

# Information Research FAQ v.4.7 (Part 5/6)

**Source:** <http://sci.tech-archive.net/Archive/sci.answers/2004-12/0051.html>

---

**From:** David Novak ([david\\_at\\_spireproject.com](mailto:david_at_spireproject.com))

**Date:** 12/29/04

Date: 29 Dec 2004 05:28:15 GMT

Archive-name: internet/info-research-faq/part5

Posting-Frequency: monthly

Last-modified: April 2002

URL: <http://spireproject.com>

Copyright: (c) 2001 David Novak

Maintainer: David Novak <[david@spireproject.com](mailto:david@spireproject.com)>

## Information Research FAQ (Part 5/6)

100 pages of search techniques, tactics and theory  
by David Novak of the Spire Project (SpireProject.com)

Welcome. This FAQ addresses information literacy; the skills, tools and theory of information research. Particular attention is paid to the role of the internet as both a reservoir and gateway to information resources.

The FAQ is written like a book, with a narrative and pictures. You have found your way to part five, so do backtrack to the beginning. If you are lost, this FAQ always resides as text at <http://spireproject.com/faq.txt> and <http://spireproject.co.uk/faq.txt> and with pictures at <http://spireproject.com/faq.htm>

\*\*\* The Spire Project also includes a 3 hour public seminar titled  
\*\*\* Exceptional Internet Research. This is a fast paced seminar  
\*\*\* supported with a great deal of webbing, reaching to skills and  
\*\*\* research concepts beyond the ground covered on our website and  
\*\*\* this FAQ. <http://spireproject.com/seminar.htm> has a synopsis.  
\*\*\* I am in Europe, seminaring in Ireland and Europe though I  
\*\*\* will be returning to the US shortly, and South Australia for  
\*\*\* a seminar this October.

Enjoy,

David Novak – [david@spireproject.com](mailto:david@spireproject.com)

The Spire Project : [SpireProject.com](http://SpireProject.com) and [SpireProject.co.uk](http://SpireProject.co.uk)

Search Tactics.

Section 7

The Pharaoh called on Shakh to negotiate the annual royal donation with the priests of Karnak temple complex. The Pharaoh was not wise in such matters and had previously given far too much to the detriment of the state. It was not wise to voice such sentiments. Shakh instead set about negotiating a figure ample to their needs but insufficient to further expand the temple complex.

Shakh wisely chose to negotiate up river at the Kom Ombo temple – away from Karnak. Choosing words carefully, he deftly rejected the initial estimate of the temple's needs, then spoke calmly, eyes tight, that the Pharaoh had decided Karnak should supply the priests to the Egyptian army – at current expenses.

It was a clever ruse. The negotiated royal donation was significantly reduced and the priests were happy to be excluded from military duty.

-----

If searching be a combination of science, art and experience, then the science of searching is the easiest of the three. There are just a few search elements to remember and search techniques to apply.

Firstly, there are the tactics associated with free text searching; that of Boolean, proximity, truncation, field searching, target searching and further enhancements.

Secondly, there are the basic classification schemes: the Dewey decimal system (for books) The WIPO and US Patent Classification Systems (for patents), the Standard Industrial Classification (SIC) Codes (for industry) and a number of additional classification systems founded on the same principles.

Thirdly, there is the way information is organized. A book has a table-of-contents and an index, large directories like Kompass and Gale Directory of Databases are arranged with so many indexes (geographic, subject, product, name) that the contact information is often separated and numbered, then referenced as a number. The results are initially confusing. Statistics similarly have ways of presenting information (pie charts, line charts, charts with ranges which do not reach zero) and again, this can be confusing the first time you see them.

Let's start with the technique associated with searching a text database.

Straight Word Searching:

All search situations allow you to ask for the presence of words in a block of text. Obviously it helps if you ask for the right word or words. If you ask for the right words, they you will quickly locate the information you desire. For best results you obviously want to choose a

word or words which accurately describes what you are looking for. Prepare to search the text several times with different terms, and consider the possibility of different spellings for the same words.

Straight word searching is fairly ubiquitous on the internet. You can always search a webpage with the search function of your web browser. Alternatively, you can search by placing a large amount of text into a word processor and using the in-built search functions. Your word-processor can handle large files like website traffic logbooks and archived files of past mailing list discussion. There are also specialist tools like the shareware WinGrep (<http://www.mindspring.com/~bgrigsby/wingrep.html>) for searching many files on your computer hard drive. (Alternatively, consider AgentRansack <http://www.agentransack.com>).

#### Text Fragments:

The simplest refinement to straight searching involves searching for parts of a word – if you are interested in surfing, search for surf better yet, search for " surf" with the space in front of the word.

#### Truncation:

Some search engines don't allow searches for text fragments, and you must explain your intention by adding a truncation mark (usually \* or ?) to the ends of words. For most professional researchable databases, alga? will include both algae and algal (as in algal bloom). I was once badly lost because of the spelling difference between aging and ageing. There are a number of improvements on this concept to. Sometimes there are special symbols for a non-space character car?a, sometimes there is automatic awareness of multiple spellings (colour & color). Sometimes there is even automatic awareness of synonyms. Often you are initially unaware important information is indexed under slightly different spelling, so truncation is strongly suggested for most searching.

#### Thesaurus:

An improvement on truncation is the opportunity to look directly at a list of words, either keywords, or descriptors. This allows you to see the range of spellings before you search. This is also ideal for searches of company names or proper places so you can select only the words you are interested in. In a simple way, some library catalogues present subject searches in this way: a list of subject categories arranged alphabetically.

#### Boolean operators:

Changing tack, searching for multiple words calls for "and, or, not" concepts. I want this word and that word, but not another word. It is simple enough. Many of the search engines allow for this with the -sign, and commercial databases often add brackets. Use of the not symbol is frowned upon in textbooks (too easy to dismiss information you are interested in it is said), but the 'and & or' is absolutely necessary for complex questions like I want [(spaghetti or noodle) and pasta] or (Italian and cuisine). With most internet search engines, but

not all commercial searches, you will find 'and' is assumed.

#### Proximity operators:

The next dramatic improvement fixes the position of words relative to one another. In this category we have adjacent (often written as adj, next, or "inserted in quotes"), near (by how many words), or in the same sentence. Often it is wise to stretch the distance a little (within two), but where available, proximity is best way to remove the drross without affecting the value of information. "Patent near Research" is much more precise than "Patent and Research".

#### Fields:

By separating information into different fields, we can selectively search different portions of the information. I want the title to show the words "Patent" and the abstract to include the words "Patent Research". Field searching is a common way to refine a search, but be aware searching titles is very likely to remove some desired information, where as searching descriptors and not abstracts may dramatically improve the content.

#### Date Fields:

Are you really interested in information more than 15 years old? Library catalogues frequently have many aging books, and date limiting is very wise.

#### Further Enhancements:

Ranking and the ability to search multiple databases are some of the further enhancements that select databases permit. There are also advances that do not have a grand impact – like natural language. Natural interpretation allows the searcher to phrase a question with common sentence structure. The computer then interprets what you want. In theory natural language is liberating but in practice the strengths of Boolean, proximity and field searching far exceed the benefits of natural language searching. Lastly, there are special techniques like target searching available on a few systems that bear discussing. Sorting allows you to shape the presentation of the information. When applied to financial information, this is particularly valuable. Alerts allow you to automatically repeat a previous search and have the information sent to you. Multiple database searching allows you to search a collection of databases concurrently. Ranking positions certain information at the top. These techniques can be valuable in certain circumstances.

These technical options improve the blunt system of simply asking for a word. You will find most search functions allow for some of these options and all commercial quality databases provide for numerous functions. The good news is an experienced searcher can accomplish wonders – collecting articles of 70%+ interest regularly on expensive database. The bad news is most of the best of search technology is not implemented on all the databases you will search and only occasionally on databases free on the internet.

### Classification

There are several search techniques associated with library catalogues. Beyond the simple author/title/subject search, we should also consider searching by Dewey number, and searching first for any title – then selecting the subject fields.

### Dewey Searching

The Dewey decimal system is similar in many ways to the patent classification system. Each step is divided into 10 – getting more and more specific. See this CAL State Dewey list (<http://www.calstatela.edu/library/guides/Dclass.htm>) to get an idea of its structure. This number here refers to a book called Australian government assistance to local government projects:

The Dewey system is arranged by Discipline, not subject groupings. Each digit to the right becomes progressively more detailed. The system works well in organizing books – and libraries expand it to suit their needs – but it is different from a subject catalogue. Because it is arranged by discipline, subject fields may be split.

In searching, we want to duplicate the walk to the shelves and browsing other publications that share similar numbers. We do this electronically by searching/browsing books that share most of a number. Drop a digit – expand the field of interest.

The Dewey system is a bit congested in certain areas, giving rise to very long numbers. For this and historical reasons, several national libraries do not use the Dewey system. The Library of Congress, for example, has its own classification scheme (Outlined here <http://lcweb.loc.gov/catdir/cpsolcco/lcco.html> ).

### Subject Searching

We can do better than searching the subject index of a library catalogue. Try instead to search for a book which interests you – which you can usually find easily with a simple title search – and then selecting the subjects that book are indexed under.

Many of the library catalogues are making this particularly easy by incorporating links into the catalogue results. A quick look at the Library of Congress, for example, will show how all the subject fields are linked to further searching.

We can show this in action by looking at the book Earth Time [1] by David Suzuki, at my State Library. As you can see down the bottom, it is indexed under Social Ecology [2] and Human Ecology [3].

This kind of 'locate then expand' is an effective search technique used in a number of situations. In commercial databases, we may search for a company then expand to make sure we catch any different company spellings. We may also wish to search for a book, then search for books by the same publisher.

[1]

<http://henrietta.liswa.wa.gov.au/search/asuzuki+dauid/1.2.46.B/frameset&asuzuki+dauid+t+1936&11..45>

[2]

<http://henrietta.liswa.wa.gov.au/search/dsocial+ecology/-5,-1.0.B/browse>

[3]

<http://henrietta.liswa.wa.gov.au/search/dhuman+ecology/-5,-1.0.B/browse>

-----

#### Patent Classification

All patents are given a special number. Unfortunately, each country has a distinct numbering scheme: US patents are assigned a consecutive patent number (currently 6 million+). Australian patents have an alphanumeric which includes the year. Canadian patents are numbered.

Above these numbering systems, we have the International Patent Classification (IPC), by the World Intellectual Property Organization (WIPO). Most every country uses the IPC to classify patents, save the US. US Patent Classification is similar in many ways.

#### International Patent Classification

Thanks to the World Intellectual Property Organization (WIPO), the International Patent Classification (IPC) works as a universal classification for patents. Started in 1975 and periodically updated, we currently use IPC 7th Edition.

Section, Class & Group. The International Patent Classification looks like this: A 02 J 1/00

At the heart of the IPC is the unique coding of every invention by its specific form or function. The system is highly specific and logical, and includes numerous cross-references to other codes of similar form or function. Think of this as the Dewey Decimal System for patents.

The first letter is the section – one of eight broad categories labeled A through G. 'A' represents Human Necessities. 'B' covers Transport.

Each section is divided into Classes. Each class includes two numbers. In addition, each class is divided into subclasses, the letters which follow the first number.

Each subclass is then divided into groups and subgroups. The number before the slash is the group, the number after the slash is the subgroup. Subgroups only have two digits, with further numbers considered as resting behind a decimal point: 3/46 then 3/464, then 3/47.

Thus A 47 J 27/09 includes the safety device on your rice cooker and B 63 G 11/00 covers your various aircraft carriers.

The IPC system is fully described in these published directories:  
The Official Catchword Index by World Intellectual Property

Organization.

International Patent Classification: Guide, Survey of Classes & Summary of Main Groups

International Patent Classification: Section G – Physics

International Patent Classification: Guide

Thanks to the World Intellectual Property Organization (WIPO), these full documents are online. We now have direct access to the International Patent Classification (7th Edition): Official Catchword Index, Guide to the IPC, and the complete Class and Section books.

Note: The International Patent Classification includes plenty of internal references – indicating this group is similar to another group; motorized boats take precedence over boat function. These internal references are important to effectively searching databases. There is more to the IPC, and we strongly recommend you read the Introductory Manual to the International Patent Classification (IPC) found on the WIPO website.

US Patent Classification

US Patents are classified with 400+ main classes and thousands of subclasses. Sound similar to the International Patent Classification? It is. US patents are numbered sequentially.

This means you can find US patents:

- by full text searching through the USPTO database CASSIS (found at US patent libraries),
- by bibliographic & abstract text searching online through the USPTO or IBM Patent Library,
- by US Patent number by US Patent Classification class & subclass – to list similar patents by an effective combination search
- by the searching recent notices in the Official Gazette... available online.

The USPTO allows you to search or browse the US Manual of Classification online. The Internet Patent Search System lets you to browse US Patent titles by class/subclass.

A little more information can be found with the Patent Guide to using CASSIS, at the University of Michigan.

Patent Search Strategies

Here are the avenues open to you:

- 1\_ Full text search and retrieval through a commercial database.
- 2\_ Free bibliographic & abstract searching online followed by selective patent perusal/ordering.
- 3\_ Paging manually through the relevant official gazette (the US gazette is searchable).
- 4\_ Retrieval of the titles & abstracts within appropriate class/subclass then selective review and patent perusal/ordering.

This last avenue is particularly resourceful and swift. Start by reaching for The Official Catchword Index, a book by World Intellectual Property Organization (WIPO). This will tell you the possible class/subclasses that will interest you. You could word-search a patent database and note all the class/subclasses found. Lastly, you can always reach for the three separate printed guides that lead you from section to subclass.

The result should be a collection of class/subclasses that may interest you.

With this information, you can now browse all the patents in the class/subclass. This process will help you locate all the patents that may interest you since patent classification is more reliable than free text search. (Note, both British and American spelling appears in patent databases.) This also allows you to quickly review the patents in other countries.

If you are undertaking a novelty search – is a patent sufficiently unique from other existing patents – then you must review more than one country. There can be a significant delay before patent applications reach other countries without affecting the protection. Case in point: Australia only accounts for 7% of the world's patents.

#### Further Search Strategy

Patent search strategy is further discussed in the Introductory Manual to the International Patent Classification (IPC) found on the WIPO website. You may also wish to reach "Searching for Patents" (<http://www.ummu.umich.edu/library/PTO/newpatsearch.html>) from the University of Michigan, and "Patents" by Simon Fraser University Libraries (<http://www.lib.sfu.ca/kiosk/nelles/patents.htm>).

-----

#### Trademarks

Trademark law is designed to protect consumers from confusion. The law can work to protect business investment in brands & slogans, but only if the business behaves in particular ways which protect consumers from confusion: actively using the trademark, working to restrict the trademark from becoming generic, routinely searching for unauthorized use. For a very clear description of trademark use, and the responsibilities of trademark owners, read the short webpages A Guide to Proper Trademark Use, and How are Marks Protected both by Gregory Guillot.

Trademark Law has implications for searching: Just because a potentially conflicting trademark has been found does not mean it should concern you. It may be simple to show or argue that trademark ownership has lapsed and become abandoned unintentionally.

A common law search involves searching records other than the federal register and pending application records. It may involve checking phone directories, yellow pages, industrial directories, state trademark registers, among others, in an effort to determine if a particular mark is used by others when they have not filed for a federal trademark registration.

The system may appear particularly legalistic, and it is. Recent Australian Trade Marks Office Decisions (<http://www.austlii.edu.au/au/cases/cth/ATMO/recent-cases.html>), information ultimately supplied by IP Australia, displays this vividly. However, much trademark activity is self-evident. In Australia, A\$350 and a minimum of seven and a half months will usually earn you a registered trademark. Should you choose a trademark and find another has used it, you will most likely receive a 'cease & desist' letter and forfeit the value you may have invested in the trademark.

This leads us to the importance of commercial trademark databases, watching services and other commercial services. Searching both prevents investment in an unusable trademark and inadvertent infringement by others – a responsibility of trademark owners.

#### Trademark Classification

A concise list of the 42 classes of the International Trademark Classification codes courtesy of Master-McNeil Inc. WIPO is in charge of the full class description, currently The 7th edition of the Nice Classification, but this is rather lengthy. IP Australia has a simple search feature of classification terminology.

Trademarks are assigned to a particular class of product or service. A slogan or mark, for example, could be registered for use in movies but not computer products. The situation has changes recently but let us explain the difference down the page a bit.

Originally, all goods and services were broken down into 42 classes. These classes are international divisions organized by WIPO (World Intellectual Property Organization), so are the same from country to country. Registered trademark documents will explain at length the types of products & services covered by a particular trademark.

There is some bleeding between categories, and trademark examiners are unlikely to grant requests for nearly identical trademarks in similar categories, but class plays a role in granting trademarks.

Recently it became necessary to list specifically the products or services to be covered, and the 42 classes have been expanded to a collection of specific sub-classes, which is reminiscent of patent classification, but far less useful.

Class is important as trademarks are class-specific. You can search by class in certain registered trademark databases, but this is not

particularly a good search technique: you are far too likely to miss a comparable trademark.

#### Trademark Picture Descriptors

Search Image Descriptors, by IP Australia, here abbreviated, needs basic words – simple like bird or butterfly.

One difficulty with trademark searches is that all the tools apply best to words which appear in trademarks. What of the picture? The solution appears to be image descriptors. I am uncertain of the international nature of image descriptors, but at least in Australia, there is a standard set of image descriptors. IP Australia allows you to search for other trademarks with a particular picture element – irrespective of the words involved. But to do this, you must first select the appropriate image descriptor.

#### Conclusion

Trademarks are just one element of intellectual property rights; patents, copyright, industrial design rights, circuit layout rights and plant breeders rights. As certain registered trademark databases are free online, some trademark research can be accomplished quite simply by the novice.

#### Why search?

- 1\_ To find existing trademarks similar to one you plan to register.
- 2\_ To find existing trademarks similar to one you plan to use as a trademark.
- 3\_ To see if a trademark is similar to a business name you consider using.
- 4\_ To search for possible infringing trademarks.

This is further explained in this help file by IP Australia.

#### Further Assistance

Misc.int–property has a lively usenet discussion on Intellectual Property. Access the newsgroup directly: misc.int–property or search the past discussion through Deja.com's usenet archive).

For a lively discussion of how trademark law affects internet domain names, consider the trademarks–I mailing list at Washburn University (read the Scout Report description <http://scout7.cs.wisc.edu/pages/00000138.html>).

-----

#### Industry Classification

Lastly, we have not yet researched the categorization of industries using standard SIC or NAICS codes. In simple terms though, all industries are given a specific code. Sub–industry is given a more specific code. More and more specific codes refer to the production of more and more specific items. Of course, some companies will be

involved in a collection of industries.

Two competing standards, the SIC and NAICS, have different codes but the same coding system. Each code system can be mapped on the other, so will cause you no undue concern. Trade statistics, digital business directories, and national statistical bureau industry data will all use the industry codes.

### Information Quality.

#### Section 8

Information has value. It also has other qualities that will assist you to judge information you may consider buying.

**Accuracy:** the factual nature of the information presented. If the statistics purport to show a particular trend – how large is the margin of error? How large is the sample size? How likely are there to have been factual errors in their development? The measurement of statistical error is now a refined science in some fields. A statistical result can be inaccurate when the sample size is too small, if the margin of error is too large, the sample collection procedure incorrect, or a number of other situations.

**Reliability:** the support for trusting the solutions, both from additional resources and from being able to duplicate the conclusions. This includes the reputation of the researchers. No matter how inaccurate and biased you may believe certain facts to be, successful independent support of a suggested fact does improve its value.

**Bias:** conscious or subconscious influences that affect information. Bias can occur in collection, preparation and presentation of information. Most information you find will be tainted. Secondary information is deeply affected. Statistics are not necessarily less biased.

We counter bias in several ways. Firstly, we try to be aware of bias. Where is bias likely? Which direction would the bias affect the information? Secondly, we try to collect information with different bias. This is why research based solely on government research, no matter how accurate and reliable, is less valuable. Often information from different countries can counter bias. Thirdly, we need to accept bias is likely to exist. This is why primary sources are often more valuable than secondary sources. This is why tertiary sources, like experts, can rarely stand alone.

**Age:** The date information was created or compiled will feature prominently in the value of information. Dates given sometimes mean the date information was created, or the date information was compiled. How old is a book compiled in 1995, which took the author 10 years to finish? I find statistics often forecast information, prominently displaying recent compilation dates but still use old census data or

the like to draw their conclusions. Information on the internet typically has no date, and can be severely challenged because of this.

Purpose: purpose merits further discussion. When you are uncertain about potential bias, you can look for reasons to distrust the information instead. Suspicion is not equivalent to bias, but it can be thought provoking. Privately, I have heard repeated rumours important national statistics have been fudged in different countries. A government research report investigating the price of books in Australia would have a political purpose, a purpose that provides the climate for some potentially significant bias. A tell-all book by industry experts often includes a tremendous quality of insider experience difficult to find elsewhere. While there may be a purpose of self-aggrandizement, the purpose is less a climate for significant bias. Medical research has perhaps the greatest climate for significant bias, and this suggests the greatest standard of proof and external, reliable support.

Accuracy, reliability, bias, age and purpose are very important in research. This is what leads us to an appraisal of value. For years, the tobacco industry funded 'independent' research finding smoking minimally harmful to health. It is now likely there may have been errors brought on by accuracy, and bias. Certainly, purpose was in doubt. As new studies show smoking is harmful, we can also say the original research lacked reliability. In some topics, like the internet, research is perpetually suspect because it also ages so quickly.

I have seen further discussions that add 'Coverage' and 'Authority' to this checklist. Both have bearing on the value of the information contained. By coverage, we mean how much detail is invested in covering a specific topic. Sparse or shallow coverage is closely tied to missing critical aspects of information. News stories frequently have limited coverage.

Once you are acclimatized to these elements, you begin to see potential for error in a whole range of information. Real-estate association figures, expert opinions, Toothpaste advertisements and National GDP figures all occasionally display some degree of warping and manipulation, clouding the truth. The solution is awareness, comparison and careful analysis. As a personal aside, this is part of the reason for my personal dislike for market research: it is often taken far more seriously than warranted and mean far less than suggested.

---

This document continues as Part 6/6

---

Copyright (c) 1998–2001 by David Novak, all rights reserved. This FAQ may be posted to any USENET newsgroup, on-line service, website, or BBS as long as it is posted unaltered in its entirety including this copyright statement. This FAQ may not be included in commercial collections or compilations without express permission from the author.

Please send permission requests to [david@spireproject.com](mailto:david@spireproject.com)