Information Retrieval & Management

A guide for students

KTH, Royal Institute of Technology

2014 (2012)

CONTENTS

Information Retrieval & Management

This self-study material is provided by the Information Retrieval Unit at the KTH Library as an introduction to the principles of information retrieval and information management. Here, we describe the various kinds of information sources, how to run a search for information, and how you can use the results of a search in a professional manner; basically, we wish to improve your Information Literacy. In short this can be seen as your ability to

- Identify your information needs
- Search for information
- Critically evaluate information
- Creatively use information

In order to gain practical experience of searching in a variety of sources, we suggest that you take one of the KTH Library's class-room courses at the ECE-school. Please note that this material is primarily intended for students at technical universities. The examples are given with them in mind, but the principles are valid for everybody.

When you have finished reading this material and have correctly answered the questions, you should understand:

- The differences between various types of information
- The need to evaluate the sources of information, and that you should primarily cite only primary information
- When to use which information resource (the Internet, reference databases, handbooks etc.)
- How to create a search question, using synonyms, thesauri, Boolean operators, proximity operators and truncation to optimize your results and to avoid linguistic pitfalls
- The difference between free text searches, searching with free search terms and using controlled search terms , as well as the pros and cons of using field searching
- How to find full text documents available online and in the library, and how to order what you can't find yourself
- How, when and why to cite in order to avoid plagiarism, and how reference management software can help you with this.

1 Introduction

Information Literacy has been defined by different organisations in different ways. In the Prague Declaration of 2003, information literacy is defined as: "…knowledge of one's information concerns and needs, and the ability to identify, locate, evaluate, organize and effectively create, use and communicate information to address issues or problems at hand"; furthermore, "it is a prerequisite for participating effectively in the Information Society, and is part of the basic human right of lifelong learning."

1.1 Objectives

In this section you will

- Learn about different types of information
- Learn about the origins of reference databases and the difference between searching in a reference database and on the Internet
- Obtain an overview of the search process

1.2 Why is it important to learn about Information Retrieval & Management?

Throughout your career, you will need to search for information. As a student, you may need to find background material for an essay or a project report; in research, you may want to identify new developments in your area of research. If you work in industry, you may be asked to find which products your competitors have on the market.

The aim of this introductory text is to show how you can find answers to such questions—in an effective way—by formulating your information needs, identifying relevant information sources, and analysing the results of the search. The focus lies on searching in reference databases, since that is where you most easily find tracks from scientific information, but you will also learn to find what you need from the KTH Library.

1.3 Different types of information

1.3.1 Scholarly

Scholarly or academic information is typically written by academics for academics within the same research area. It uses the technical terms for that field of research. Within the areas of science, technology and medicine this type of information is typically found in scientific journals.

The scientific journal was invented in the latter half of the 1600s. Initially, there were only a few such journals, but by about 1850 the number of journals had risen to around 300, and scientists needed a tool by which to scan scientific articles without actually reading them in full. The solution was to print serial publications that contained short descriptions of the contents and articles in the ordinary science journals. For most such serial publications, the journals to cover were chosen according to subject. These new serial publications were called abstract journals or index journals; they are typical secondary information publications. (See "Little science, big science" by Derek de Solla Price, publ. by Columbia University Press (1968)).

A century or so later the first computerised versions of the abstract journals appeared. They came to be called reference databases, bibliographic databases or even bibliographic reference databases. They became commercially available for the first time in 1970

1.3.2 Popular

Popular scientific texts are aimed at people who are not experts in the area. Journals such as *Scientific American*, *New Scientist* or *Illustrerad Vetenskap* supply such articles: they provide a good picture of a subject using non-scientific terminology and without the details of a scholarly text.

1.3.3 Patents

A person who has made a new invention may apply to the national patent office for a patent, which is a legal document granting the inventor the right to exclude others from making, selling or importing the claimed invention for a certain period of time. In order for the application to be accepted, the invention must fulfil the criteria of novelty (being different from what is currently known), inventiveness (not an obvious development of an existing product or technology) and usefulness (industrially applicable); in exchange, the patent owner must disclose how to make and use the invention so that others can use the information for further development.

It is important that information on the invention has not been published before the patent application is received by the patent office. Thus, patent documents can be the first (and sometimes only) publication on newly-developing technologies. They are widely used by commercial organisations to identify the areas their competitors are working in, or to guide their own research, and are very useful as background material for a study of emerging technologies.

1.3.4 Commercial

Commercial information, such as a trade journal or an advertisement, may come from a company informing its customers about new products or updates, or from someone trying to get you to buy a product. Commercial information is more likely to be biased than scholarly information and you should always be aware of the purpose behind it. It can, however, be a good source of information; if you want to know the price of a product from different companies, then product catalogues are examples of useful commercial information.

1.3.5 Primary and secondary information sources

You should learn to differentiate between primary and secondary information sources. A primary information source is the original source for a specific piece of information. A research article in a scientific journal discussing new experimental results is an example of a primary information source.

Newspaper articles reporting directly from wars or natural disasters are also often primary information sources, but they are not scientific documents.

A secondary information source uses primary information but is not the original source: it could be a review article discussing already published datasets, or someone reciting a story they've heard from someone else. Wikipedia is a well-known secondary information source, as are trade journals or the science sections of daily newspapers. Sometimes, important information present in the primary source may be excluded or misinterpreted in the secondary source. When you want to use information you have heard or seen somewhere, you should always make sure that you refer to the primary information. You should, of course, have read the primary information when you cite it; you should never cite anything you have not read yourself.

1.3.6 Peer Review

Many scientific journals apply a Peer Review process. Peer review means “examination by equals”, and is intended to be a form of quality control. In the world of academic publishing, peer review means the examination of a manuscript intended for publication in a journal to ensure that it meets the standards of quality and ethical behaviour set by the publishers. In the Sciences (peer review also occurs in the Arts and Humanities), the reviewer is another person active in the same area of research. The task of the reviewer is to ensure that the author has used scientific method in the research, that the conclusions drawn from the results are reasonable, and that the author has correctly cited relevant references

.

1.4 The search process

The search for scientific information is often a multistep, iterative process. In the following sections we will go through each of these steps:

- Define the search request as precisely as possible
- Choose appropriate information resources (reference database(s), library catalogue, Internet)
- Choose and list relevant search terms
- Modify the search terms to suit the chosen information resources
- Combine the search terms to create a search query
- Run the search
- Modify your search query based on the previous results and re-run the search. Good searches include iterative behaviour
- Save your selected search results in a reference management system

Searching is often an iterative process (Figure 1): any given source may not contain all the information you require, and you may find that the search terms you have initially chosen are not the best, simply because you do not know the terminology. By running a modified search using terms extracted from the previous search, you can often greatly improve the relevancy of the results you obtain. Even if you do not use the best search terms from the start, this way you can make up for your limited knowledge of the terminology. It is good search behaviour to modify and re-run the search until nothing new interesting appears anymore, and you reach “Good enough” according to Figure 1.

For a comprehensive search, you must be prepared to try two or more different information resources, using the iterative process in each one. You may need to modify your query as you move between search resources, to take into account different operators or fields available for searching.

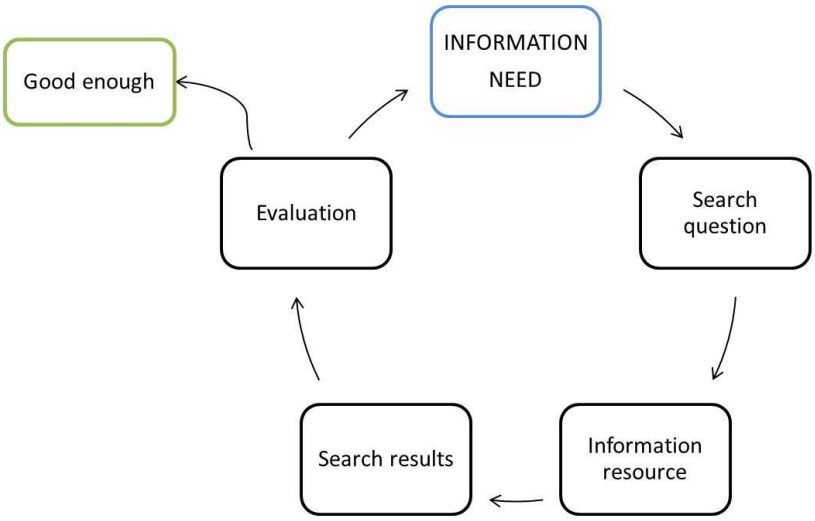


Figure 1. The search cycle.

If you are searching for background information for a project, start searching for information as soon as it becomes clear to you what the project is all about. Finding relevant literature can take time. To acquire the full text of documents that are not immediately available at the library may also take time, sometimes weeks, and it is not a good idea to rely entirely on information which is immediately available. As your work progresses, you will find that your information needs will change and you will need to make new searches.

Remember to document your search step by step in case you need to repeat or modify the search; many databases automatically record each search in the current search session and also allow you to save the searches for reuse later on. See Save your search history, below.

1.5 The KTH Library web site

The KTH Library contains a large collection of periodicals (journals, conference proceedings or book series) or monographs (books, dissertations or reports), in printed and electronic form. The library also gives students and faculty access to a large number of electronic databases which can be accessed on campus or off campus. The main library building is usually open from 8 am to 8 pm Monday-Thursday and 8 am to 6 pm on Fridays; the electronic library is open all hours. The KTH Library uses the discovery system Primo, a search tool that provides access to the KTH Library's online and print collections.

1.5.1 Register for a library card

The KTH Library card and PIN code allows you to:

- Request books from within the library
- Extend your loans and check the details of your current loans
- Order articles that are not available online or at the KTH Library (see 5.4 Document ordering)
- Order books from other libraries (inter library loans)
- Reserve a group study room.

When you are using a computer on campus you have full access to all the electronic resources (e-journals, reference databases etc.) available through the KTH Library. To get the same access off campus, you need to log in to the KTH Library service using your kth.se account.

On the KTH Library web site you can also find a link to Search Support where you can book a 45 minute consultation with an information specialist. The information specialists at the KTH Library all have

higher degrees in subjects taught at KTH and are experienced in finding high quality academic information in their subject areas.

1.5.2 The kth.se account

All current students and staff members of KTH possess personal kth.se accounts. As long as you are an active KTH student or member of staff you have access to the KTH Library electronic resources. This includes most of the reference databases, e-journal archives, collections of conference papers—as well as of conference proceedings—and e-book collections. If you are logged in to your kth.se account you have 24/7 access from anywhere on the Internet. Using any campus computer always grants you access to these resources.

2 How to prepare your search

2.1 Objectives

In this section you will

- Learn to prepare a search by analysing your information needs and choosing a suitable information resource
- Understand the general idea of how reference databases are constructed
- Learn to use the thesaurus in a reference database

2.2 Identify the main concepts

What problem are you trying to solve? What question are you trying to answer? These are the types of questions you will have to answer before you can start a search. To be able to run a successful search, you need to know how and where to start. Keep in mind that it is easier to solve one problem at a time and that the information you need to find at the start of a project will be different from the information you need as the project develops. It is also possible that what you are looking for at the start of a search is not what you later accept as the search answer, since your information needs may have developed further as your search progresses and you obtain new results.

Define the search request as precisely as possible. For example, consider the following question: “What are the effects of global warming on coasts?”

Start by breaking this phrase up into a number of concepts, in this case “effects”, “global warming” and “coasts”. Think carefully about which search terms could be used to define each concept, together with other words or phrases (search terms) with a similar meaning. Then group those terms together, as in Table 1, where terms belonging to the same concept are placed in the same column.

Concept 1	*Concept 2*	*Concept 3*
Effect	Global warming	Coast
Influence	Greenhouse effect	Coastline
Results	Climate change	Shore
Impact	Changing of climate	Seashore
Consequence		Beach

Table 1. Examples of search terms for three different concepts

2.2.1 Stop words

Stop words are commonly-occurring words that are ignored by the search engine and therefore do not need to be included in the search; examples are *about*, *again*, *all*, *almost*, *also*, *of* and *the* as well as operators such as *and, not,* or *or*. If you do need to search for a phrase that contains stop words, place the phrase within braces or quotation marks (depending on the database interface, check HELP to see what applies). For instance, enter *{block and tackle}* or *"block and tackle".*

2.2.2 Sources of search terms

If you find it difficult to choose good search terms, you can consult the reference literature available at the KTH Library. This includes bibliographies, encyclopaedias, reference works and dictionaries. A great deal of reference literature is accessible online. You can also talk to your teacher or other experts on the subject. If you already have access to a publication on the particular subject, use that to assist your search for search terms. In this early stage of the search process, Internet resources such as Google or Wikipedia can also be a great help for checking terminology.

Take some time to think about the kind of information you need: a summary of the latest developments or detailed information; the very latest or older information; in English only, etc. The better you prepare yourself, the better the search will be and the more relevant the answers.

2.3 How to select an information resource

Depending on what kind of information you need, different sources of information are appropriate (Table 2). If you are looking for the melting point of tin, a handbook is probably the best source of information. If you want to find background information on a new subject area, you will probably need to look for a book or a review article. To find cutting-edge new scientific results, you will need the latest scientific paper on the subject. And if you want to know what a company is planning for the future, you might want to look up their patents.

If you are searching for a	*Look in*
Book	KTH Library catalogue, Libris, Primo
Doctoral thesis	DiVA, SwePub, Dissertations and Theses
Journal article	Reference databases, Primo
Master thesis, Candidate thesis	KTH Library catalogue, Libris, Primo, DiVA (from KTH), SwePub (all Swedish universities), uppsatser.se (in Swedish), essays.se (in English)
Report	KTH Library catalogue, Libris, Primo, SwePub
Patent	Reference databases, DiVA (only from KTH)

Table 2. How to select an information resource

2.4 Understanding reference databases

To perform a successful search in a reference database it helps to know something about how such databases are constructed.

A reference database is a collection of document descriptions (bibliographic references) that has been processed, organised and stored to allow for searches specifically on – for instance –the contents and the authors of the described documents. A genuine reference database seldom contains the actual

documents, but nowadays often links to them. Reference databases often contain data on a specific subject, for example chemistry, physics or geology. The database is built by a database producer, who gathers and analyses selected documents and then puts bibliographic information in the database. Bibliographic information means details on author or authors, title, source details (for instance, the name of a journal or a conference), date of publication and so on. In each reference this information is sorted into various individually-searchable parts called fields. If the original document includes an abstract (a summary of the content) this is often also included, or the database producer may write one. In addition, some database producers index the documents with terms chosen from a thesaurus. The databases are continually updated with new documents, usually once a week.

The most common reference databases in engineering and science are bibliographic databases, which refer to documents from periodicals, conferences, reports, dissertations, books, patents and standards. There are also databases that contain the complete texts of articles—full text databases—or compilations of facts, such as numerical data, chemical structures or physical facts. These databases are often called databanks (or source databases), because they do not – as opposed to reference databases – force the user to go elsewhere to find the actual information that has been found. One way to look at reference databases and databanks is shown in Figure 2.

The reference databases are usually stored at a database host. The host acts as an intermediary between the producer and the searcher. Often a database host can store several hundred databases. The search interface is the same for all reference databases on the same host.

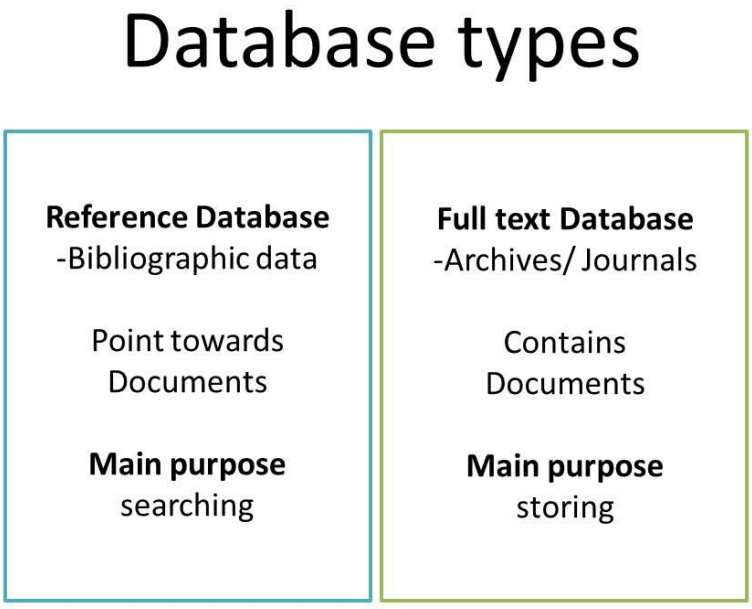


Figure 2. Types of databases

2.4.1 Coverage

Different reference database suppliers select the documents to include in the database in different ways: they may have a basic list of journals from which they take in all articles as well as other items (such as letters to the editor), and a second list of journals from which they only select articles which cover areas of particular interest to the database supplier. For instance, the supplier of SciFinder (a chemistry database) takes in all articles (cover-to-cover) from a core list of 1,500 chemistry journals,

but only selected articles from another 8,000-9,000 journals. Many computerised reference databases came into operation in the 1970s and contain references from that time onwards. However, many reference database suppliers have extended their coverage back in time, and it is now possible to locate documents that were written in the early 1800s. Note that some databases do not allow you to do a date-specific search all the way back: for instance, Scopus includes documents published in 1823, but you can only specify a search for publication years which are 1960 or later.

2.4.2 How to choose a database

The KTH Library provides access to a number of databases. From the main menu at the KTH Library web site, you can find a list of accessible databases by clicking on “Databases” under the “Search Tools” tab. Then, for example, select a “Category” (followed by a “Sub-category”) to find a group of subject-specific databases. Or search in “Keywords” to find a database using a search term, for instance “pharmaceuticals”.

When you have identified a suitable database for your search you click on the name, which will link you to the database native search interface. Please note that the KTH library interface forces you to open (by right-clicking) the database link in a new tab for an optimal search. If you have problems identifying a suitable database or running a search, contact an information specialist at the KTH Library. They are usually available for a quick consultation at the information desk in the main library. You can also book a 45 minute personal consultation through the Search support. The Search support consultation is booked via the tab “Services”.

2.4.3 Reference database profiles

Although most reference databases cover a specific subject area, for purely practical reasons, this subject area cannot be too strict. Areas of scientific learning do not have strict limits but flow into and complement one another. It is therefore useful to test reference databases that lie in the "intermediate zone" with respect to the subject. This zone may contain significant amounts of the sort of the information you are looking for.

For our search for the effects of global warming on coasts we chose Inspec, since this is one of the most popular and comprehensive reference databases amongst engineers. Inspec contains records mostly from journals and conference papers, and covers the whole period back to 1898.

Other reference databases that might be even better suited for this search include: ArtikelSök, Dissertations and Theses, Environmental Sciences and Pollution Management, NASA Technical Reports, NTIS, Scopus and Web of Science.

2.5 Where to search in a database

The most common way to search for information in a reference database is to use free terms, words based on your own knowledge of the subject and your imagination, to search in pre-defined fields. That is, you determine the search terms yourself and enter them in the single search box. This search box corresponds to one or more fields determined by the database producer.

2.5.1 Free text searching

If all the material in the database is searched when you enter your search terms, it is a free text search. What is covered by a free text search depend on the database. In a full text database the whole text is

searched, whereas a free text search in a reference database is limited to the information available in that database. For some databases the default search option is free text searching.

2.5.2 Field-based searching

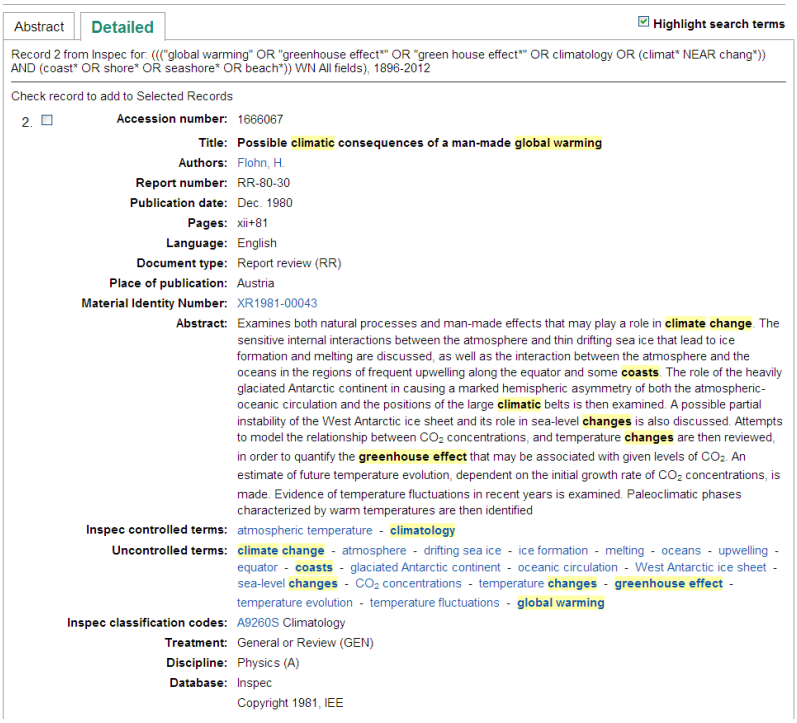
Each record in a database is divided into several fields such as author, publication year etc. It is possible to limit a search to just one or more particular fields. Although this is not recommended for a general search, it can be useful when you are searching, for instance, for a particular author or a specific publication, or if you want to limit your search to a certain year. Searching for indexed words (from the thesaurus) in the indexed word field is also an example of field searching.

The number of fields varies between databases but some of the commonly occurring ones are:

- Abstract
- Article title
- Author
- Classification codes
- Controlled terms (index terms, keywords)
- Journal name
- Publication year
- Publisher

The reference database producers and hosts are constantly improving the organization and ease of use of their databases. This includes increasing the number of searchable fields. While this is an improvement, it is also a reason for caution. As the number of searchable fields increases, new publications are immediately sorted into the latest organization. Older publications, however, may not have been indexed in all fields, since at the time they were indexed there were fewer fields to index in.

Figure 3 shows a record from the bibliographic database Inspec, with examples of searchable fields.


Abstract | Detailed — Highlight search terms

Record 2 from Inspec for: (((“global warming” OR “greenhouse effect*” OR “green house effect*” OR climatology OR (climat* NEAR chang*)) AND (coast* OR shore* OR seashore* OR beach*)) WN All fields), 1896-2012

Check record to add to Selected Records

2.

Accession number: 1666067
Title: Possible climatic consequences of a man-made global warming
Authors: Flohn, H.
Report number: RR-80-30
Publication date: Dec. 1980
Pages: xii+81
Language: English
Document type: Report review (RR)
Place of publication: Austria
Material Identity Number: XR1981-00043
Abstract: Examines both natural processes and man-made effects that may play a role in climate change. The sensitive internal interactions between the atmosphere and thin drifting sea ice that lead to ice formation and melting are discussed, as well as the interaction between the atmosphere and the oceans in the regions of frequent upwelling along the equator and some coasts. The role of the heavily glaciated Antarctic continent in causing a marked hemispheric asymmetry of both the atmospheric-oceanic circulation and the positions of the large climatic belts is then examined. A possible partial instability of the West Antarctic ice sheet and its role in sea-level changes is also discussed. Attempts to model the relationship between CO_2 concentrations, and temperature changes are then reviewed, in order to quantify the greenhouse effect that may be associated with given levels of CO_2. An estimate of future temperature evolution, dependent on the initial growth rate of CO_2 concentrations, is made. Evidence of temperature fluctuations in recent years is examined. Paleoclimatic phases characterized by warm temperatures are then identified
Inspec controlled terms: atmospheric temperature - climatology
Uncontrolled terms: climate change - atmosphere - drifting sea ice - ice formation - melting - oceans - upwelling - equator - coasts - glaciated Antarctic continent - oceanic circulation - West Antarctic ice sheet - sea-level changes - CO_2 concentrations - temperature changes - greenhouse effect - temperature evolution - temperature fluctuations - global warming
Inspec classification codes: A9260S Climatology
Treatment: General or Review (GEN)
Discipline: Physics (A)
Database: Inspec
Copyright 1981, IEE

Figure 3. Individual record from a search in Inspec showing field names in the left-hand column

In some cases a reference database may be available through different database hosts: the content will often be the same but the search possibilities may differ. For example, the Inspec database contains the same articles on both the STN (Scientific and Technical Information Network) host and the Engineering Village host; however, it is possible to search in 50 fields in Inspec via the STN host, but in only 25 fields via the Engineering Village host.

2.5.3 Thesaurus

Some databases offer the possibility to search with controlled vocabulary from a thesaurus. A thesaurus is a specific list of terms developed by the reference database producer and used to index ("tag") each reference. Note that different reference databases have different thesauri. In addition to controlled and free text, words and phrases you may also encounter the concepts controlled terms, or descriptors, and free terms. The meaning of 'term' here would be words, phrases, numbers, formulas, abbreviations etc. In controlled text searching, you search for the terms in a specific field.

Definition: Controlled vocabulary is an established list of standardized terminology for use in indexing and retrieval of information; it is used to harmonize the several definitions which may exist for a given term. One example, found in the CAS REGISTRY chemistry database, is benzoic acid, 2-(acetyloxy)-, which is the standardised name given to the substance variously known as acetylsalicylic acid, 2-(acetyloxy)benzoic acid, 2-carboxyphenyl acetate, Aspirin or Magnecyl.

Words or phrases contained in a thesaurus are indexed alphabetically and have a precise relationship with each other. This relationship is hierarchical, progressing from broader to narrower concepts. Figure 4 shows the entry for the word *climatology* in the Inspec thesaurus.

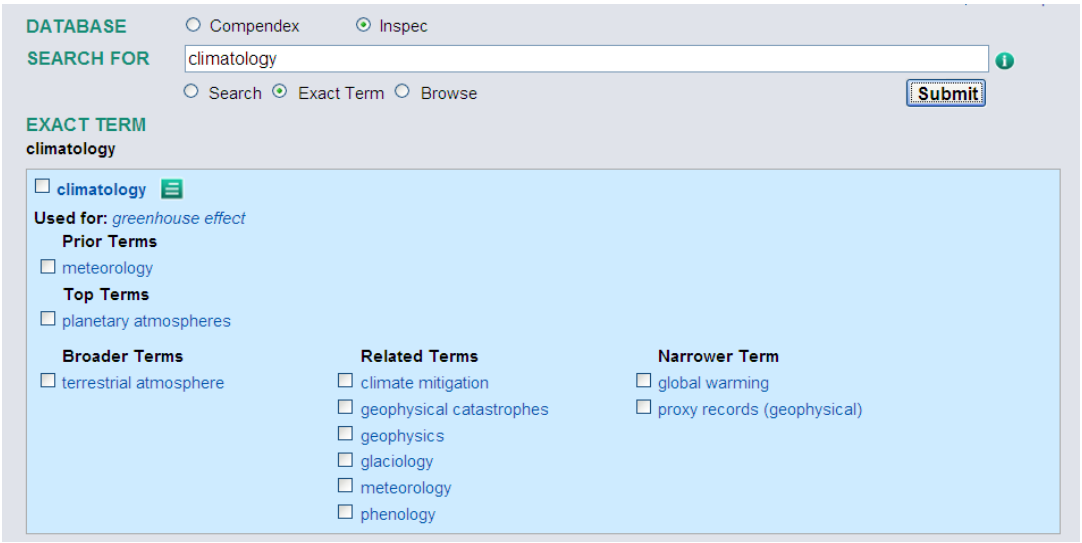


Figure 4. Thesaurus entry for "climatology" in Inspec

Other headings which appear in the thesaurus listing are "Used for" and Prior Terms". In the Inspec database, *climatology* is the preferred index term to use when you wish to search for *greenhouse effect*, and *climatology* has replaced the prior (older) term, *meteorology*. By clicking on the square information button you can see that the term *climatology* was introduced as the preferred index term in 1973. Thus, if you wish to obtain documents published before 1973, you should include both *climatology* and *meteorology* in your search.

In this example, the term *planetary atmospheres* is a Top Term in the hierarchy; that is, there are no broader terms than *planetary atmospheres*.

A thesaurus can also contain Related Terms, which indicates closely-related areas where you might find relevant information.

Different database producers update their thesauri more or less frequently, to account for changes in terminology and to remove redundant or add newly-appearing terms; for example the producers of the MEDLINE and Inspec databases update their thesauri once a year.

Since the thesaurus is unique for a database you may need to rephrase your search terms when you switch database.

In addition to the controlled terms in the thesaurus, some database producers add a field with uncontrolled terms. The terms in the uncontrolled terms field are not in any way organized but are chosen from the full text document and the author's own keywords. The terms found in the uncontrolled terms field are often useful in a free term search. It is also possible to search the uncontrolled terms field.

2.5.4 Pros and cons with different search types

There are advantages and disadvantages with both free term searching and controlled term searching. It is usually a good strategy to start searching with free terms to see what sort of result you obtain and to get ideas for other search terms. You can also use words or phrases from the thesaurus as free search terms to obtain more relevant search results.

With free text searching you risk obtaining fewer relevant references, and getting more mismatches and irrelevant results. On the other hand, you can find more references on the fringes of your topic that may be interesting, and you can search in several reference databases with the same search terms.

With a controlled term search you retrieve more relevant references, and the risk of mismatches and irrelevant results is small. There is, however, a risk that the search is too precise, and you may miss results that are not highly relevant for your subject but which may anyway be of interest. There is also a risk that the thesaurus is not fully up to date with the development in all research areas, such as when a large research field differentiates into several smaller. The terms found in the thesaurus can thus sometimes also be too broad for a precise search.

The best results can often be obtained by combining controlled term searches and free term searches.

3 Phrasing the question

3.1 Objectives

In this section you will

- Learn to use Boolean operators and proximity operators
- Learn the importance of considering synonyms and linguistic aspects when phrasing the search question
- Understand how to truncate

3.2 Boolean operators

The absolute majority of reference databases and Internet search services use the Boolean operators AND, OR and NOT. AND is used between different search concepts, OR is used between synonyms within a concept. NOT excludes a term or concept from the search.

Here, the three answer-sets retrieved using search terms A and B serve as example (Figure 1):

- If A and B are two synonyms and you want your answer set to include everything that contains *at least one of the terms* , use OR between the two terms.
- If A and B are two different concepts and you want only the answers which contain *both* of these concepts, use AND.
- If you are *only* interested in A and not at all in B, you could use NOT, that is A NOT B:

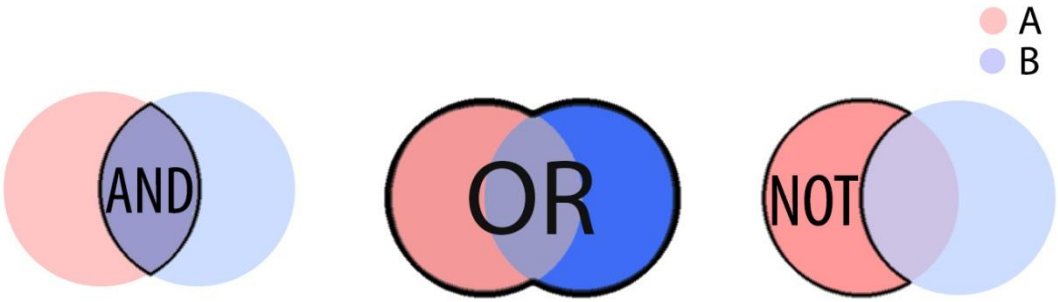


Figure 5. Boolean operators

We recommend you to always think twice before using NOT. As you can see from the example above NOT excludes some of the answers containing A, because they also contain B. If you are aiming to cover everything concerning A, using NOT would ruin your search.

3.2.1 Proximity operators

Proximity operators make the search more exact than the AND operator. AND requires that the search terms appear somewhere in the reference, but does not require the search terms to be connected to each other in any semantic way. Proximity operators, on the other hand, can be used to specify how close the words are to each other and their relative order in the text.

Databases have different user interfaces and do not follow a common standard. Therefore, every time you begin a search using an interface that you are not familiar with, we advise you to read the instructions which can be found under the HELP button.

Figure 6 shows a few examples of proximity operators and their uses.

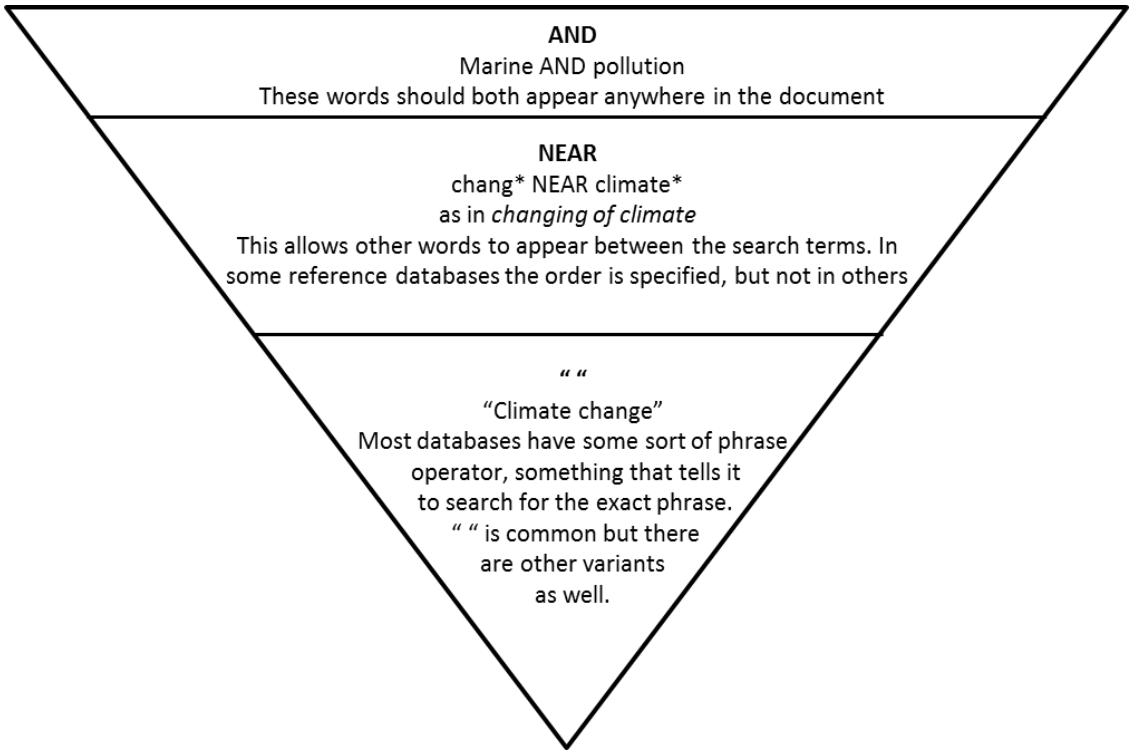


Figure 6. Narrowing the search using proximity operators

3.3 Linguistic aspects

When you choose your search terms, always think about the following linguistic aspects:

3.3.1 Compound wording (when words are joined together)

A compound word can evolve over time from two words written separately (Table 3), then joined by a hyphen, and finally written as a single word; all variants should be considered in a search.

ground water	ground-water	Groundwater
sea shore	sea-shore	Seashore
on line	on-line	Online
wind power	wind-power	Windpower

Table 3. Illustration of development of compound words

3.3.2 Synonyms or terms related within the search context

Synonyms are words that describe the same thing. When we talk about synonyms in a search context the word is used with a broader definition. Synonyms in a search context are words or phrases that can be used to describe a part of the search question; they are not necessarily synonyms in the strict sense.

For example: *durability - endurance - stamina* can be considered synonyms in the traditional sense of the word whereas *ship - boat - naval* are not truly synonyms but may be considered synonyms in a search context.

Sometimes, opposite terms can act as synonyms when it comes to formulating your search question:

Keep tight - prevent leakage

3.3.3 Spelling variations

In most reference databases the records are written with American English spelling, but they may also include documents written in British English. Some databases automatically search for both American and British spelling, but most do not. If the database you are working in has this function you will find information about it in the database HELP. To be on the safe side, you should search both words if there is a possibility of spelling variations (Table 4).

Equivalents often contain		For example	
British spelling	*American spelling*	*British spelling*	*American spelling*
S	Z	Computerise	computerize
Ou	O	colour mould	color mold
Re	Er	Fibre	fiber
Gue	G	Catalogue	catalog
Y	I	Tyre	tire
Ae	E	Leukaemia	leukemia
Lling	ling	Modelling	modelling
Or	Er	Divertor	diverter

Table 4. Variations between British and American English spelling

For some words the spelling may also have changed over time; for example, *vitamin* was originally spelt *vitamine*.

You can find a British English - American English dictionary in the library.

3.3.4 Database interface and record languages

From time to time you may notice that the appearance of the interface itself will be changed, perhaps to accommodate new search and display features introduced by the database producer.

Although the great majority of databases use English as the working language, some of the original documents making up the database may have been written in another language. However, the title and sometimes the abstract will be translated into English, and the index terms will always be in English. You should also be aware that there are a few databases which use a language other than English as their working language.

3.3.5 Variations in terminology

The same word can have different meaning in different contexts or different countries (Table 5).

British English	*American English*	*Swedish*
Pavement	Sidewalk	Trottoar
Road surface/surfacing	Pavement	Vägbana
Windscreen	Windshield	Vindruta
Fuel/petrol tank	Gas tank	Bensintank
Bonnet	Hood	Motorhuv
Boot	Trunk	Bagagelucka

Table 5. Examples of different words for the same concept

3.3.6 Acronyms, abbreviations and trade names.

Some terms exist as both abbreviations or acronyms and fully-written words. Note, however, that acronyms can be expanded to very different phrases, depending on the context.

For example:

LED	Light Emitting Diode
LCD	Liquid Crystal Display
PTFE	Polytetrafluoroethylene (Teflon)

Table 6. Common acronyms

Some trade names are very common and should also be included in a search, such as Teflon for PTFE.

3.4 Truncation and wild cards

Truncation means to shorten by cutting off a part; to cut short; to maim, mutilate (Oxford English Dictionary). During a search we often want to use the same keyword to cover a range of inflections (different tenses and cases of the same word), as well as a number of compound words built up from the same word. In these instances we use a "stem" or “root”, which is the part of the word common to all the inflections and compound words, and add a truncation symbol to indicate that one or more letters may be attached to the stem.

The truncation symbol varies depending on the user interface and is often *, ?, $ or #.

In those databases which offer truncation, right-hand truncation is the rule: the truncation symbol instructs the system to look for all the words starting with the stem, followed by a number of letters. Some search systems also allow left-hand truncation, and a few offer simultaneous left- and right-hand truncation.

Here are a couple of examples of right-hand truncation:

climat* finds: climat -e, climat -es, climat -ic, climat -ical, climat -ology, climat -ological

coast* finds: coast -s, coast -al, coast -line, coast -s, coast -er, coast -ing, coast -guard, coast -wise

Example of left-hand truncation:

*national finds: supra –national, inter –national, multi -national

3.4.1 Wild cards

Wild cards are symbols used within a word or phrase to replace one letter by zero, one or two other letters. In the examples below, “?” replaces one letter, “$” replaces zero or one letter.

By writing “wom?n” you will find: "woman" and "women".

By writing “colo$r” you will find: "color" and "colour".

Note that the characters used in truncation and as wild cards differ between reference databases; use the HELP function when in doubt.

3.4.2 Good truncation

Always be careful when truncating. Be especially careful when truncating very short and compound words.

Example: purif* finds purify, purified, purification

pur* finds also puritan, purple, purpose as well as pure and the words above

Poor truncation can give you too many irrelevant references. Sometimes it may be wiser simply to search the terms as synonyms separated by OR.

3.4.3 Autostemming and Lemmatization

Some databases offer a feature called autostemming, which could be considered as an alternative to truncation. When you enter a term, an algorithm will automatically create the word stem for you and then search for variations of the word. In Inspec, for instance, the system automatically autostems all search terms except for Author names and words in quotations and/or braces: for example, management returns manage, managed, manager, managers, managing and management.

If you would like to avoid this, you can select 'Autostemming off'.

Lemmatization is related to autostemming but also finds irregular, but semantically related, variants such as go, going, went and gone.

3.5 Thesaurus check

You should also check which of your search terms that can be found in the reference database's thesaurus. If necessary, modify and complement your search terms with search words or phrases from this thesaurus. In Inspec, you can open the thesaurus by clicking on the tab on the search page.

Your own search terms	*Words or phrases from the thesaurus*
Greenhouse effect Green house effect	Climatology
Global warming	global warming
Climate change Changing of climate	climate mitigation
Coasts Coastline	
Shores Seashore	
Beaches	

Table 7. Examples of thesaurus terms from Inspec

4 The search cycle

4.1 Objectives

In this section you will

- Learn to build your search
- Understand how to refine a search based on the initial results
- Learn to evaluate your results

4.2 Building your search

So far, you have identified the key concepts, created a list of synonyms, chosen a reference database and checked the thesaurus, and you know how to truncate the search terms to optimize your results. Now it is time to create the initial search strings.

When you have two or more different concepts you get a better overview of the process if you search each concept separately. Arrange each search term and its synonyms in a logical OR group and run the

search; repeat with each of the other concepts. You can then test many different AND combinations of these groups to obtain results which appear to solve your search question.

The following is one example of a stepwise search strategy modified for searching in the Inspec database.

1. "global warming"
2. "greenhouse effect" OR "green house effect"
3. climatology
4. climat* NEAR chang*
5. #1 OR #2 OR #3 OR #4
6. coast*
7. shore* OR seashore* OR beach*
8. #6 OR #7
9. #5 AND #8

The advantage of running a search in several steps is that it is easy to re-run the search, correcting mistakes and iteratively adjusting the search strings after the preliminary results have been obtained. Figure 7 shows the results from the individual search steps 1-9, which resulted in 3,957 answers.

Search history Hide

Combine Searches: e.g., (#1 AND #2) NOT #3 Search SORT BY Relevance Publication year

No.	Type	Search	Auto-stem	Sort	Results	Year(s)	Database	Add Email Alert	Save Search
9.	Combined	((((({global warming}) WN ALL) AND (1896-2012 WN YR)) OR ...	Off	Relevance	3,957		Inspec	☐	☐
8.	Combined	(((coast*) WN ALL) AND (1896-2012 WN YR)) OR (((shore* ...	Off	Relevance	95,811		Inspec	☐	☐
7.	Quick	((shore* OR seashore* OR beach*) WN All fields)	On	Relevance	53,402	1896 - 2012	Inspec	☐	☐
6.	Quick	((coast*) WN All fields)	On	Relevance	46,190	1896 - 2012	Inspec	☐	☐
5.	Combined	((((global warming}) WN ALL) AND (1896-2012 WN YR)) OR (...	Off	Relevance	55,671		Inspec	☐	☐
4.	Quick	((climat* NEAR chang*) WN All fields)	On	Relevance	31,120	1896 - 2012	Inspec	☐	☐
3.	Quick	((climatology) WN All fields)	On	Relevance	36,759	1896 - 2012	Inspec	☐	☐
2.	Quick	(("greenhouse effect*" OR "green house effect*") WN All f ...	On	Relevance	1,474	1896 - 2012	Inspec	☐	☐
1.	Quick	(("global warming") WN All fields)	On	Relevance	6,077	1896 - 2012	Inspec	☐	☐

Clear Search History View Saved Searches

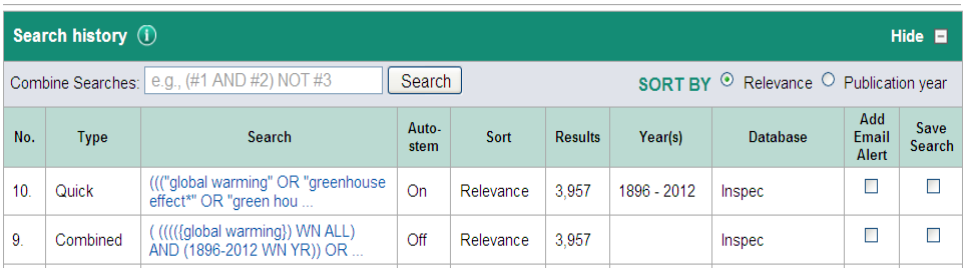
Note: This Search history will contain the latest 50 searches you perform in this session.

Figure 7. Search history from the searches run in Section 4.2

If you choose, instead, to run the search in a single step, using both AND and OR operators, it is important that all the search terms grouped by an OR operator are enclosed within brackets.

10. ("global warming" OR "greenhouse effect*" OR "green house effect*" OR climatology OR (climat* NEAR chang*)) AND (coast* OR shore* OR seashore* OR beach*)

Note that the single-step search (No. 10) resulted in exactly the same number of answers (Figure 8).



Search history Hide

Combine Searches: e.g., (#1 AND #2) NOT #3 Search SORT BY Relevance Publication year

No.	Type	Search	Auto-stem	Sort	Results	Year(s)	Database	Add Email Alert	Save Search
10.	Quick	((("global warming" OR "greenhouse effect*" OR "green hou ...	On	Relevance	3,957	1896 - 2012	Inspec	☐	☐
9.	Combined	((((({global warming}) WN ALL) AND (1896-2012 WN YR)) OR ...	Off	Relevance	3,957		Inspec	☐	☐

Figure 8. Results of combining individual search steps and a single-step search

In Inspec you use double quotation marks “ ” to search for an exact phrase or phrases containing stop words. Other reference databases have similar solutions for this. Use the HELP button when you are searching in a database that you are not completely familiar with.

4.3 Examining the initial results

Once you have run the first search it is time to examine the initial results. How many references did you find? Do they seem relevant to the subject? Are they in a language you understand?

Sometimes the initial search yields very few or no references at all. First check that you have spelled the search terms correctly and that you have placed the operators “AND” and “OR” and brackets, if you use them, correctly. The reason could also be that there is very little written on the subject. More often, however, you have chosen search terms that are too limiting or the wrong reference database for the subject. Study the titles and the keywords associated with some of the references that you have obtained to see if there are *additional* terms or synonyms you can use in the search.

There could be several reasons for obtaining too many results: there may be something wrong with the search strategy, or there is a lot written on the subject, or simply that the search terms you have used are very broad. If you have used acronyms as search terms, you may obtain both relevant and irrelevant results; for instance, NDA means New Drug Application in the pharmaceutical industry, but to a lawyer it can mean Non-Disclosure Agreement. Study the titles and the keywords or index terms associated with some of the references that you have obtained to see if there are *other, narrower* terms you can use in the search.

4.4 Refining and developing

To refine your search, regardless of the number of references you have obtained, it is always a good idea to use more or better search terms.

If you have *too few* references in your initial list you should revise your search question and try to add more synonyms or use broader terms to find more references. You can also investigate if there are more suitable reference databases.

If you have *too many* references you should also revise your search question and try to find more precise terms to use as search terms or perhaps an additional concept.

There are also some simple *quantitative measures* that you can apply on the set of search results in order to judge whether you need to widen your search or to narrow down your search. These measures are normally computed from the variables: total number of hits, number of relevant hits in the hit lists, number of relevant hits in the database, and so on.

Sometimes these measures stem from experimental evaluation of different search systems on special test collections (test/evaluation databases, also known as “corpus” in the humanities). In these test collections we can (by some external procedure) obtain a well-defined set of relevant documents for

each information need, which means that we can compare this set of document with the documents empirically retrieved by different search systems and queries. Two of the most classical quantitative measures are *precision* and *recall*, but many other measures exist.[1]

You can also try one or more of the following:

- Use proximity operators
- Limit the search to the index field (words from the thesaurus, for instance)
- Limit the search to a specific time frame or language
- Require the search terms to appear in the title or abstract

You can also explore the clusters (facets) that many reference database interfaces create from your search results (Figure 9). Each cluster contains documents with one parameter in common, such as an author name, publication year, subject term, and so on. By clicking on one of the titles/links in the Author cluster, for instance, you can immediately limit your results to those written by that author.

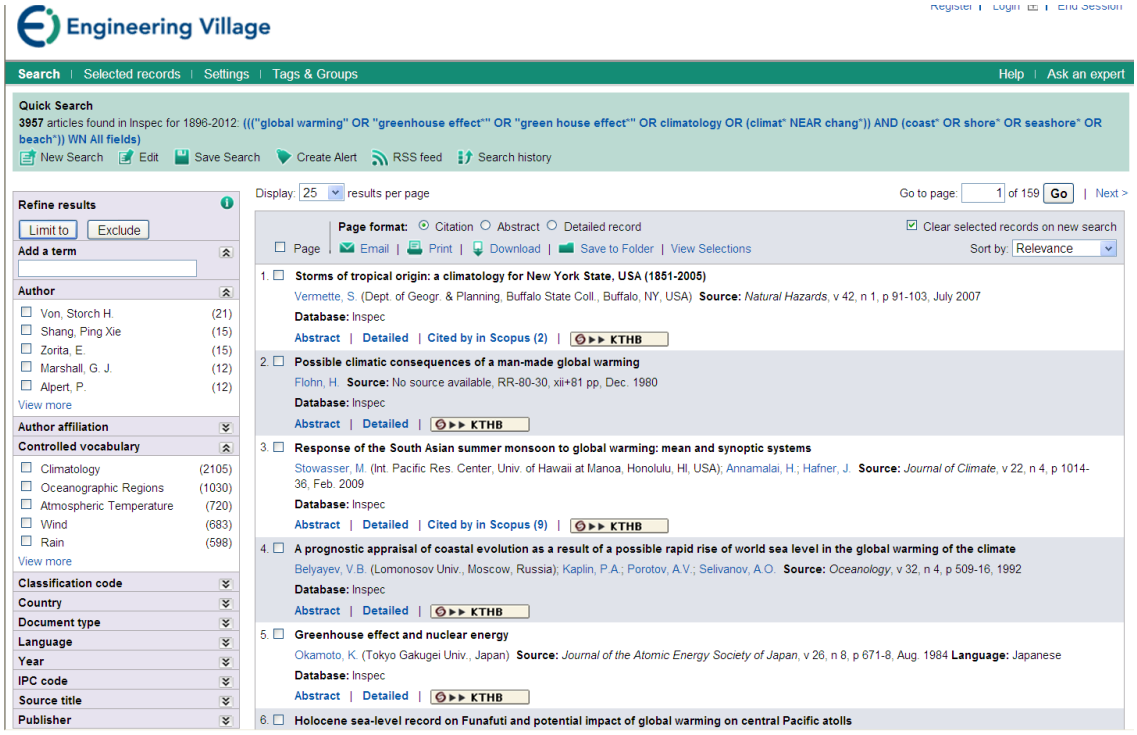


Figure 9. Inspec results page showing clusters in the left-hand column

4.5 Evaluating the results

Now you have to start thinking about quality control. Do your search hits appear to come from reliable sources? Would you feel comfortable about citing them in a report? Always use your own knowledge of the subject to evaluate the information you find.

A separate document on the analysis and evaluation of search results, entitled "Concerning Literary Criticism", is available from the KTH Library website.

[1] As a starting point, view the web page: http://en.wikipedia.org/wiki/Precision_and_recall or the book "Introduction to information retrieval" by Manning, Raghavan and Schütze, published by Cambridge University Press, New York, USA, 2008.

4.5.1 Results from reference database searches

Many of the records in a reference database will be selected from "peer-review" journals, where the articles have been assessed, evaluated and scientifically scrutinized before publication; this is a sign of good quality. However, the same database may also contain articles from journals where this quality-control process has not been applied. Thus, it is your responsibility to critically evaluate the search results. Perhaps the document that you have found is several years old, and there are newer articles which contradict it. If the article you find is only the application of a method someone else has developed, then you will need to find the original article yourself. The journal editors, peer reviewers and database producers that have caused the article to appear in the database are also only human, with their own beliefs and agendas; after all, you can find recent scientific research articles arguing both sides of the climate change issue.

4.5.2 Results from Internet searches

You should be aware that information you find by using Internet search engines may not have been subjected to quality control: anyone can publish whatever they like on the Internet. In addition, the same search can provide different results with different search engines and at different times.

If you plan to cite a document, it is your responsibility to identify who the original author is and to assess the credibility of the material; this is more difficult for web-based documents. The Internet contains a lot of second, third or fourth-hand information. Are you able to check whether, for example, a table or a text is correctly reproduced? It may be poorly reproduced or even intentionally misrepresented. You need to find the primary source to be sure.

If you cannot identify the original author, or if the author does not seem trustworthy on the subject, you ought to view the information as dubious. Do not cite information where you cannot identify the author. You also need to assess the author's credibility. Are they, or do they represent, a reliable authority on the subject? A person's authority is very subject specific. A professor in astronomy may be a very reliable authority on black holes, but not on cancer treatments. If the authors do not seem reliable as a reference, do not use the information.

This does not mean that you cannot use information found on the Internet, just that you have to consider it extra carefully before you do. It is up to you to judge whether the information presented is reliable.

4.5.3 Wikipedia

Wikipedia can be described as an online encyclopaedia, containing free content that, except under certain circumstances, anyone can edit, use, modify, and distribute. Nowadays, Wikipedia is a common resource to use for all kinds of information in a wide variety of languages.

You should not cite a Wikipedia entry as a source of information—not because it lacks a formal reviewing process but because what is published on Wikipedia is not primary information. You can use Wikipedia to look something up and then go to the reference list to (hopefully) find the original source.

4.5.4 Open educational resources

Open Educational Resources (OERs) are seminars, textbooks, courses, etc. that are freely available online. The number of available OERs is increasing rapidly but, as with Wikipedia, you rarely get primary information from these sources. So while many OERs are good tools for learning, they are generally not information sources to be cited.

4.6 Save your search history

In most databases, a search history is generated automatically as you run the search. Many databases allow you to save the search profiles (search terms, operators and limiters) you have used for information retrieval. Note that you may have to register with the database publisher in order to save the search history beyond the current search session. There are several good reasons for saving your search histories:

- A documented search history is an excellent way to remember how you ran your search
- In most cases, you can return to the database at a later date and run the search again without having to re-enter all the words, operators and limiters you used the first time
- If you search stepwise (for example, one group of synonyms followed by another group), you can easily combine or exclude intermediate search results in order to improve the relevancy of your results

In our KTH Library information retrieval courses, we usually require you to show exactly how you ran a particular search. It helps us to understand how you have searched, and it allows us to advise you on how you might search more efficiently.

See Figure 7 and Figure 8 for examples of search histories.

4.7 Selecting documents to read

Once you are satisfied with your search results, you need to select some documents to read. Based on the documents' titles and abstracts you should select a number of documents to retrieve in full text. If you are a newcomer to the area, it is a good idea to start with review articles or reference works, which will provide you with background and history of the subject. You can then progress to the narrower journal articles for more in-depth information about a particular aspect.

Most databases can help you with this selection process, by allowing you to filter your results in different ways, such as type of document, publication date or relevance, etc. (Figure 9). In this way it is easy to make sure that the documents you choose for reading are written in a language you understand and are not out of date. It is also a good idea to mark all those records that seem particularly interesting and to save them to a separate list. Once you have read the first few articles, you will realize if you need to get more full text documents from your search results (also a very good reason to save the list) or if you need to run a modified search in order to better match your information needs.

The KTH Library has agreements with a large number of publishers which allow active students and staff at KTH to use their kth.se account to download full text copies of journal articles, most often in pdf format; look for the Adobe Acrobat Reader or pdf symbol. If the full text is not available from the publisher, you can order a copy through the KTH Library Article Service (and you may have to pay for it). See Section 5.4 for more information.

4.8 Saving to reference management software

As a final step you should export the bibliographic information for your selected references to a reference management program; this makes it easier to keep track of your references and to insert them into your report once you start writing. Most reference databases have a function that allows you to directly export the references in the correct format for a variety of reference management systems. For more information, see the section on Reference Management.

5 Accessing full text documents

5.1 Objectives

In this section you will

- Learn to navigate among the different full text options available through the KTH Library
- Learn about document identifiers
- Learn how to order full text documents not available online or directly at the library

5.2 Finding original documentation

There are different options for finding the original documentation. Depending on what you are looking for, you can search in:

- Reference databases, journals and patents
- Library catalogues, for example the KTH Library catalogue, Primo or LIBRIS
- Full text archives, for example DiVA and SwePub

5.2.1 Reference databases – find journal articles and patents

Use the reference databases to find journal articles, review articles and patent documents. Most databases have links either to full text in pdf format or to the publisher's web site.

The KTH Library subscribes to a large number of journals, and the full text articles or reviews from many of these journals are directly available online.

If you know exactly which article you are looking for, you can also try to search for it directly on the journal's website, in ScienceDirect or in Google Scholar.

Full-text patent documents from countries and patent authorities around the world are available free-of-charge through the Esp@cenet database at the European Patent Organisation.

5.2.2 KTH Library catalogue

The library catalogue will tell you where to find a specific printed book, report, dissertation or journal and contains direct links to e-books and e-journal archives. You enter search terms (words or phrases) describing the area you are interested in; the words may be in the title or in the catalogue index terms or may be the author's family name. Note that you cannot search for individual journal articles in the library catalogue as you can in a reference database.

5.2.3 Primo

Primo is a search tool that provides a single access point to the KTH Library's online and print collections including

- scholarly journal articles
- print and electronic journals
- some reference databases
- print and electronic books
- conference proceedings
- theses and dissertations and more.

Primo contains everything found in the KTH Library Catalogue, plus millions of full-text articles. It retrieves citations from most, but not all, of the library resources. Just enter your search terms and use the options on the left-hand side of the results screen to refine your search. From there you can quickly get to the full-text online or see if the KTH Library has a print copy.

5.2.4 LIBRIS

If you cannot find the book or periodical you seek in the KTH Library, or if it is out on loan, try searching in LIBRIS. This is the search service of the National Library of Sweden (*Kungliga biblioteket*), and allows you to find documents held by all Swedish university libraries, most Swedish research libraries and a few public libraries. KTH's own collection of electronic journals and e-books is also accessible in LIBRIS. See below for rules on interlibrary loans via LIBRIS.

5.2.5 DiVA and SwePub

Digitala Vetenskapliga Arkivet (*Digital Scientific Archive*) is the Swedish national publication database, which comprises information on undergraduate and doctoral theses, reports, journal articles, conference papers, books and patents. The KTH component of DiVA contains a steadily increasing number of references to theses, articles and other documents (some in full text) written by KTH researchers and students. This is partly a consequence of the KTH publication policy, which states that KTH theses should be published open access.

SwePub is the portal to about 30 of the Swedish university publication databases. You can limit your search to a single university or you can search several universities at the same time. Note that SwePub contains articles, doctoral theses and conference papers. If you wish to find degree project reports, you can search in Uppsatser.se (theses written in Swedish) or Essays.se (theses written in English). You can find resources such as such as LIBRIS, KTH DiVA and SwePub, via the database list under the "Search Tools" tab at the KTH Library website.

5.3 Document identifiers

5.3.1 Digital Object Identifier, DOI

Many scholarly publishers use the Digital Object Identifier (DOI®) System to uniquely identify electronic documents, such as journal articles and books or book chapters, and where they can be found on the Internet. The idea is that although information about a document may change over time, its DOI remains the same and will always lead you to the document.

One example is doi:10.1038/nature10112, which refers to an article published in Nature on June 8, 2011. To go directly to the article, place http://dx.doi.org/ before the DOI address, thus: http://dx.doi.org/10.1038/nature10112. You can also use DOIs to locate a specific scientific article using Google: simply search for either doi:10.1038/nature10112 or just enter 10.1038/nature10112.

Although you seldom find a DOI in a reference list, it is a useful way to cite an "article in press", which is an article that has been accepted for publication by a journal but has not yet been published. Such an article is usually available as a pre-view online on the journal's website before it is officially published in the journal and the DOI assigned to the pre-view version remains after the article is published.

5.3.2 ISSN, ISBN and ISRN

These are unique numbers for identifying a specific journal, book or report, and can be used for effective searching in library catalogues:

- ISSN - International Standard Serial Number, identifies a specific journal.
- ISBN - International Standard Book Number, identifies a specific book.
- ISRN - International Standard Report Number, identifies a specific report.

The same publication can have more than one ISN; a report can also be a book and may then have both an ISRN and an ISBN. There are different ISBNs for the paper-back, hardcover and electronic versions of a book.

5.4 Ordering documents

When the document you are looking for is not available at the KTH Library you can order it. If a book is on loan you can reserve it or order it from another library if it is not held by the KTH Library. Specific journal articles can also be ordered.

5.4.1 Articles

The KTH Library subscribes to many, but not all, journals. A majority of these are available online and are free to download for students and staff at KTH with a valid library card. Many publishers have created backfiles, scanning old issues of journals to make them available online in addition to the new material.

Most of the journals delivered in printed form to the KTH Library have been relocated to the external stacks in Bålsta; the only ones accessible in the KTH Library building are those issued since 2005.

If you wish to obtain a copy of a journal article which was published before 2005, or which is published in a journal not available through the KTH Library, you can order it via the KTHB Electronic Library: follow the links *Services* > *Loans and Copies* > *Article Copies.*

For copyright reasons, the KTH Library can only supply you with a paper copy of an article, not an electronic copy.

5.4.2 Books and reports

Once you locate the publication, you can apply for an interlibrary loan via LIBRIS. Please note that the KTH Library will not accept requests from students for interlibrary loans of books that are held by other libraries within Greater Stockholm or for course textbooks held at one of the local KTH libraries; you must visit the library in question yourself. Nor will we accept loan requests for books which are held by the KTH Library but are currently out on loan. If the publication is a periodical, you can order a copy of the article instead.

You can place yourself on the waiting list for a book which is out on loan; when the guaranteed minimum loan period is up, the current borrower will be asked to return the book to the library for the next person in the queue.

5.4.3 Costs

You may borrow books and reading articles and other material that is available in the library's collection, online or at the library, without cost. You may also borrow books from other libraries within the Nordic countries through our interlibrary loan service (via LIBRIS) and request an article from the external stacks free of charge. However, you will have to pay for articles that are not available in the KTH Library collection and for books that we need to borrow from libraries outside the Nordic countries. You can find the price list on the KTH Library web site.

6 Using information

6.1 Objectives

In this section you will

- Understand the need to cite your sources and what information is needed in a reference
- Know how to avoid plagiarism
- Learn about reference management software

6.2 Copyright

In Sweden the Act on Copyright in literary and artistic Works [SFS 1960:729] states, in the official English translation, that

"...Anyone who has created a literary or artistic work shall have copyright in that work, regardless of whether it is

1. "a fictional or descriptive representation in writing or speech,
2. "a computer program,
3. "a musical or dramatic work,
4. "a cinematographic work,
5. "a photographic work or another work of fine arts,
6. "a work of architecture or applied art,
7. "a work expressed in some other manner."

"Maps and other works of a descriptive nature executed as drawings, engravings, or in a three-dimensional form, shall be considered as literary works."

The owner of the copyright shall have "the exclusive right to exploit the work by making copies of it and by making it available to the public, be it in the original or an altered manner, in translation or adaptation, in another literary or artistic form, or in another technical manner." In addition, the making of copies means "any direct or indirect, temporary or permanent preparation of copies of the work, regardless of the form or through which method this is carried out and regardless of whether it concerns the work in whole or in part."

In other words, a writer, photographer, sculptor or other person who creates a work of art owns the copyright to that work of art. The copyright owner can prevent other people from making and selling copies of that work, either in original form, translated into another language or converted into another medium, for example, from a book to a film. The person owns the copyright during his or her lifetime, and the descendants for 70 years after the death of the copyright owner. However, the copyright owner may grant other people the right to make copies or modify the original work in return for payment.

Most other countries have similar laws.

In Sweden there are three organisations which look after copyright and ensure that the original copyright owners receive compensation for the use of their works: STIM (*Swedish Performing Rights Society*) grants permission to perform music publicly, and collects and distributes remuneration to the performers, composers and songwriters; BUS (*Bildkonst upphovsrätt i Sverige*) is the corresponding organization for original art and Bonus Presskopia collects, administers and distributes remunerations for copying from books, newspapers and periodicals and other literary and artistic works especially for educational purposes.

In accordance with the Swedish Act on Copyright in literary and artistic Works: making a copy of an entire book is forbidden, copying is only permitted for limited parts of books or other literary works, the copy should be for private use only, and may not be sold to another person.

The KTH Library has agreements with the suppliers of electronic books and electronic journals which limit the amount of each which may be copied.

KTH has an agreement with Bonus Presskopia covering reimbursement of documents which are copied by teachers for use as educational purposes, which states that a teacher may copy not more than 15 % of a book, or 15 pages.

For more information, follow the links to:

- Act on Copyright in literary and artistic Works [SFS 1960:729]
- STIM
- Bonus Presskopia

6.3 Why cite?

When you write a piece of work, for example a report, an essay, or when you create a web page, you often use or refer to material created by someone else. This material includes for example, quotes, pictures or diagrams. When you do this, you are morally bound to provide the name of the original author or source.

The aim of a reference list is to make it possible for the reader to find the publication, the original document, you are referring to. Therefore the written reference must contain enough information to allow someone else to find the document. There are many different standards for writing references but the differences are merely in how the information is organized; the contents are the same.

For printed material you need to include author, title, publication year and source for the document you are referring to in your reference list. The information you need to provide about the source varies with the type of publication:

- Books - name of the publisher, the publishing area, part, volume or print/reprint and ISBN
- Journal articles - journal title, volume, number/issue and page reference (sometimes ISSN)
- Reports - publisher, publishing area and report number
- Conference paper - conference name, title, venue, date, pages, publisher, ISSN/ISBN

Similar rules apply to web pages that you refer to. However, since web pages are being constantly updated and moved about, it is not certain that you will be able to find the same web page again in a later search. For this reason you should provide the following data (where known) when you refer to sources of information found on the Web:

- Name(s) of author(s)
- Title of the article within quotation marks
- Original publishing date/latest update of the web page
- Name of the organisation that owns the web page
- When the page was last updated
- Date you visited the web page
- URL (Web page address)

For example:

“Finding information in science and nature – OpenLearn – The Open University” 2012. p. Private course on information searching. http://openlearn.open.ac.uk/course/view.php?id=2376 (accessed 21 August 2012).

6.4 Plagiarism

Plagiarism is defined as submitting someone else’s work as your own. It is a form of cheating, and can have serious consequences. KTH has published guidelines on “Guiding Students Away from Plagiarism”, together with a list of contact persons at each School to whom you can turn if you have questions about plagiarism. A rule of thumb is that as soon as you include in your essay or report something that you have not written or created yourself, you should include a reference to the original source.

6.5 Reference management

One way to avoid plagiarism is to ensure that you collect and maintain a reference list as you work through your project.

There are many software programs which can be used to manage your references. Some examples are EndNote, Zotero, BibTeX and Mendeley. Most of these programs work in similar ways, and instruction manuals/tutorials can be found freely on the web.The programs are usually web-based, although some offer a client version with extra functionality or are available as an app for your mobile phone or tablet. The KTH Library offers courses in all these programs on request.

Many reference databases provide applications which allow you to directly export bibliographic information to a reference management system, thereby making it is easy to store the correct information about a reference. Once the references are stored in the system it is easy to organize them and keep track of the references for a specific project. Often, you can share references with other members of your group.

The reference manager also provides a possibility to present the references in different standard formats in your text document, in order to suit a wide range of scientific journals. You also have the possibility to create your own output styles.

There are different standards for writing references. Ask your supervisor or other members of staff which standards are used within your institution. If you are writing a research paper for a journal, check the Information for Authors and the reference lists for a couple of recently-published paper to see how the references are written in that particular journal.

7 Summary

Now that you have read through the whole material you should feel more secure about how to find and manage information. By now you should know the difference between different types of information, the fact that you need to evaluate your sources and that you should only cite primary information.

You have learnt to create a search question, using synonyms, thesauri, Boolean operators, proximity operators and truncation to optimize your results and to avoid linguistic pitfalls. You have also read about how reference databases are constructed, and the pros and cons of using field searching.

You know how to find full text documents online and in the library and you know how to order what you can't find yourself. You have also learnt why, when and how to cite, in order to avoid plagiarism and how a good reference manager can help you with this.

7.1 Summary for how to create a search question:

- Define the search request as precisely as possible
- Choose information resources (reference database, library catalogue, Internet)
- Choose relevant search terms
- Modify the search terms to suit the chosen information resources
- Combine the search terms to create a search query
- Run the search
- Refine your search query based on the results and run the modified search
- Run your modified search, possibly also in additional databases
- Save your selected search results in a reference management system.

Example: "Find information on substances used for the accelerated healing of eye wounds". Table 8 shows which search terms you can extract from the question, possible synonyms, and how you can use truncation to find alternative forms of the search terms.

Keyword	*Synonyms*	*Truncations, where #=0-1 letter, * 0 or many letters*
Healing	Treatment	Heal* (heal, heals, healed, healing) Treat*(treat, treats, treatable, treated, treating, treatment)
Eye	Ocular, cornea	Eye# (eye, eyed, eyes) Ocular Corne## (cornea, corneal)
Wounds	Trauma, injury, regeneration	Trauma* (trauma, traumas, traumatic, traumatized) Injur* (injure, injured, injury, injuries) Regener* (regenerate, regenerated, regeneration, regenerative)
Accelerated	Rapid	Rapid* (rapid, rapidly, rapidity, rapids) Accelerat* (accelerate, accelerating,)

Table 8. Possible terms, synonyms and truncation for a search on healing of eye wounds

Always remember the HELP button! Different search interfaces employ different search strategies. Therefore, if you want to search proficiently and use the search tools properly, click on the HELP button and read the instructions provided.

8 The authors

The original version of this material was created in 2000-09-01 by:

Lena Fahlén, Information Specialist; Economics & Commerce
Andrzej Olkiewicz, Information Specialist; Mechanical Engineering & Geology
Janet Wamby, Librarian
Christian Åström, Web Programmer

It has since been updated by:

Lena Fahlén, Information Specialist; Economics & Commerce

Kjell Jansson, Information Specialist; Physics & Mathematics

Christian Åström, Web Programmer

Göran Hamrin, Information Specialist; Mathematics, Philosophy and Computer science

Peter Engström, Information Specialist; Construction, Transportation and Urban planning

The 2012 major update and reconstruction was done by:

Roger Bishop, Information Specialist; Biochemistry

Mikael Danielsson, Information Specialist; Construction, Transportation and Urban planning

Maria Unger, Information Specialist; Analytical and Environmental Chemistry

Rosa Lönneborg, Information Specialist; Biochemistry

Thomas Nyström, Librarian and Information Specialist; Computer Science

Bengt Haraldsson, Information Specialist; Physics

Göran Hamrin, Information Specialist; Mathematics, Philosophy and Computer science

The 2013 minor update was done by:

Roger Bishop, Lecturer and Information Specialist; Biochemistry

Göran Hamrin, Lecturer, Librarian and Information Specialist; Logic, Mathematics, Philosophy and Computer Science

The 2014 minor update was done by:

Göran Hamrin, Lecturer, Librarian and Information Specialist; Logic, Mathematics, Philosophy and Computer Science

9 Acknowledgements

We would like to express our gratitude towards all our friends and colleagues at the Royal Institute of Technology Library for their kind help and assistance in creating this introduction to information retrieval and management.

10 Index

The definitions of terms in this index are collected from online encyclopaedias such as Wikipedia or the Online dictionary for Library and information Science.

Bibliography: A list of the books of a particular author, printer, or country, or of those dealing with any particular theme; the literature of a subject.

Conference proceedings: A collection of academic papers that are published in the context of an academic conference

Coverage: The extent of the information found in a database; how much of a subject area that is covered; which years etc.

Encyclopaedia: A type of reference work holding a summary of information from either all branches of knowledge or a particular branch of knowledge. Encyclopaedias are divided into articles or entries, which are usually longer and more detailed than those in dictionaries.

Handbook: A type of reference work that often contains data that are useful within a particular field or data about a particular technique. Could also be referred to as a manual.

Information retrieval: The activity of obtaining information relevant to an information need from a collection of information resources.

Information Society: A society where the creation, distribution, diffusion, use, integration and manipulation of information is a significant economic, political, and cultural activity. The aim of the information society is to gain competitive advantage internationally, through using information technology (IT) in a creative and productive way.

Inter library loans: Loans between libraries. The KTH Library provides inter-library loans for KTH students and employees for books that are not in the KTH Library's own collections.

Internet: Global network of computers using the IP protocol for communication.

Iterative: Describes a process that is repeated, usually with the aim of approaching a desired goal. In the context of information searching, an *iterative search process* refers to a set of search queries on a subject, where new terms are added to the queries and the search syntax is systematically revised in light of the previous search results.

Peer-review: The process of subjecting an author's manuscript to others who are experts in the same field, before an article describing this work is published in a journal. The experts (referees) may accept it, suggest changes or reject the manuscript based upon its scientific quality.

Prerequisite: Required as a prior condition.

Pros and cons: Advantages and disadvantages.

Reference work: A book or serial publication to which you can refer for confirmed facts e.g. catalogues, dictionaries, encyclopaedias etc. The information is intended to be found quickly when needed.

Scientific journal: Journal where scientists publish their results. The articles are often peer-reviewed.

Serial publication: Materials in any medium issued under the same title in a succession of discrete parts, usually numbered (or dated) and appearing at regular or irregular intervals with no predetermined conclusion.

Synonyms: The linguistic definition says that synonyms are words with identical or similar meanings. In the context of information searching however, words are also considered synonyms if they can describe aspects of the same concepts.

Technical terms: Words with a specialized meaning in a particular area of expertise.

Trade journal: A magazine published with the intention of target marketing to a specific industry or type of trade.