

Finding Technical Information Over the Internet

Brenda F. Bell
CSA/Aerospace
59 John St., 7th Fl.
New York, NY 10038-3709
bbell@csa.com

Most of us have tried (at least once!) to find something highly specific over the Internet – sometimes to no avail. If the information *is* indeed available over the ‘Net (and most things are), your difficulty finding it may have to do with how you look for it – or *where* you look for it. This is especially true if you’re looking for technical information, such as the latest research on a new material or tests on a potential AIDS vaccine. In the following pages, we’ll take a look at places that specialize in providing well-indexed – or at least searchable! – technical information.

I. Introduction to Searching

No matter where you search for information, you need to have a coherent *search strategy* in mind before you sign on to a site. This is particularly true with some of the transactional database aggregators, where charges can mount up dramatically with each minute you’re on the site, each database you search, and each document you view. So before we start looking at search sites, database providers, and publishers, let’s go over some of the basics of searching.

First, you want some idea of what you want to find, and how to express it in text form. This can be as simple as “*What is quantum computing?*” or as complex as “*How would a missing heat tile have affected the Space Shuttle **Columbia**’s thermal profile at the point of reentry?*” It may help to write this down on a scrap of paper that you can refer to as you craft your search.

Now look at that paper and isolate the words and strings (phrases) that define what you want to find out. In the first case, it’s pretty simple: you want to find out about *quantum computing*. In the second, there are several concepts that need to be found **at the same time**: *Space Shuttle **Columbia***, *heat tiles*, and *reentry thermal profiles*. Making it more difficult, an author might express these concepts in several different ways. For example, *Space Shuttle **Columbia*** might be expressed as “*U.S. Space Shuttle*”, “*NASA Space Shuttle*”, “*Space Transportation System*”, “*STS-107*”, “*Space Shuttle Orbiter 102*”, or “***Columbia** Orbiter*”, among others.

Once you have these basic concepts identified, you can begin to craft your strategy.

A. Approach: *Granularity* (Specificity)

At this point, you’ll want to identify several levels of specificity – called *granularity* in search terminology – and decide which level you’ll do best to start with. The most usual strategies are (1) starting with the most general, and progressing to the most specific, and (2) starting with the most specific, progressing to the most general. Each method has its place, as does starting somewhere in the middle and choosing a direction (more specific or more general) based on the results of that first search.

Most of the time, general searches work best if you don’t know much about the topic you’re searching for – for example, “*What is quantum computing?*” From your general results, you will

hopefully get some ideas of what to include and what to exclude to find what you're really after. Highly-specific searches work well if you know in advance not just what you're looking for, but that you will find it where you'll be looking for it. For example, you're more certain to find something on **Columbia's** entry profiles in a database that specializes in aeronautics and astronautics – such as [CSA's Aerospace and High Technology Database](#) – or on a site devoted to the technical aspects of space flight, such as <http://www.space.com/> or any of a number of [NASA](#) sites, or the highly-specific [Columbia Accident Investigation Board](#) site.

Most of the time, you'll want to start somewhere in the middle – searching for thermal profiles of *any* of the Space Transportation System orbiters, for example – and going more- or less-specific, depending upon your results.

B. Syntax: Boolean Logic

The next thing to keep in mind is the syntax you'll need to search a site. Yahoo!, AltaVista, and most database aggregators allow you to use some form of Boolean logic as well as any site-specific syntaxes they may provide. In a Boolean search, simple operators such as AND, OR, and NOT help refine a search. For example, you might search for

Columbia AND shuttle to search for the Space Shuttle

or

Columbia NOT gem to avoid lyrics for *Columbia, the Gem of the Ocean*

Using Boolean logic, you can even search for

(Columbia AND “space shuttle”) NOT “Gem of the Ocean”

to look for a generic reference to a space shuttle, specific reference to **Columbia**, and avoid those pesky song lyrics!

Some sites also give you the options of Boolean operators NAND (“not AND”, or “and NOT”), NOR (“not OR”), and NEAR (proximity) to help you further refine a search.

N. B. Google uses a default AND search. You cannot perform an OR search on Google.

C. Search Aid: Controlled-Vocabulary Indexing

Let's say what you're **really** looking for is a graphic on a set of equations that show you how the temperature along **Columbia's** wing surface varies with location at a specific time into the reentry. You might refer to this as a “temperature profile”, a “thermal profile”, or a “temperature distribution” – however, if you search for “thermal profile” and the author of the document you're looking for wrote “temperature distribution”, you might have a problem finding the document.

Enter the technical indexer and the controlled-vocabulary index. The indexer's job is to determine the basic concepts of what the document (or part of a document) is about, and to assign to that item a number of key words and phrases by which it can be retrieved. A **controlled vocabulary**, or **thesaurus**, when it is used, restricts the indexer to specific wordings for each concept included in the thesaurus. In the above example, if the thesaurus term for “temperature distribution” is “thermal profile”, then regardless of what the author wrote, you will be able to find the document using the term “thermal profile”.

“How do I know what wording the database uses for this concept?” you might ask. If you can't find anything about the specific thesaurus on the database's site, you might want to speak with your corporate librarian or call the database provider's customer service line. If the database

you're searching uses one of several standard thesauri (such as the *NASA Thesaurus*), you may be able to look for that thesaurus (either in your library or online) and browse, borrow, or purchase a copy of it.

D. Strategy: Field Search

Because most database records are composed to fit a specific structure, a database usually gives you the option of limiting your search to specific fields of that database. For example, in CSA's **Aerospace and High Technology Database**, I can search for an article on Space Shuttle **Columbia** by searching for the specific index terms "Columbia Orbiter" or "Space Shuttle Orbiter 102" – or I can look for "Columbia AND (space shuttle)" in either the title or the abstract.

To most effectively conduct a field search, you'll want to know which of a database's fields are searchable, what their names are, and how to write the field-search query for the database you're searching. Fortunately, most database aggregators have a number of ways for you to find this information. We'll come to that later.

II. A Sample Search

In order to get a better example of what technical databases can provide as opposed to the Web at large, we're going to do some sample searching. Since we're looking for something with a bit of academic probity, I'm going to go trawling for something specific and in the news: the relationship of wing thermal profiles to the Space Shuttle **Columbia** disaster.

A. Searching on Yahoo!, Lycos, Google...

Most of the main search engines and Web directories have a common symbology for expressing Boolean logic. The operators + (and) and - (not) are common to a number of default-OR search engines.

Using these expressions, I searched Yahoo!, Lycos, and Google using the following strategies:

Columbia AND wing AND temperature distribution

Space shuttle AND thermal profile

Space shuttle AND disaster

The first pages of the results may be found in attachments 1, 2, and 3.

1. Problems with searching on Yahoo!, Lycos, Google...

a. Irrelevant Results and False Hits

- The first problem most of us have noticed in searching the publicly-available search engines and directories is the large amount of extraneous data in the search results. For a search on either of the Space Transportation System disasters, or the 9/11/01 tragedies, you're likely to find a lot of personal-recollection sites and memorial sites. The increasingly-commercial nature of the Internet makes it likely that you'll find a number of "sponsored sites" – where the siteowner pays for a top position – at the top of your search results, which may or may not be relevant to what you're looking for. In addition to these legitimate but irrelevant results, you may still come across a number of false hits

from sites promoting pornography, discount-rate pharmaceuticals, financial “services”, and other completely irrelevant things.

- Even if you can completely eliminate these results, you may still have a problem with what’s left.

b. Academic Probity of Relevant Hits

- While more and more researchers are publishing their results directly to the Web, there are still a number of corporations, research labs, and universities which, while publishing their current findings, wish to keep their corporate intelligence and proprietary information... proprietary. Persons working for those organizations are more likely to present their results in papers presented at technical conferences or published in technical journals than they are to make them freely available on the Web.
- On the other side of the coin, if you’re a serious researcher, you want to know that the information you’re using as a basis for a grant request or a contract proposal is legitimate information, that the source of the research is reliable, that the methods used are reliable and acceptable to the field of endeavor, that appropriate and sufficient information is collected from the research, and that the conclusions are consistent with the information collected. In short, you want to make sure that the paper has been reviewed and/or **refereed** by one or more independent parties knowledgeable in the field.
- If you wish to restrict your search to refereed sources, or if you need to research items published in technical journals and conference proceedings, you’ll generally need to look them up either directly from the publisher, or you’ll need to go through a technical database (or set of databases) geared towards the field in which you’re looking.

B. Searching Publisher Sites

One place where you can get information from refereed sources is to go directly to the publishers. Kluwer, Wiley, and Springer all have extensive sites where you can search their journals, monographs, and proceedings. In some cases, you’ll need a subscription to access titles, abstracts, and/or full-text (HTML or .PDF) of the articles and papers you’re interested in.

Since primary publishers usually secure reprint rights for the items they publish, you should not have problems getting copies of the documents you require.

You should also be able to talk to a primary publisher’s rights department to secure permission to quote an article or reprint a graph as necessary.

1. Problems With Searching Publisher Sites

- You may need a subscription to search the site. In some cases, the base levels of subscription are free (Wiley, Institute of Physics Press, and SPIE all allow access to titles and abstracts for free), but deeper levels require paid access.
- Some publishers charge for access to abstracts (IEEE).
- You usually can’t access multiple publishers’ publications from a single site. You may or may not be able to search across a publisher’s various imprints (e.g. Springer’s Springer, Springer-Verlag, and Birkhauser lines)
- You may need to know in advance that what you’re looking for (a specific journal title, for example) comes from that publisher, or you may only be able to search one journal title at a time

C. Searching Aggregator Sites

One of the newest classes of entrant into the technical-search field is the “aggregator”, which is a secondary publisher which breaks up primary-publication information into “chunks” of information tightly indexed and with the appropriate copyright/source information attached. As of last year, *World of Science* was one such aggregator. (They appear to have either folded or been merged into another company.) Some primary publishers, such as Kluwer, have been working at “infochunking” their own publications to create a similarly-searchable database (in Kluwer’s case, Science Direct <http://www.sciencedirect.com>).

1. Problems With Searching Aggregator Sites

- Not all aggregator sites give access to an entire article, just to paragraphs and parts of the article.
- Some primary publishers refuse information or limit the rights of aggregators to include information from their publications
- Some sites require subscriptions
- Document-delivery services may or may not be available

D. Searching Technical Databases and Database Collections

Before the Internet became the universally-accessible information-and-communications source it is today, database collections – and before them, abstracts journals – were one of the most likely sources from which to find information on technical publications. While the field has narrowed, the services these players offer have broadened to include full-text, author contact information, and increasingly varied methods of document delivery.

Some of the major players in the database collection/database aggregator segment include Ei (Elsevier Engineering Index, <http://www.ei.org/eicorp/eicorp>), DIALOG (<http://www.dialog.com>), and Silver Platter (http://www.ovid.com/site/products/tools/silverplatter/access_tools.jsp?top=2&mid=3&bottom=8&subsection=15), as well as the company I work for, Cambridge Scientific Abstracts (CSA) (<http://www.csa.com>). Ei and CSA are a bit of the oddballs in this segment, as both companies produce a number of the databases they host. (I work on the team that produces CSA’s **Aerospace and High Technology Database**.)

1. Advantages to Searching Database Collections

- Academic probity: source documents are heavily weighted towards refereed journals and conference proceedings
- Multiple databases allow you to search multiple disciplines at a time
- Controlled-vocabulary indexing means that as long as you use the thesaurus, you **will** be able to find what you’re looking for
- Human indexers and abstractors oversee the production of each record.
- Created **for** the technical community, **by** the technical community

2. Disadvantages to Searching Database Collections

- You need an account or a subscription for each database or set of databases you wish to search, which can run into the thousands of dollars per year

- The collections provide abstracts; full-text, facsimile, and .pdf document delivery (if available) cost more
- CD-ROM-based database collections (like Silver Platter and DIALOG on Disc) may only allow you to read one disc or one database at a time
- The amount of time it takes to acquire, select, and process a conference proceedings or a journal issue may make the database less current than other collections.

III. Where to Find Technical Information Online

A. Primary Publishers Available Online

1. Freely Available

These publishers allow you to access at minimum titles and abstracts for their publications, without charging a fee. They may, however, require you to register with them first.

a. Institute of Physics Press

- Home page: <http://www.iop.org>
- Journals: <http://www.iop.org/EJ/>
- Books: <http://www.iop.org/bo.html>

b. SPIE

- Home page: <http://www.spie.org>
- Journals Online available through the American Institute of Physics (AIP) [Online Journals Publishing Service](#)
- Search Proceedings Volumes and read abstracts: <http://www.spie.org/app/Publications/index.cfm?fuseaction=proceedings>

c. Wiley

- Home page: <http://www.wiley.com>
- Browse titles: <http://www.wiley.com/cda/sec/0,,918,00.html>
- Journal abstracts on Wiley Interscience: <http://interscience.wiley.com/>

Subscriptions are available for fulltext of journals and books

d. Kluwer

- Home page: <http://www.kluweronline.com>
- Browse journal titles and abstracts: <http://journals.kluweronline.com/>
- Read abstracts and purchase e-books: <http://ebooks.kluweronline.com/>
- Paid-subscriber access to reference works: <http://reference.kluweronline.com/>

Custom books and pay-per-view available

e. Elsevier

- Home page: <http://www.elsevier.com>
- Browse all publication types from here: <http://www.elsevier.com/inca/tree/>

- Browse journal titles and abstracts:
<http://www.elsevier.com/inca/tree/?prod=J&key=SSAN>
- *ContentsDirect* delivers titles and abstracts by e-mail:
<http://contentsdirect.elsevier.com/Welcome.html>

f. Springer-Verlag

- Home page: <http://www.springer.de>
- Journal and book abstracts by paid subscription

g. Russian Academy of Sciences

- Home page: <http://www.maik.rssi.ru/>
 - Search for journals:
<http://www.maik.rssi.ru/eng/journals/index.htm>
- Abstracts available free for online journals

2. Available to Members and/or Paid Subscribers

a. IEEEXplore

- Home page: <http://www.ieeexplore.ieee.org>
- A service of the IEEE <http://www.ieee.org>
- Journal and Conference abstracts, fulltext, and .pdf files
- Institutional and individual subscriptions
- Access determined by subscription, Society membership, purchase of Conference Proceedings volumes

b. Wiley Interscience

- Home page: <http://interscience.wiley.com>
- Fulltext and .pdf access available by subscription

c. Science Magazine

- Home page: <http://www.sciencemag.org>
- Subscriber access to current events and to abstracts for all journals published
- Fulltext and .pdf access to subscribed publications
- Individual and institutional subscriptions

d. Nature Publications Group

- Home page: <http://www.nature.com>
- Subscriber access to abstracts for all journals published
- Fulltext and .pdf access to subscribed publications
- Individual and institutional subscriptions

e. AIAA

- Home page: <http://www.aiaa.org>
- Publications page: <http://www.aiaa.org/Research/index.hfm?res=0>

- Journal abstracts available at <http://www.aiaa.org/Research/index.hfm?res=9>
- Full-text articles available to subscribers
- Meeting papers available online to subscribers

B. For-Pay Database Collections

1. Cambridge Scientific Abstracts (CSA)

- Home page: <http://www.csa.com>
- Database list: <http://www.csa.com/csa/ids/databases-collections.shtml>
- Products page: http://www.csa.com/csa/e_products/elprod2.shtml
 - Internet Database Service: <http://www.csa.com/csa/ids/ids-main.shtml>
Home to over 50 databases published by CSA and its publishing partners. Available by subscription. Includes indexed Web Resources as well as journal- and conference-based records.
 - Hot Topics: <http://www.csa.com/csa/hottopics/hottopics-main.shtml>
Papers on topics ranging from [Alzheimer's Disease](#) to [Global Warning](#) to [Quantum Cryptography](#)
 - BiblioAlerts: <http://www.biblioalerts.com/>
“Canned” bibliographic searches and abstracts for selected topics.
- Sales contacts: <http://www.csa.com/csa/support/contact-csa.shtml#sales>

2. DIALOG

- Home page: <http://www.dialog.com>
- Products page: <http://www.dialog.com/products/productline/dialog.shtml>
 - Products for multiple market sectors
 - Multiple delivery products and options
 - Web
 - Intranet
 - CD-ROM
 - Telnet

3. Ovid

- Home page: <http://www.ovid.com>
- Products page: http://www.ovid.com/site/products/product_catalog.jsp?top=2&mid=3
- Products available over the Internet or on CD-ROM

4. Silver Platter

(now owned by Ovid Technologies, Inc.)

- Home page: http://www.ovid.com/site/products/tools/silverplatter/access_tools.jsp?top=2&mid=3&bottom=8&subsection=15
- Gained their reputation publishing CD-ROM based databases

5. STN

- Home page: <http://www.cas.org/stn.html>
- A product of the American Chemical Society <http://www.cas.org>
- Database list: <http://www.cas.org/ONLINE/DBSS/dbsslist.html>

6. Engineering Information, Inc.

(now owned by Elsevier)

- Home page: <http://www.ei.org/eicorp/eicorp>
- Products page:
<http://www.ei.org/eicorp/eicorp?menu=productsservicesmenu&display=productsservices>
 - Engineering Index
<http://www.ei.org/eicorp/eicorp?menu=engineeringmonthlymenu&display=engineeringmonthly> one of their best-known products and one of the top players in the abstracts-database market
 - Compendex
<http://www.ei.org/eicorp/eicorp?menu=compendexmenu&display=compendex> another well-known product

IV. About the Author

Brenda Bell has over 16 years' experience selecting source material and preparing records for publication in the Aerospace and High Technology Database (formerly The Aerospace Database), one of the databases published by Cambridge Scientific Abstracts. She has presented the talk, "Searching the Internet – How to Find What You're Looking For" at the Trenton Computer Festival in 2000, 2001, and 2002.

V. About Cambridge Scientific Abstracts

CSA is a privately-owned information company located in Bethesda, Maryland with offices in Hong Kong, France, and the UK. CSA has been publishing abstracts and indexes to scientific and technical research literature for over 30 years.

Information about CSA and its products may be found on the Web at <http://www.csa.com>. The sales representative for New York, New Jersey, and Pennsylvania is Chris Vander Groef. He can be reached by e-mail at cvandergroef@csa.com.