Chemical Information Retrieval
CAS & SciFinder
Introduction to Subject Searching

_Chemical Abstracts_, which is published monthly by the Chemical Abstracts Service (CAS), is a collection of abstracts of chemistry-related articles published all over the world. The CAS databases are extensive:

- Over 16 million bibliographic records from over 8,000 journals (from 1967)
- Over 22 million chemical substance records (from 1957)
- Almost 4 million chemical reactions from over 100 journals (from 1985)

SciFinder Scholar provides access to the CAS databases, which are updated daily.

_**SciFinder functions VERY differently from standard web search engines and database query processors.**_ Instead of using truncation and Boolean operators to assemble concepts in a search query, SciFinder uses a _Natural Language Query (NLQ)_ system to break your query into a set of discrete concepts and compare the concepts with its database indices.

Here is some guidance for using the SciFinder NLQ system:

- Identify the concepts for your topic
- Link these concepts with preposition or conjunction connectors
  - Examples: after, among, as, at, between, for, from, in, into, of, on, upon, with.
  - For example: nanotubes with boron doping as semiconductors
- Adjacent words not linked by a connector are assumed to be a phrase
  - You may use quotation marks to identify an exact phrase, but it is not necessary
- Keep your query simple
  - No more than three or four distinct concepts
- **Do not** truncate words. In most cases SciFinder truncates words automatically.
- **Do not** use the Boolean operators AND and OR. SciFinder will ignore them.
- Instead of using the OR operator, you can specify synonyms by enclosing them in parentheses
  - For example: “group theory” (symmetry) in chemistry
- SciFinder is case-insensitive
  - aids = AIDS
- SciFinder recognizes standard abbreviations
  - NMR = nuclear magnetic resonance
- SciFinder does not distribute modifiers
  - Use ovarian cancer or liver cancer; don’t use ovarian or liver cancer
- Slight differences in the way you phrase a query can retrieve markedly different result
  - You should try searching in several different ways

---

2 Scholar is the academic version
3 www.lib.utexas.edu/chem/scifinder_faq.html#cite
Combining Concepts in SciFinder Searches

In the first Library lesson, I discussed the strategy for a successful search. The strategy for SciFinder is the same, with one significant exception

1. Formulate the search question

Ask yourself, "What is my topic?" and "What is the purpose of my search?" Suppose for example that the purpose of the search is to find articles about semiconductors made by boron doping nanotubes.

2. Identify the important concepts

Break your topic up into important concepts and choose words associated with each concept that you are likely to find in the articles. In our example there are four:

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
<th>Concept 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>semiconductor</td>
<td>nanotube</td>
<td>boron</td>
<td>doping</td>
</tr>
</tbody>
</table>

3. Prepare an initial search query

The search query is the series of words you enter to tell the search application what you are looking for. To write an effective search query, we do NOT truncate words. SciFinder’s NLQ system is very good at figuring out different variations of words, so you do not need to use truncation with wildcard operators. It a

We also do NOT use Boolean operators. Their use can defeat some of the effectiveness of the NLQ system. Instead we connect concepts with prepositions or conjunctions, such as after, among, as, at, between, for, from, in, into, of, on, upon, with

The SciFinder query for the four concepts above might be the following:

semiconductors by doping with boron in nanotubes

Notice that we have connected the four concepts with prepositions. Although it may help you to use prepositions that make sense to connect the concepts, this is not necessary. SciFinder only uses the connecting prepositions to figure out what the concepts are. The linguistic meaning of the prepositions is not analyzed.

---

Even though the SciFinder NLQ system does not use the Boolean OR operator, it is possible to designate alternative words. In the above example, suppose we are interested in doping with either boron or phosphorus. We can designate an alternative word by putting it in parentheses.

semiconductors by doping with boron (phosphorus) in nanotubes

boron (phosphorus) is equivalent to boron OR phosphorus

*It is very important to connect concepts with prepositions or conjunctions.* Simply entering a string of concepts without connectors can get you in trouble. Here is an example of what I mean.

Suppose you want to find references on the effect of atomic size on forces. You might expect the word “radius” to be used instead of size. There are three concepts in this search:

- atomic
- size or radius
- forces

A proper SciFinder NLQ system query might be: **atomic with size (radius) in forces**

With the query above, I found 4817 references with the three concepts closely associated with each other.

This query is NOT written properly: **atomic size (radius) forces**

This query finds 7887 references, *many of them irrelevant to your search*. It has assumed that “atomic size” and “radius” are alternate terms. The concepts used in this search are:

- forces
- “atomic size” or radius
Exploring References

You already should have registered for SciFinder as directed in my email message. In a browser (IE, Firefox, or Safari), go to the URL http://scifinder.cas.org. Sign in with your username and password and then Accept the conditions to use the application.

Let’s start on the “Explore References” option, which you should reach after you accept the conditions. As an example, assume again that you are doing a research paper on Bisphenol-A’s (BPA’s) role as an endocrine disruptor in humans, children in particular. This topic has three associated concepts:

- BPA or Bisphenol-A
- endocrine
- disruptor
- humans (with children as an alternate term)

It is important that your initial query not be too complex, because that may limit the search too much. I generally prefer to start with at most three concepts. If you need to use more, you can always refine the search later. Let’s start with the first three in the list.

**BPA (Bisphenol-A) as disruptor of endocrine**

You will see that SciFinder returns a set of results, starting with the most specific. A few of the sets are shown below.
The top set of results looks the most promising. Each reference in the set contains all the terms in the query and they are closely associated with each other. Select the check box in this row and click on **Get references** at the bottom of the page.

## Use Titles and Abstracts as a Guide

This search returns over 2500 results and obviously needs refining. At this **point use the titles and abstracts as a guide**. Read information about some of the higher ranked references to get an idea of ways to tighten up the search. **The better you understand what you have found, the better will you be able to improve the search.**

If we look at the first couple of references, we see three things

- The first reference is in Japanese, which won’t be very helpful.
- The first one doesn’t involve the effect of BPA on humans, children in particular, which was the fourth concept in the list on page 2.
- The abstract of the second reference suggests that estrogen and estrogenic activity are involved. Other abstracts in the list also mention estrogen.

We can refine the search to take all three of these observations into account.
Refining a Search

Click on the Refine tab at the top of the right column.

We noticed that the language on at least one reference is Japanese. We can eliminate other languages by selecting English only.

This cuts the list of results down to around 2100, still a lot.

We also realized that several references that do not involve the effect of BPA on humans, children in particular. We can further refine the results by Research Topic, looking only for the effect on humans with children as a synonym.

We also noticed that estrogen appears to be a factor in the activity of BPA. We decide to include this in the refinement as well.

Remember to connect the concepts with a preposition

estrogen in humans (children)

This is a significant refinement, down to a little over 350 results.
We might also specify the types of documents we want to use, mainly journals and related types of documents.

This doesn’t have much effect on the search results.

At this point, we have done a fairly thorough job of narrowing the search down.

Removing Duplicates

SciFinder actually pulls its results from two databases: CAS and Medline. The latter database is the U.S. National Library of Medicine's database that contains over 19 million references to journal articles in life sciences with a concentration on biomedicine.

Because SciFinder is pulling from two databases it sometimes will pull the same reference from both databases, so a list of search results frequently will contain duplicates. To remove duplicates from your list, select Tools and then Remove Duplicates.
Reverting to an Earlier Point in a Search

You may find that a refinement is not fruitful and would like to return to where you were in the search before the refinement was made. You can do this by clicking on the step to which you wish to return in the breadcrumb trail of the search. For example, if you would like to return to the display of references before the “estrogen in humans (children)” refinement, you can click on this point in the breadcrumb trail at the top of the page.

After you have done this, use the breadcrumb trail to return to the final point in the search, after you removed the duplicates.

Saving an Answer Set

This has been a successful search. We have more results than we need, but it is relatively simple to go through the list and select the references in which we are interested. Select a few items from the list by checking the box to the left of the their title.

When you have selected the references you want to save, click on Save in the line above the list of references.

In the pop-up window choose Only selected answers, give the answer set a title (maybe BPA_1), and enter a short description.

When you are ready, click on OK.
If you click on **Saved Answer Sets** in the upper right corner of the window, you will see information displayed about the search.

![Saved Answer Sets](image)

**Doing an Alternate Search**

You also might want to do a new search, phrasing the query in a different way. To do a new search click on **Explore References**.

![Explore References](image)

For example you might do a more specific search for

**BPA (Bisphenol-A) on estrogen in children**

Limit the search to a preprint, report, journal, or review written in English.

This returns far fewer results than before. In this case, we might look at the second set of 38 references that contains all the search concepts, but not necessarily closely associated. Remove the duplicates in the list.

Select a few articles from this list and save then as a second answer set named BPA_2.
Combining Answer Sets

We now have two answers sets on the same subject. It would be useful to combine them. Click on Saved Answer Sets to view the sets you have saved.

In the list of saved answer sets, select both sets by checking them.

Click on the Combine Answer Sets icon, which becomes active when you have selected more than one set in the list.

You can choose the way you want to combine the answer sets.

Make your choice and click on Combine Answer Sets.
Changing the Display of a List

You have the option of sorting the list of references in various ways. The default way of sorting is by **Accession Number** is a number assigned by CAS to uniquely identify the reference. In general, a higher accession number corresponds to a newer reference. You also can sort by Publication Year and other fields. The “Citing References” field stores the number of other articles that reference the one under consideration.

You also have control over the amount of abstract information displayed for each reference.
Displaying a History of Your Searches

Your memory probably is better than mine, but in case it is not you may find it useful to go back and review what you have done. To see a history of your searches, click on History in the upper right corner of the screen. This will display the history of your searches during the current session.

Note that you also can print out a record of the searches done in Previous Sessions.
Another Search

Before you leave this lesson, do the search described below.

Quantum dots are tiny particles, or “nanoparticles,” of a semiconductor material, traditionally selenides or sulfides of metals like cadmium or zinc, that range from 2 to 10 nanometers in diameter (about the width of 50 atoms). These small particles emit visible light when excited; the wavelength of the light depends not on the material from which the quantum dot is made, but on its size. Quantum dots can therefore be “tuned” during production to emit any color of light desired. The smaller the dot, the closer it is to the blue end of the spectrum, and the larger the dot, the closer to the red end. Dots can even be tuned beyond visible light, into the infra-red or into the ultra-violet. ⁵

In the past few years, some researchers have been looking into the possibility of using quantum dots in medicine. Do the following search, one step at a time.

1. Search for the application of quantum dots in medicine
   a. Look for only English-language journals, preprints, reports, or reviews
   b. Remember to connect concepts with prepositions
2. In the search results, look at the list of references that include all the search concepts, but not necessarily closely associated.
3. Refine the search to the Research Topic of diagnosis, with diagnostic as a synonym
4. Refine the search gain to the treatment of cancer
5. Limit the references to those published between 2010 and 2102
6. Remove duplicates
7. Select all of the references in the list and save them as an answer set
8. Look at the history of your search
9. Think of another way you can rephrase this search
   a. Do the search again as you have rephrased it.

⁵ http://www.nanocotechnologies.com/content/AboutUs/AboutQuantumDots.aspx