SEARCHING THE "SUNDIAL MAIL ARCHIVE"
by Carl Sabanski / edited by Mac Oglesby

Daniel Roth began the "Sundial Mailing List" (SML) in February of 1996 to serve as a means for the members of the world sundialling community to share their thoughts and ideas. Over the years many thousands of messages have been sent to the SML by the ever-increasing membership. The majority of the messages are directly related to sundialling but there have also been many "off-topic" messages of interest to many of the members. Seeing the need to make these messages always available, Mac Oglesby was instrumental in having the existing messages and any future messages stored for future reference in the "Sundial Mail Archive" (SMA). Its URL is:

http://www.mail-archive.com/sundial@uni-koeln.de/

But with thousands of messages stored in the SMA how do you find the ones relevant to the topic you are interested in. Or perhaps you remember the message was from a particular person or it had been sent during a specific period of time. How do you find it?

When you enter the SMA you will find at the top of the page the following items:

- "Thread": This selection lists entries in the SMA by topic or "thread" and is the default list when you enter the SMA. "Threading" is discussed later under "How to Prepare an SML Message".
- "Date": This selection lists entries by date starting from the most recent message and going to earlier messages.
- "Search": This selection lists messages based on the entry in the search box. This technique is discussed in detail later.
- "[Earlier Messages]": This selection is available only when using "Thread" or "Date". When searching by "Thread" it will list messages previous to those presently on the screen. When searching by "Date" it will list messages with dates previous to those presently on the screen. This selection cannot be made if you are on the last page of a listing.
- "[Later Messages]": This selection is available only when using "Thread" or "Date". When searching by "Thread" it will list messages after those presently on the screen. When searching by "Date" it will list messages with dates later than those presently on the screen. This selection cannot be made if you are on the first page of a listing.

Using the "[Earlier Messages]" and "[Later Messages]" selectors will only allow you to view, at 200 per page, the first 3000 messages sorted by "Thread" or "Date". This may appear to be a lot of messages but nowhere near the total number stored in the SMA. And this number keeps increasing continuously. All other messages are available using "Search".

Using "Thread" or "Date" is not an easy way to search for a specific message or topic. To help you do this there is the SMA search engine. The search engine is a powerful tool and, if used effectively, is an efficient way to find what you are looking for. Searching is a word game and

is a lot of fun. You will always win unless what you are searching for does not exist. Winning does not necessarily mean finding hundreds of messages related to what you are searching for but perhaps the few specific ones that will really help you. If you learn the techniques of searching and use them, you will quickly find that the more relevant messages will appear on the first page and not buried where you may not even find them. To search effectively you must also practice.

SMA SEARCH TECHNIQUES

The first step is to identify keywords that describe the message(s) you want to find. For example:

- If you want to find messages that deal with horizontal and vertical sundial designs specifically for the Southern Hemisphere the keywords may be:

horizontal   vertical   sundial   design   Southern Hemisphere

You may wonder why "sundial" is included when this is a "Sundial Mail Archive". If you search for only the words "horizontal" and "vertical" you will find messages that do not even relate to these types of sundials, such as messages with "horizontal surface" or "vertical wall".

Think about your search words carefully and any synonyms or variant word forms the creator of a message may have used. You may not use all the words in a single search but they will provide options for subsequent searches if you do not find what you are looking for.

The next step is to determine what query to enter in the "Search" box. This is done following specific rules of engagement or "search syntax". This is a lot easier than it sounds! A query is broken up into "terms" and "operators".

<u>TERMS</u>

There are two types of terms.

Single Term: This is a single word such as "sundial" or "shadow". A search on a single word will return messages that contain that word.

Phrase: This is a group of words surrounded by double quotes such as "horizontal sundial" or "shadow plane sundial design". A search on a phrase will return messages that contain the entire phrase as it is written.

Terms can be entered in upper or lower case. They will be treated as the same by the search engine.

"Single Term" and "Phrase" searches provide a means of performing simple searches but may result in many messages being returned. If you wish to narrow the results to messages more relevant to the topic of interest the use of "operators" is necessary.

OPERATORS

The terms can be combined with Boolean or logic operators such as AND, OR, NOT and "-". It is important to remember that Boolean operators must be "ALL CAPS" or they will be treated as search words.

AND　　　The AND operator is the default operator. If there is no Boolean operator between two terms, the AND operator is used. (This is true, despite what the Lucene query parser syntax page says.) The AND operator will return messages that contain **all** of the terms. For example:

horizontal AND vertical　　　　　　　　　**horizontal vertical**

will both return messages that contain **all** terms; "horizontal" and "vertical".

horizontal AND vertical AND polar　　　**horizontal vertical polar**

will both return messages that contain **all** three terms; "horizontal" and "vertical" and "polar".

OR　　　The OR operator will return messages that contain **either** of the terms. For example:

horizontal OR vertical

will return messages that contain **either** "horizontal" or "vertical".

horizontal OR vertical OR polar

are the same query and will return messages that contain **either** "horizontal" or "vertical" or "polar".

NOT　　　The NOT operator excludes messages that contain the term after NOT. For example:

horizontal NOT horizontally

will return messages that contain "horizontal" but not "horizontally".

\-   The "-" (implied NOT) or prohibit operator excludes messages that contain the term after the "-" symbol. For example:

horizontal -horizontally

will return messages that contain "horizontal" but not "horizontally".

Using Boolean operators will greatly increase your ability to search and find messages that deal with your topic of interest. Practice using them alone or in various combinations. For example:

sundial AND design NOT construct

will return messages that contain both the terms "sundial" and "design" but not "construct".

There is a method to further increase the power of Boolean operators.

Grouping: Using parenthesis "( )" groups clauses to form sub queries. For example:

(dialling OR dialing) AND (guide OR template)

will ensure the return of messages that must contain either the term "dialling" or "dialing" and must also contain either the term "guide" or "template".

You can now enter your query in the "Search" box and select the "Search" button. It's just that easy!

When you perform your first search, at the top of the page you will find the message "1-10 of ### matches" where "###" will be the total number of matches for your search query. At the bottom of the page you can select from any of 10 pages. Also note that the messages are not listed in chronological order. Each message is given a score calculated using a complex formula based on a variety of factors. This score determines how a message is ranked and where it appears in the listing after a search using your query string is completed.

The SMA search engine will also allow you to search particular email fields.

FIELDS

The email fields that can be searched include "from", "date", "subject" and "message". If a field is not specified in the query then all four fields are searched by default. The field must be entered completely in lower case or no matches will be found.

from: If you know the user name of the person who sent a particular message you can search this field as follows:

from:"Dialling Diva"

date: If you know the precise or approximate date a particular message was sent you can search this field as follows: **date:yearmonthday**

date:19980121 will return all messages sent on January 21, 1998.

date:199801* will return all messages sent in January, 1998.

date:[19980121 TO 19980610]

will return all messages sent between January 21, 1998 and June 10, 1998 inclusive. Note that "TO" must be entered in upper case or no matches will be found.

It is possible to enter a range of dates that is too wide. If you receive a message stating that no matches could be found do not assume this to be true. Reduce the date range and try again.

subject: You can search the subject line for particular terms as follows:

subject:"sundial kits"

The subject line can be a very useful way to find messages dealing with particular topics. However, this will only happen if members of the SML provide good descriptors in the subject line that will allow this happen.

message: You can search the body of a message for particular terms as follows:

message:"sundial kits"

And it even gets better! You can combine fields and operators in the query. For example:

date:[19980121 TO 19980610] from:"Dialling Diva" AND "shadow plane"

will return messages sent by "Dialling Diva" between January 21, 1998 and June 10, 1998 and contain the term "shadow plane" anywhere in the message.

There are number of term modifiers that provide even a wider range of searching options. These are listed below but only one will be discussed in detail. Wildcard and range searches have been used in the "date" field examples.

- Wildcard Searches
- Fuzzy Searches

- Proximity Searches
- Range Searches
- Boosting a Term

Wildcard Searches: You can search a message using single character and multiple character wildcard searches as follows:

- To perform a single character wild card search use the "?" symbol. This search looks for terms that match with the single character replaced. For example:

  te?t

  will return messages with the terms "test" and "text".

- To perform a multiple character wild card search use the "*" symbol. This search looks for terms with zero or more characters in place of the "*". For example:

  sun*

  will return messages with the terms "sun", "sundial", "sundialling", "sunny", "sunshine", "Sunday", etc.

  Wildcard searches can also be used in the middle of a term. For example:

  sun*ing

  will return messages with the terms "sundialing" and "sundialling".

  Note that you cannot use a "*" or "?" symbol as the first character of a search.

HOW TO PREPARE AN SML MESSAGE

If you are a contributor to the SML there are steps you can take that will help someone looking for "your" message find it more easily.

When a message comes into the SMA it must be threaded. Whether a message is threaded to an existing topic or is the start of a new topic will depend largely on the contributor. If you are sending a message on an entirely new topic described in the subject line a new thread will be started. Anyone replying to this message will have their message threaded to the original. At times the original topic of a thread may evolve into a new topic. If you reply to the original message but change the subject line to reflect the new topic a new thread will not be started. The threading subsystem must decide whether to track "who replied to what" or to track the subject line. The SMA goes with the former, as it is more reliable. That is why you see two different topics threaded together.

Any time you want your message to start a new thread be sure to start a new message with a new subject line. This is the best way to ensure that your message does not get threaded to an existing topic.

Take the time to include a good subject line that is relevant to your topic and uses descriptive terms found in the body. These are the terms that someone will be using to find your message when your message is no longer in the top 3000. These are terms that the SMA search engine may use to rank your message and make it easier to find.

NOTES:

• The Mail Archive FAQ concerning searching the archives is at:

   http://www.mail-archive.com/faq.html#search

• The Lucene query parser syntax page is at:

   http://lucene.apache.org/java/docs/queryparsersyntax.html