MEDLINE®/PubMed® Baseline Repository (MBR) Query Tool User's Manual

Date Last Updated: Wednesday, May 30, 2012

The MEDLINE/PubMed Baseline Repository (MBR) Query Tool allows users to search any of the currently stored MEDLINE Baselines (2002 - 2012). Users may use any combination of MeSH® Headings (MH), MeSH Subheadings (SH), Supplemental Concepts/Chemicals (NM), and MH/SH combinations to build their query searches. Users may also specify a date range to help limit the search.

The results of the query search can be provided as a list of matching PMIDs, as the matching records/citations themselves. If the user chooses to receive the matching citations, they can select between XML MEDLINE or ASCII MEDLINE formats. The user may also elect to have any of the results randomized and split into Testing and Training subsets.

The ability to create subsets was designed to provide researchers a simple way to develop test collections for their work. Depending on the researcher's needs, we have the capability to create a single subset of the requested size, or we can automatically split the results into Testing and Training subsets based on the researcher's requested distribution. If requested, we can also randomize the ordering of the results before we create the subsets.

Using the Query Tool, a researcher could very easily develop a test collection based on the following scenario:

> Using the 2003 MEDLINE Baseline, I would like all citations in Full MEDLINE format that include the MeSH Heading "Liver" and the MeSH SubHeading "drug therapy" (not necessarily in "Liver/drug therapy" combination) that were completed (using Date Completed/DCOM date) between January 1, 2002 and August 31, 2002. I need to have a randomized test collection that includes Testing and Training subsets with 90% of the results placed in the Training subset and the remaining 10% in the Testing subset.

The results are not provided in real-time due to the time involved in searching the databases and building the results set(s). You submit your request and are then notified via email when we have completed processing your request.

Please Note: Results in "Full Citations" format require extra time to generate the results due to the increased amount of data being brought together. This should normally be 30 – 60 minutes.

The MBR Query Tool web site is available using the IP address you provided to NLM at the following URL:

`http://mbr.nlm.nih.gov`

The MBR Query Tool website is designed to be contextually active. The decisions you make while filling out the form determine what fields are presented. This allows you to focus on the fields that are specific to your needs and saves you time by not having to weed through fields that may or may not be relevant. For example, if you elect to receive the results as a list of PMIDs, you have no need to select the format of the resulting citations, so that question is not presented to you.

NOTE: *Look for the following symbol*❓. *If you select this symbol you can retrieve additional information about the field this symbol is associated with.*

Details

Available Baselines: 2002 - 2012.

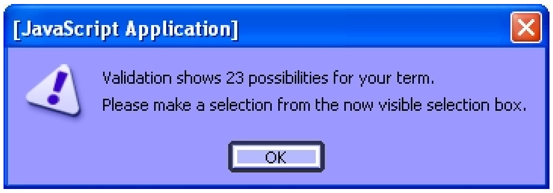
You may create a search using subject/chemical search terms, date limits, or a combination of the two:

Subject/Chemical Searches:

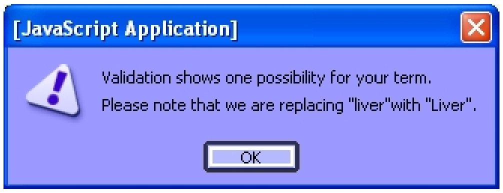
If you would like to perform a search based on any of the subject/chemical search types (MH, SH, MH/SH, or NM), just select "Yes" from the "***Perform Subject/Chemical Search(es)?***" selection list. You will then be allowed to enter up to five different Search Terms to search for in the database along with any associated conditionals. If you want to add an additional Search Term, select "Yes" when prompted with "***Another Subject/Chemical Search Term?***".

Query Search Types: MH, SH, MH/SH, and NM

Search Term:
To enter Search Terms, you can specify all or part of the word (case insensitive). ***When you leave the Search Term field, the contents will be validated against the live database.*** If there is more than one possibility for what you have specified, you will be provided with a selection list showing the various terms matching what you have entered to remove the ambiguity. You will be notified via an alert message as to how many possibilities there are available in the new selection list.

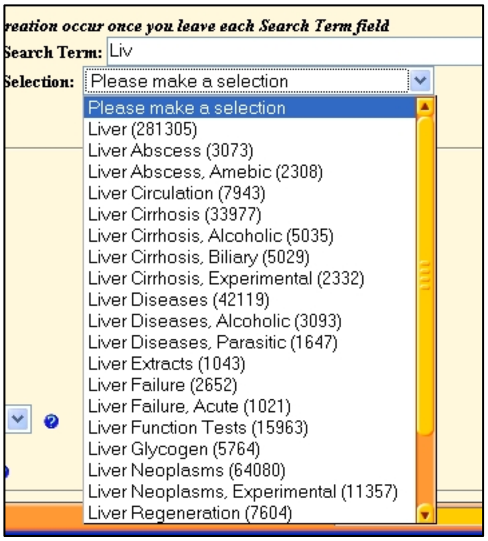

[JavaScript Application]

Validation shows 23 possibilities for your term.
Please make a selection from the now visible selection box.


[JavaScript Application]

Validation shows one possibility for your term.
Please note that we are replacing "liver"with "Liver".

The numbers in parentheses following each term in the selection list are the number of items available for that term in the database. The validation is not case sensitive. In the case of misspelled words, the validation process moves back along the term until it finds valid matching entries in the appropriate list. The figure to the right provides an example of how the Query Tool presents the list of valid entries matching the partial or misspelled term along with their count.


reation occur once you leave each Search Term field

Search Term: Liv

Selection: Please make a selection

Please make a selection
Liver (281305)
Liver Abscess (3073)
Liver Abscess, Amebic (2308)
Liver Circulation (7943)
Liver Cirrhosis (33977)
Liver Cirrhosis, Alcoholic (5035)
Liver Cirrhosis, Biliary (5029)
Liver Cirrhosis, Experimental (2332)
Liver Diseases (42119)
Liver Diseases, Alcoholic (3093)
Liver Diseases, Parasitic (1647)
Liver Extracts (1043)
Liver Failure (2652)
Liver Failure, Acute (1021)
Liver Function Tests (15963)
Liver Glycogen (5764)
Liver Neoplasms (64080)
Liver Neoplasms, Experimental (11357)
Liver Regeneration (7604)

Valid entries are of the form:

- **Full Term:** Liver (or liver) – *case insensitive*
- **Partial Term1:** Live -- *generates selection list of valid entries starting with -live-*
- **Partial Term2:** Live% -- *generates selection list of valid entries starting with -live-*
- **Misspelled Term:** Livre -- *generates selection list of valid entries starting with -liv-*

If you are looking for the MH "Liver" and the SH "drug therapy" – not necessarily tied together, you would need two search terms and specify MH of "Liver" and SH of "drug therapy". If you wanted to ensure the two were tied together, you would specify a single search term with "MH/SH" as your Query Type and with "Liver/drug therapy" as the Query Term.

Conditional: "AND", "OR", "AND NOT" are applied in a top down order for the searches.

For example, (**(**Search Term 1 <cond> Search Term 2**)** <cond> Search Term 3)

Date Limits:

If you wish to perform a search using Date Limits, simply specify "Yes" to the "Use Date Limitations?" prompt.

Date Field: Which of the available date fields to use for the search? Below is the list of available dates we support along with a brief description provided from the NLM Bibliographic Services Division web site (http://www.nlm.nih.gov/bsd/licensee/data_elements_doc.html).

Available Date Fields	Description
Date Created (DA)	The date processing of the record begins and an a record is distributed to PubMed and Licensees in In-Data-Review or In-Process status.
Date Completed (DCOM)	The date processing of the record ends; i.e., MeSH® Headings have been added to MEDLINE status records, quality assurance validations are completed; and the completed record subsequently is distributed to PubMed and licensees.
Date Last Revised (LR)	Identifies the date a change is made to a record either as a result of individual or global maintenance. There is no indication of what the change is on the record and only the latest revision date is distributed.
Publication Year	The Publication Year for the citation from the Publication Date field. This date field only accommodates YYYY format.

Start Date: Year, Month, Day
End Date: Year, Month, Day

Note: *These dates are inclusive. For example, if you specify:*
Start: 2001 January 15 End: 2001 January 31
The search would include records from the 15th and 31st.

MEDLINE Status filtering: You can filter the list of final results by removing specific Status categories if you wish.

Include MEDLINE Records? Can be either "Yes" or "No", "Yes" is selected by default. This includes all of the MEDLINE records that have been completed up to the time the baseline database was created. This is the status for the majority of citations in the Baselines.

Include Out-of-scope/PubMed-not-MEDLINE Records? Can be either "Yes" or "No", "Yes" is selected by default. Prior to the 2004 MEDLINE/PubMed Baseline, this status was called "Out-of-scope" and is now referred to by "PubMed-not-MEDLINE". The PubMed-not-MEDLINE status refers to less than 1% of the citations that reside in PubMed from journals included in MEDLINE and have undergone quality review but are not assigned MeSH headings because the cited item is not in scope for MEDLINE either by topic or by date of publication.

Include OLDMEDLINE Records? Can be either "Yes" or "No", "Yes" is selected by default. Starting with the 2005 MEDLINE/PubMed Baseline, OLDMEDLINE records/citations are included. The OLDMEDLINE citations are from international biomedical journals covering the fields of medicine, preclinical sciences, and allied health sciences. The OLDMEDLINE citations in the 2005 MEDLINE/PubMed Baseline were originally printed in hardcopy indexes published from 1951 through 1965 and make up approximately 11% of the total number of citations contained in the 2005 MEDLINE/PubMed Baseline.

Results: Can be a "list of PMIDs", "Citations in ASCII MEDLINE format", or "Citations in XML MEDLINE format". For information about the different Full Citations formats, please review the following URLs:

XML MEDLINE - http://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html

ASCII MEDLINE Display – list of elements in the ASCII MEDLINE Display format:
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helppubmed.table.pubmedhelp.T44

Please Note: If you elect to receive Full Citations in XML or ASCII format, you will be limited to a maximum of 100,000 citations in the results. Your query may produce more than 100,000 PMIDs in the PMID_list file, but, only the first 100,000 will be used to retrieve the Full Citations. If you end up splitting the results into Testing and Training sets, each file has the 100,000 limit separately – so you may end up with 200,000 citations.

# of Results: Can be either "All" of the results, or a "Subset" of the results.

If you request a subset, you will be asked for the size of the subset. The subset will be pulled from the head of the full list of results. **Note:** *If the total number of results is less than the size specified for the subset, we will use all of the results.*

Result Sets: You can specify a single result set or to have the results split between a Testing and Training subset.

If you specify the Testing/Training option, you will be asked for a percentage split that you would like the final result set allocated to each of the two subsets. **Note:** *The percentages must equal 100%.*

Randomize Results: You can request that the results be presented in a randomized order or in PMID numerical sorted order. If we randomize the results list, we do so **BEFORE** applying any subset size limitation (if specified) and **BEFORE** we split into Testing/Training subsets (if requested). If you have requested that we split the results into the Testing/Training subsets, once the entire list is randomized, we will select the Training subset items from the head of the list and the remainder will be selected for the Testing subset.

Before you can submit the form, the data must meet the following minimum criteria:

1) Must have selected a Baseline to use.
2) Must have an Email Address specified.
3) If you have elected to use Date Limitations, and have selected "Date Completed", "Date Created", or "Date Last Revised", the End Date (YearMonthDay) must be greater or equal to the Start Date (YearMonthDay). If you selected "Publication Year", the End Year must be greater or equal to the Start Year.
4) If you specified a Subject/Chemical search, the associated Search Term(s) must be filled in and validated along with any appropriate Conditionals selected.
5) If you did not specify a Subject/Chemical search, you must then specify a date limitation to use to limit the number of results.
6) If you specified that you wanted the results to be split between a Training and Testing subset, the percentages must total to 100.

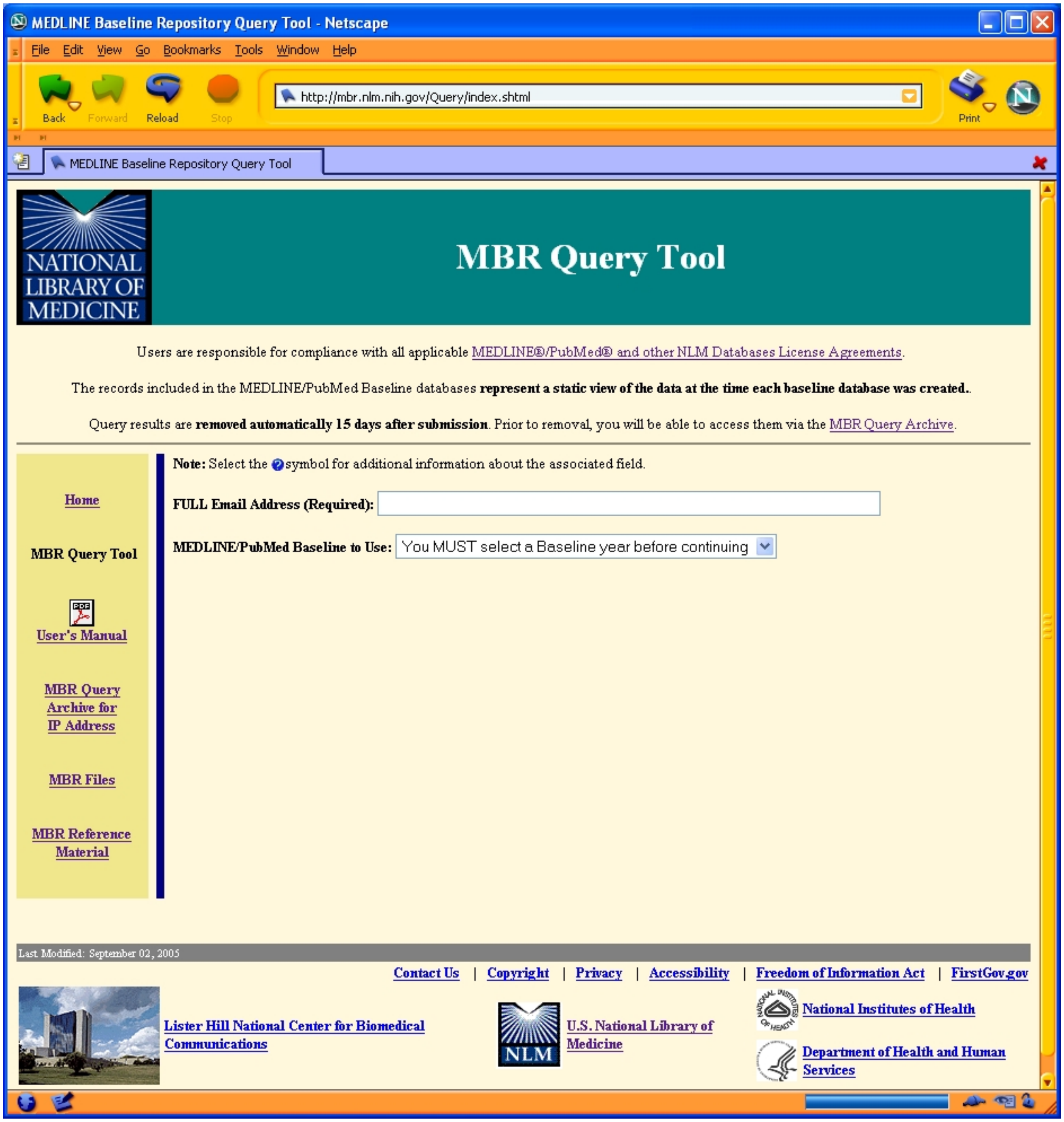
Once the form has been submitted, you will see a page that reiterates the data that was received by the program. The search you have specified will than be executed in the background. Once the request has finished processing, you will receive an email telling where to find the results. The email will provide you with a summary of the results of your search; provide a link to where you will find your results, and a list of what files should be in the results directory with a brief explanation for each file.

You may see the following files in your results directory depending on your selections:

File	Description
config	Original parameters from the web page
PMID_list	Results set list of PMIDs when Train/Test subsets NOT specified
Train_PMID_list	Results set list of PMIDs assigned to the Training subset
Test_PMID_list	Results set list of PMIDs assigned to the Testing subset
cits.txt	Single results set of citations (if ASCII MEDLINE specified)
cits.xml	Single results set of citations (if XML MEDLINE specified)
Train_cits.txt	Training subset results set of citations (if ASCII MEDLINE specified)
Test_cits.txt	Testing subset results set of citations (if ASCII MEDLINE specified)
Train_cits.xml	Training subset results set of citations (if XML MEDLINE specified)
Test_cits.xml	Testing subset results set of citations (if XML MEDLINE specified)
Summary	Summary of the run including number of entries found and echo of submission

Please Note: If you elect to receive Full Citations in XML or ASCII format, the result files will be compressed using bzip2. On Windows, you can use WinZip to uncompress the files. On Unix, you can use bunzip2 to uncompress the files.

Examples


MEDLINE Baseline Repository Query Tool - Netscape
File
Edit
View
Go
Bookmarks
Tools
Window
Help
Back
Forward
Reload
Stop
http://mbr.nlm.nih.gov/Query/index.shtml
Print
MEDLINE Baseline Repository Query Tool
NATIONAL LIBRARY OF MEDICINE
NLM

MBR Query Tool

Users are responsible for compliance with all applicable MEDLINE®/PubMed® and other NLM Databases License Agreements.

The records included in the MEDLINE/PubMed Baseline databases **represent a static view of the data at the time each baseline database was created.**.

Query results are **removed automatically 15 days after submission**. Prior to removal, you will be able to access them via the MBR Query Archive.

Home

MBR Query Tool

User's Manual

MBR Query Archive for IP Address

MBR Files

MBR Reference Material

Note: Select the ? symbol for additional information about the associated field.

FULL Email Address (Required):

MEDLINE/PubMed Baseline to Use: You MUST select a Baseline year before continuing

Last Modified: September 02, 2005

Contact Us | Copyright | Privacy | Accessibility | Freedom of Information Act | FirstGov.gov

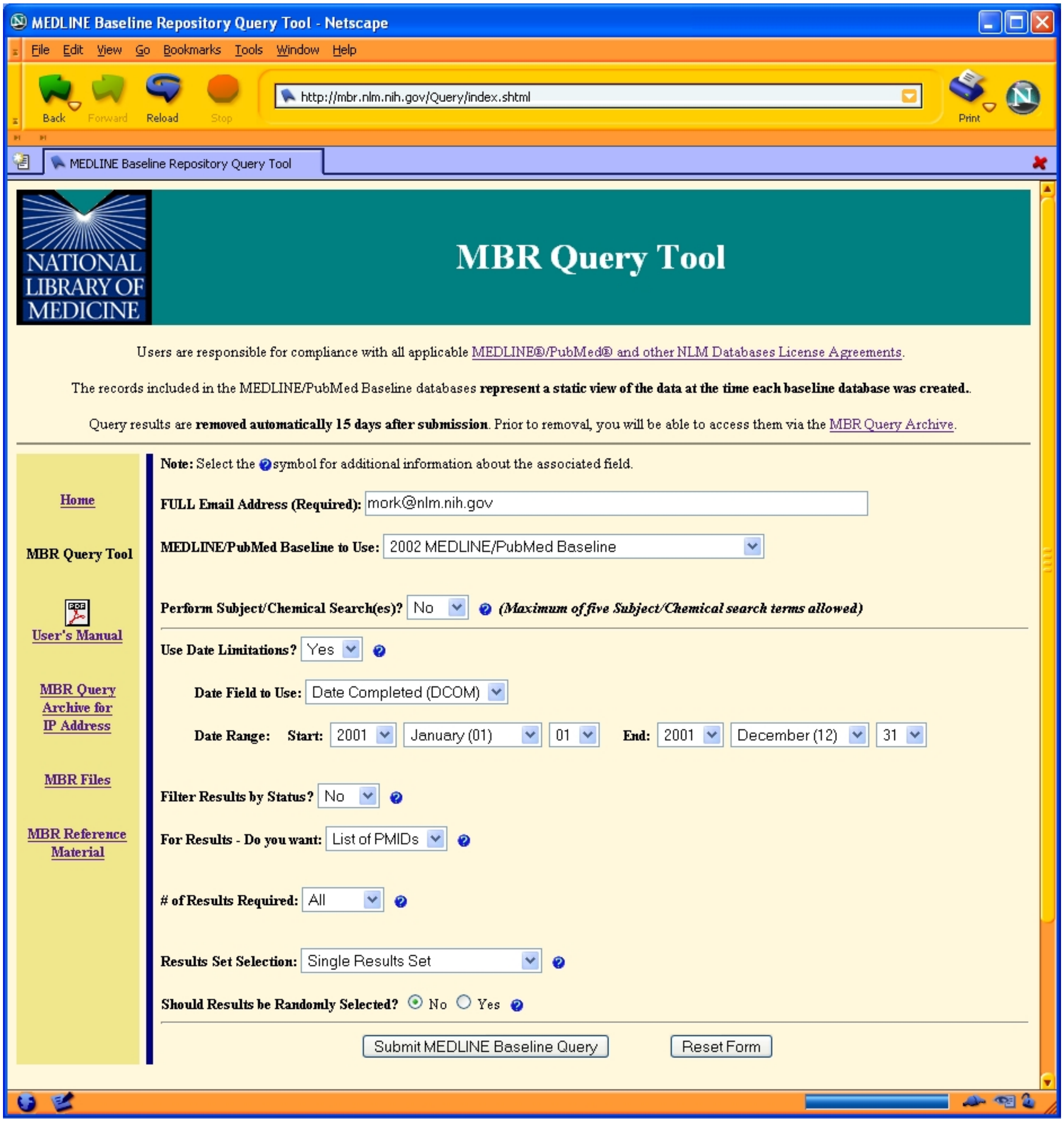
Lister Hill National Center for Biomedical Communications

U.S. National Library of Medicine

National Institutes of Health

Department of Health and Human Services

Initial screen showing that none of the fields appear until you make a selection for the MEDLINE/PubMed Baseline.


MEDLINE Baseline Repository Query Tool - Netscape
File Edit View Go Bookmarks Tools Window Help
Back
Forward
Reload
Stop
http://mbr.nlm.nih.gov/Query/index.shtml
Print
MEDLINE Baseline Repository Query Tool
NATIONAL LIBRARY OF MEDICINE
Home
MBR Query Tool
User's Manual
MBR Query Archive for IP Address
MBR Files
MBR Reference Material
Submit MEDLINE Baseline Query
Reset Form

MBR Query Tool

Users are responsible for compliance with all applicable MEDLINE®/PubMed® and other NLM Databases License Agreements.

The records included in the MEDLINE/PubMed Baseline databases **represent a static view of the data at the time each baseline database was created.**

Query results are **removed automatically 15 days after submission**. Prior to removal, you will be able to access them via the MBR Query Archive.

Note: Select the ? symbol for additional information about the associated field.

FULL Email Address (Required): mork@nlm.nih.gov

MEDLINE/PubMed Baseline to Use: 2002 MEDLINE/PubMed Baseline

Perform Subject/Chemical Search(es)? No ? ***(Maximum of five Subject/Chemical search terms allowed)***

Use Date Limitations? Yes ?

Date Field to Use: Date Completed (DCOM)

Date Range: **Start:** 2001 January (01) 01 **End:** 2001 December (12) 31

Filter Results by Status? No ?

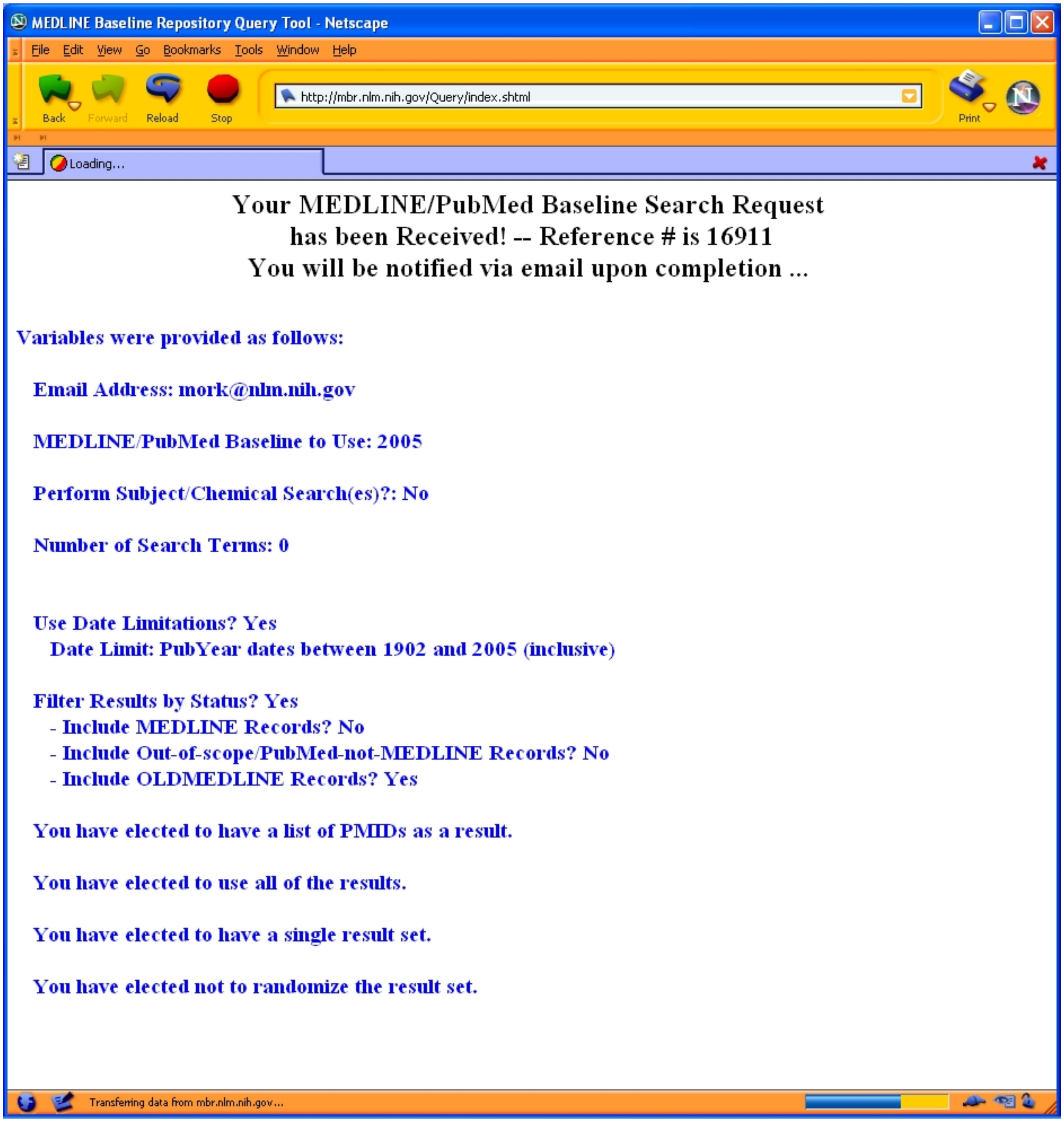
For Results - Do you want: List of PMIDs ?

# of Results Required: All ?

Results Set Selection: Single Results Set ?

Should Results be Randomly Selected? ● No ○ Yes ?

This example shows a partially expanded form where we have selected that our results be pulled from the 2002 MEDLINE/PubMed Baseline. We have requested a list of PMIDs which were completed between January 1, 2001 and December 31, 2001 inclusive. **NOTE:** *The 2002 MEDLINE/PubMed Baseline would not really have any records with completion dates after mid-November due to how and when the baselines are created.*


MEDLINE Baseline Repository Query Tool - Netscape
File Edit View Go Bookmarks Tools Window Help
Back
Forward
Reload
Stop
http://mbr.nlm.nih.gov/Query/index.shtml
Print
Loading...
Transferring data from mbr.nlm.nih.gov...

Your MEDLINE/PubMed Baseline Search Request
has been Received! -- Reference # is 16911
You will be notified via email upon completion ...

Variables were provided as follows:

Email Address: mork@nlm.nih.gov

MEDLINE/PubMed Baseline to Use: 2005

Perform Subject/Chemical Search(es)?: No

Number of Search Terms: 0

Use Date Limitations? Yes
Date Limit: PubYear dates between 1902 and 2005 (inclusive)

Filter Results by Status? Yes
- Include MEDLINE Records? No
- Include Out-of-scope/PubMed-not-MEDLINE Records? No
- Include OLDMEDLINE Records? Yes

You have elected to have a list of PMIDs as a result.

You have elected to use all of the results.

You have elected to have a single result set.

You have elected not to randomize the result set.

This is an example of the receipt you would receive upon submitting a search. The receipt lists all the relevant data from your request.

```
Subject: MBR Query - 16911 (1760574) Completion
From: <MBR_QueryTool@nls9.nlm.nih.gov>
To:

Your MEDLINE/PubMed Baseline Repository Search request has completed!

Refnum: 16911:
Accumulated Time to Process Search: 33 seconds
     -- 00 Hours
     -- 00 Minutes
     -- 33 Seconds

We found a total of 1760574 citations matching your criteria.


Please Remember:
  1) The records included in the MEDLINE/PubMed Baseline database
     REPRESENT A STATIC VIEW OF THE DATA AT THE TIME THE BASELINE
     WAS CREATED.
  2) These search results will be automatically removed in 15 days.

The results may be found at:
   http://mbr.nlm.nih.gov/baseline/16911/

File Information:
   config            -- Original parameters from web page
   PMID_list         -- PMID list from the Result Set
   Summary           -- Summary of the run
```

This is an example of the notification email you would receive when the MBR facility has finished building your test collection. If you follow the link specified, you will have access to the files only if you use the same IP address previously registered with NLM and used to create the original search. The subject line always denotes the MBR Query Tool batch number (16911) and a total count of results you received (1760574).

For additional information about the different elements within each of the Full Citations formats, please review the following URLs:

XML - http://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html

ASCII MEDLINE Display – list of elements in the ASCII MEDLINE Display format:
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helppubmed.table.pubmedhelp.T44

Examples of the different Full Citation results formats:

ASCII MEDLINE Display:

```
PMID- 15360816
OWN - NLM
STAT- MEDLINE
DA  - 20040913
DCOM- 20050607
PUBM- Print
IS  - 1569-6332
VI  - 11
IP  - Pt 1
DP  - 2004
TI  - The NLM Indexing Initiative's Medical Text Indexer.
PG  - 268-72
AB  - The Medical Text Indexer (MTI) is a program for producing MeSH indexing
      recommendations. It is the major product of NLM's Indexing Initiative and
      has been used in both semi-automated and fully automated indexing
      environments at the Library since mid 2002. We report here on an
      experiment conducted with MEDLINE indexers to evaluate MTI's performance
      and to generate ideas for its improvement as a tool for user-assisted
      indexing. We also discuss some filtering techniques developed to improve
      MTI's accuracy for use primarily in automatically producing the indexing
      for several abstracts collections.
AD  - National Library of Medicine, Building 38A, MS 54 8600 Rockville Pike,
      Bethesda, MD 20894, USA. alan@lhc.nlm.nih.gov
FAU - Aronson, Alan R
AU  - Aronson AR
FAU - Mork, James G
AU  - Mork JG
FAU - Gay, Clifford W
AU  - Gay CW
FAU - Humphrey, Susanne M
AU  - Humphrey SM
FAU - Rogers, Willie J
AU  - Rogers WJ
LA  - eng
PT  - Evaluation Studies
PT  - Journal Article
PL  - Netherlands
TA  - Medinfo
JID - 7600347
SB  - IM
MH  - Abstracting and Indexing/*methods
MH  - MEDLINE
MH  - *Medical Subject Headings
MH  - National Library of Medicine (U.S.)
MH  - *Natural Language Processing
MH  - Research Support, U.S. Gov't, P.H.S.
MH  - Unified Medical Language System
MH  - United States
EDAT- 2004/09/14 05:00
MHDA- 2005/06/09 09:00
AID - D040005122 [pii]
PST - ppublish
SO  - Medinfo 2004;11(Pt 1):268-72.
```

XML:

```
<PubmedArticle>
    <MedlineCitation Owner="NLM" Status="MEDLINE">
        <PMID>15360816</PMID>
        <DateCreated>
            <Year>2004</Year>
            <Month>09</Month>
            <Day>13</Day>
        </DateCreated>
        <DateCompleted>
            <Year>2005</Year>
            <Month>06</Month>
            <Day>07</Day>
        </DateCompleted>
        <Article PubModel="Print">
            <Journal>
                <ISSN>1569-6332</ISSN>
                <JournalIssue>
                    <Volume>11</Volume>
                    <Issue>Pt 1</Issue>
                    <PubDate>
                        <Year>2004</Year>
                    </PubDate>
                </JournalIssue>
            </Journal>
            <ArticleTitle>The NLM Indexing Initiative's Medical Text Indexer.</ArticleTitle>
            <Pagination>
                <MedlinePgn>268-72</MedlinePgn>
            </Pagination>
            <Abstract>
                <AbstractText>The Medical Text Indexer (MTI) is a program for producing MeSH
indexing
recommendations. It is the major product of NLM's Indexing Initiative and has been used in both
semi-automated
and fully automated indexing environments at the Library since mid 2002. We report here on an
experiment
conducted with MEDLINE indexers to evaluate MTI's performance and to generate ideas for its
improvement
as a tool for user-assisted indexing. We also discuss some filtering techniques developed to
improve
MTI's accuracy for use primarily in automatically producing the indexing for several abstracts
collections.</AbstractText>
            </Abstract>
            <Affiliation>National Library of Medicine, Building 38A, MS 54 8600 Rockville Pike,
Bethesda,
MD 20894, USA. alan@lhc.nlm.nih.gov</Affiliation>
            <AuthorList CompleteYN="Y">
                <Author ValidYN="Y">
                    <LastName>Aronson</LastName>
                    <ForeName>Alan R</ForeName>
                    <Initials>AR</Initials>
                </Author>
                <Author ValidYN="Y">
                    <LastName>Mork</LastName>
                    <ForeName>James G</ForeName>
                    <Initials>JG</Initials>
                </Author>
                <Author ValidYN="Y">
                    <LastName>Gay</LastName>
                    <ForeName>Clifford W</ForeName>
                    <Initials>CW</Initials>
                </Author>
                <Author ValidYN="Y">
                    <LastName>Humphrey</LastName>
                    <ForeName>Susanne M</ForeName>
                    <Initials>SM</Initials>
                </Author>
                <Author ValidYN="Y">
                    <LastName>Rogers</LastName>
                    <ForeName>Willie J</ForeName>
                    <Initials>WJ</Initials>
                </Author>
            </AuthorList>
            <Language>eng</Language>
            <PublicationTypeList>
                <PublicationType>Evaluation Studies</PublicationType>
                <PublicationType>Journal Article</PublicationType>
            </PublicationTypeList>
        </Article>
        <MedlineJournalInfo>
            <Country>Netherlands</Country>
            <MedlineTA>Medinfo</MedlineTA>
            <NlmUniqueID>7600347</NlmUniqueID>
        </MedlineJournalInfo>
        <CitationSubset>IM</CitationSubset>
        <MeshHeadingList>
            <MeshHeading>
                <DescriptorName MajorTopicYN="N">Abstracting and Indexing</DescriptorName>
                <QualifierName MajorTopicYN="Y">methods</QualifierName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="N">MEDLINE</DescriptorName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="Y">Medical Subject Headings</DescriptorName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="N">National Library of Medicine
(U.S.)</DescriptorName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="Y">Natural Language Processing</DescriptorName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="N">Research Support, U.S. Gov't,
P.H.S.</DescriptorName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="N">Unified Medical Language System</DescriptorName>
            </MeshHeading>
            <MeshHeading>
                <DescriptorName MajorTopicYN="N">United States</DescriptorName>
            </MeshHeading>
        </MeshHeadingList>
    </MedlineCitation>
    <PubmedData>
        <History>
            <PubMedPubDate PubStatus="pubmed">
                <Year>2004</Year>
                <Month>9</Month>
                <Day>14</Day>
                <Hour>5</Hour>
                <Minute>0</Minute>
            </PubMedPubDate>
            <PubMedPubDate PubStatus="medline">
                <Year>2005</Year>
                <Month>6</Month>
                <Day>9</Day>
                <Hour>9</Hour>
                <Minute>0</Minute>
            </PubMedPubDate>
        </History>
        <PublicationStatus>ppublish</PublicationStatus>
        <ArticleIdList>
            <ArticleId IdType="pubmed">15360816</ArticleId>
            <ArticleId IdType="pii">D040005122</ArticleId>
        </ArticleIdList>
    </PubmedData>
</PubmedArticle>
```