option is the most useful will depend on the information needed. If you want be to be comprehensive use both.
6. At times in the middle of search we may get a little confused as to exactly what a topic or concept is about and a definition or two may help clarify our thinking. Again Google comes to our aid with its 'define' feature. Simply put *define* + a colon in front of the

word or phrase (eg. **define:violence** ...no spaces) and Google will provide one or more dictionary style definitions with links to the source. Try it.

7. Searches can also throw up the desire to check or unearth factual points eg. you may want to establish when video games were first introduced to get a better idea of when they started having an influence. While you may strike it lucky with a loose keyword search ('video games introduced') a better strategy can be to enter the hoped for answer minus the bit we don't know (all in quotes). Try the phrase: "**video games were first introduced**" and you will see the answer rapidly emerge. This strategy is sometimes called 'fill in the blanks' searching and is really useful for digging up those quick fact type queries. Google gives higher ranking to items that have the words in the same order you enter them so it pays to think how answers might be commonly expressed, whether or not in quotes.
8. To date our search scope has included material on the Web as a whole, as this is the default Google setting. There are however useful ways to limit or filter the initial search scope and/or the results. Look at the top of the search page and you will notice some other options: *Images*, *Maps* etc. Clicking on any of these 'tab' links generates a new search screen limited to the selected form - in effect 'slicing' out a section of the total Web resources to more closely match our needs. If we wanted to find some illustrations to go with our text material simply click on **Images**, enter our core terms (**young "violent games"**) if not already there, and a host of graphic images are displayed in small ('thumbnail') form with links to the original page and full size image. Similarly if we select the **News** and do a search it will limit the results to only current news media type sources – useful if we want mass media examples on an issue. Alternatively we could filter for a particular form or view to improve the precision of an existing broad search. Select the **Show options...** link at the top of the displayed results and consider the options in the left sidebar. These provide the facility to narrow display results to a particular time period (eg. **Past year** or use the **Timeline** view to select by decade), a specific form/format of result (**Reviews**, **Videos...**) or illustrations associated with results (**Images from the page**). Return to your original result at any time by selecting **Reset options**.

There are other features of Google worth exploring - click on the [About Google](#) link beneath the search box and have a browse. It is best however to decide which you are most likely to use regularly and get know these, rather than trying to use all the bells and whistles. This will ensure consistently successful searching and probably faster development of your search skills.

BROWSING THE WEB THROUGH DIRECTORIES

In a library some needs are best met by using the catalogue to search for quite specific titles, names etc. while others are better achieved by browsing the shelves in areas of interest. While we can't just cruise up and down rows of websites we can still browse at least a selection of sites through Web directories. These are compilations of selected sites in categories - much like telephone yellow pages. They can range from a few pages of recommended links posted on Library website to the large resources dedicated to specific types of sites such as education sources. We will look at two that are consistently useful to NZ searchers.

The **Te Puna Web Directory** (<http://webdirectory.natlib.govt.nz>) is put together by the National Library and is a selected listing of NZ and Pacific websites including online magazines and newspapers. Note the 'selected' - it doesn't list all NZ sites just those that experienced staff have decided meet the criteria for inclusion. This ensures a certain 'quality' but not exhaustive coverage, and sometimes not currency.

Most directories have a subject arrangement and the obvious way to search is to select the broad category we think the topic relates to, then 'drill down' through the narrower categories until we get to lists of sites to choose from. So, if we wish to find say sites of educational organisations relating to schools then we would click on the to *Subject* list then choose *Education* then *Educational Agencies and Organisations*, then *Schools*. From there we have descriptions and links to sites plus a referral link to school websites. We also have the option of browsing through 'alphabetical' or 'hierarchical' listings instead of 'subject'.

Note that we can easily move up and down the Subject levels to retrace our steps by clicking on the navigation ('breadcrumb') trail near the top left of the screen.

Google Directory (<http://www.google.com/dirhp>) has a similar structure but is a significantly larger resource that aims to cover the whole of knowledge. While it does serve up listings of a more uniformly higher quality than an open Web search, many of the categories are inconsistently covered, some quite dated and most heavily US weighted. One area where it is of value for schools is in searching for sites for younger kids – it can be a challenge to find age appropriate stuff from a straight Web search. Go to the *Kids & Teens* category (which has sites selected for under 18s) and you can then drill down to areas of interest, or search across just the *Kids & Teens* sites using the search box + radio button at the top. The *School Time* category is a good place to hunt for school project science and social studies stuff, and can be a quick way to getting a list of age suitable sites that students can browse themselves from the link page. Given the variable quality of the directory and the weak NZ coverage it should be complemented by other sources for NZ topics. Larger NZ public libraries such as Christchurch or Auckland, provide useful links of sites of more direct NZ interest.

Such human compiled directories may though provide some assurance of quality or at least a representative selection. And they can save you having to wade through pages of a Google search trying to select sites. You do however need to be familiar enough with a subject to select the right options as you drill down.

If you want information on very specific topics a Google Web search is a better strategy. But, if you need just a few good sites on a broad topic (such as iwi, pacific history, nutrition...), to familiarize yourself with a subject, or get a listing of related sites then directories can be a useful option.



GOOGLE SEARCH STRATEGIES CHECKLIST

Start with the most specific word(s) that describe the topic. Scan the first page of results and then build or refine your search by applying the following strategies:

Use More Words

Use other words associated with the topic to make your search more precise
eg. **erebus air disaster** is better than just **erebus**

Is it a Phrase?

If two or more words do they usually occur in a set order (names, titles, expressions)? Search for them as a phrase by enclosing in "double quotes" eg. **"hongi hika"**

A NZ Focus?

Select the **o pages from New Zealand** 'radio button' and/or add 'zealand' to your search. Which is preferable depends on what you need.

Provide for Alternatives

Need to provide for alternate ways of expressing something?
Use an **OR** (uppercase) between either/or alternatives: eg. **solar OR sun power**

If it's a concept that has various alternatives place a tilde (~) in front of the term to search for related words: eg. **solar ~power** (finds: **power or energy or electricity...**)

Need to Define Your Term?

Uncertain of what a term actually means? Try the 'define' feature eg. **define:energy**

Fill in the Blanks

Need a quick answer to a factual type query? Enter the hoped for answer (not the question) minus the missing bit: eg. **"largest wind farm in new zealand"**
A looser option is possible with just an asterisk: **largest wind farm in new zealand ***

Filter the Search Results to Improve Precision

Select **Show Options...** (top of displayed list) and limit results to:

- A publication time period (eg. **Past year**) or use the **Timeline** view to select by decade
- A type or format of result (eg. **Reviews** or **Videos**)
- Associated Images, illustrations (select **Images from the page**)

Return to your original result view by selecting **Reset options**

If you start interested only in a specific form or source then select from the 'tabs' top left of search screen to limit your searches to:

- Pictures, graphs, illustrations – select **Images** tab
- News media sources - select **News** tab
- Maps - select **Maps** tab
- Books – select **Books** tab
- Scholarly, research level sources - click on the **more** tab, then select **Scholar**.

Return to a full Web search at any time by selecting **Web**