Windows MBox Viewer
User Manual 1.2

Table of Contents

1 Modification History

Revision	Date	Comments
1.0	February 28, 2018	Initial Release
1.1	March 12, 2018	Described new layout of the header in the Message Window. Described new global options settable via GUI.
1.2	April 9, 2018	Enhanced search scope to include from, to, subject fields, message text and text attachments. Eliminated false negative searches by decoding base64 and quoted-printable content blocks. Enhanced mboxview index file to keep track of text content blocks to optimize message and attachments search performance. Deprecated searching of RAW mail data.

2 Feedback

To help to improve mbox viewer, please post the review what works and what doesn’t, create bug

reports and request for enhancements. Provide as many details as possible, such as country, screenshots, etc.

3 Overview

MBox Viewer (or mboxview) is Windows GUI based program to view mbox files such as Thunderbird Archives, Google mail archives or simple Eml files.

The mbox viewer is a simple but quite powerful email viewer that supports the following features:

1. large file support > 4Gb
2. fast parsing of mbox archives
3. quick access to attachments
4. export of single mail in Eml
5. search by date, subject, sender or raw email
6. sort by date, from, to, subject and size
7. support for mail's body encoded with different character sets
8. support for searching emails sorted by header fields

4 Installation

The executable and the source code can be downloaded from Sourceforge or Github.

https://sourceforge.net/projects/mbox-viewer/files/

https://github.com/eneam/mboxviewer/releases

Download the executable package, for example the latest mbox-viewer.exe-v1.0.2.1.zip, to the working directory and unzip. This will create the mbox-viewer.exe-v1.0.2.1 directory and the stand-alone mboxview.exe executable in that directory. No administrator privileges are required.

5 Running the MBox viewer

The mboxview program can be executed by double clicking the executable from the explorer window or from the command line by typing the command name and argument list to it.

5.1 Argument List Summary

The mboxview accept the following command line options.

-FOLDER=folderPath – the full path to the folder containing one or more mbox or eml files. Persists in the registry across multiple runs.

-MAIL_FILE=fileName – full file path or just file name of mbox/eml file to load. If the full path is not specified, the full path is created from the folderPath and the fileName. Doesn't persist in the registry across multiple runs.

-EXPORT_EML=y|n – enables or disables automatic export of eml files generated from mbox

archive. Improves traversing performance of mail if set to ‘n”. Persists in the registry across multiple runs.

-PROGRESS_BAR_DELAY=seconds – search progress bar is activated when the search duration lasts longer than the value set for PROGRESS_BAR_DELAY. Applicable to searches of RAW mail. If set to **-1**, the progress bar is disables. Default value is set to **1** seconds. Persists in the registry across multiple runs.

If desired, instead of typing directly desired command line options, a simple .cmd or .bat file can be created for convenience and run from the explorer.

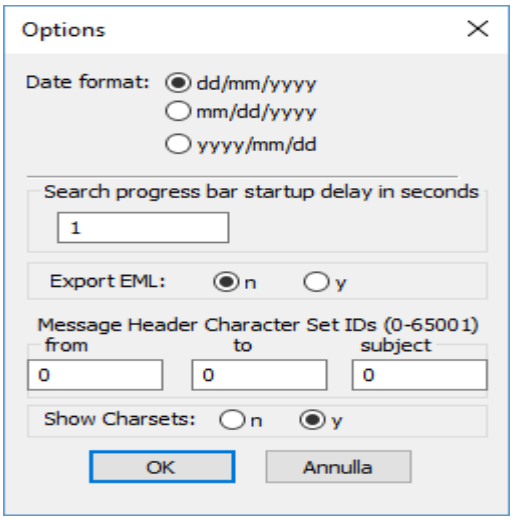
5.2 Setting Options from GUI

The **EXPORT_EML** and **PROGRESS_BAR_DELAY** values can be changed on the fly at any time from GUI.

Click on the “File” to open the drop menu and then select the “Options” option. The below dialog will be presented. Set the desired values followed by OK button.

Note that if exporting of EML files is disabled, you can still export the eml file for the selected mail by selecting “View EML” under the “View” drop menu.

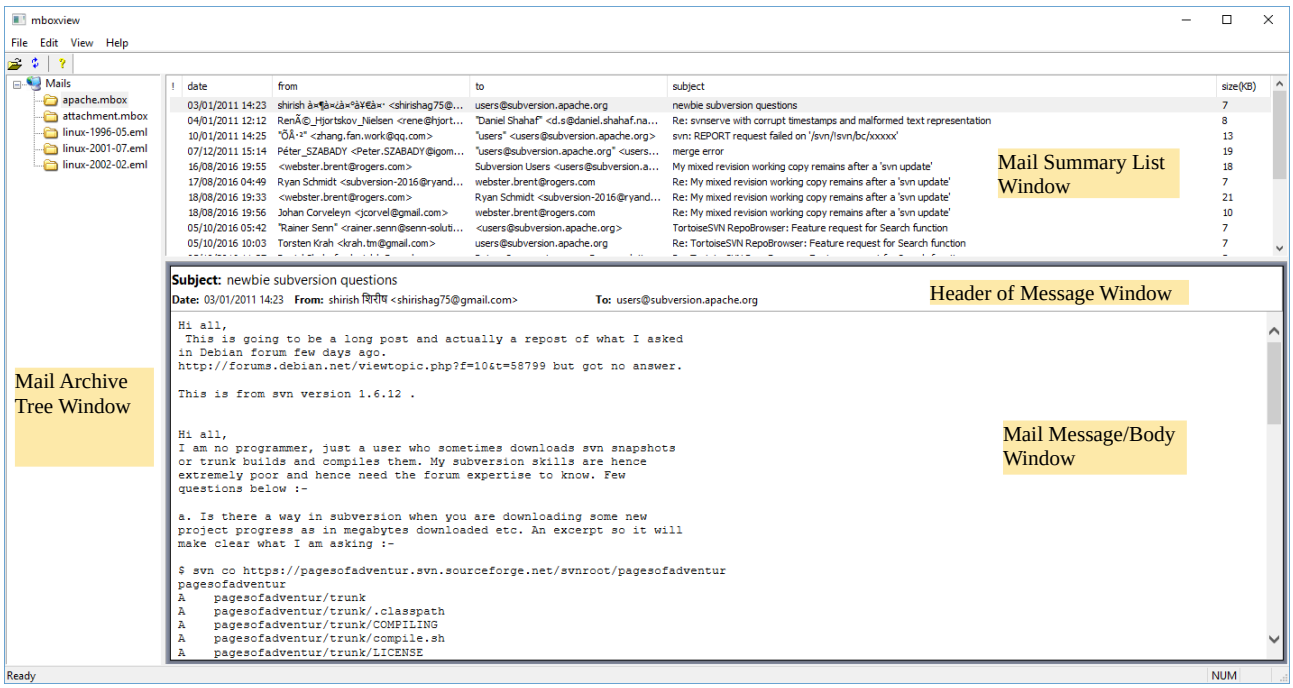
Two options, “**Show Charsets**” and “**Set Character Sets for Header Fields**”, can help to properly display header fields in the Message Window. See 5.4.1.1 for additional details.



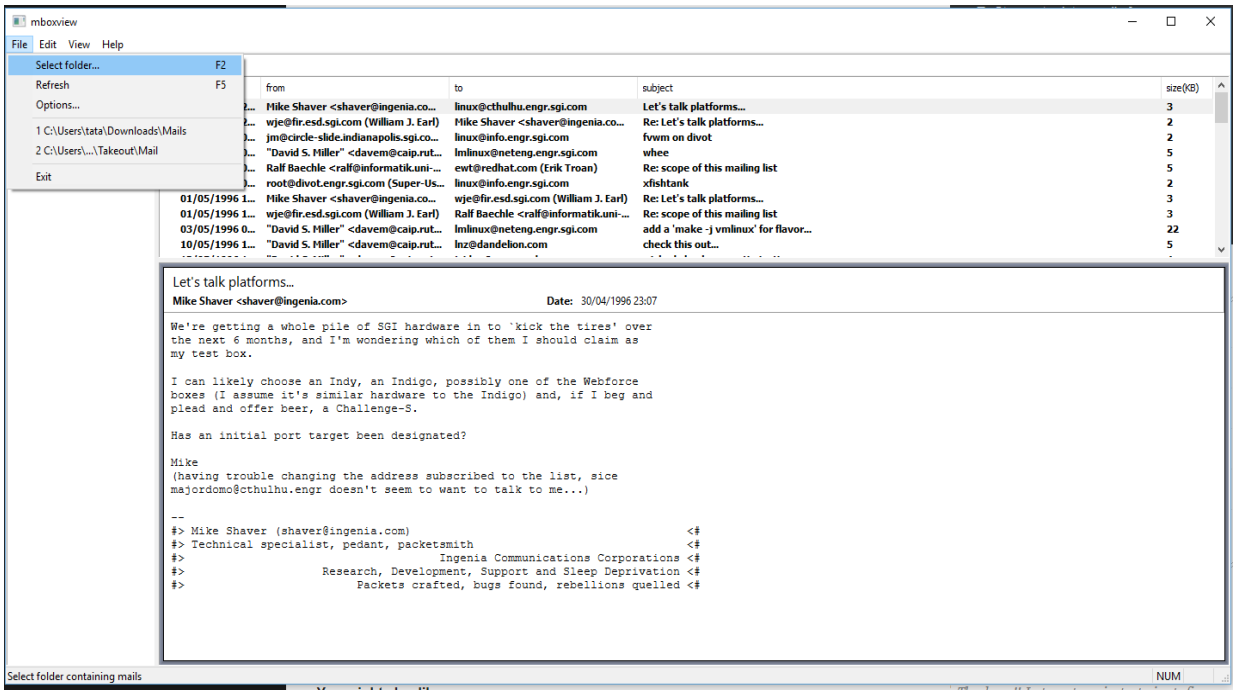
5.3 Basic Use Case

Double click of the mboxview to start the viewer. The picture below shows the screen created by mboxview after few steps performed by a user. Mail Archive, Summary and Message windows are initially blank after startup. Steps to load an archive and display the content are described next.

NOTE: The screenshot below shows the new header layout in the Message Window. Remaining screenshots were updated only when needed.



1. Click on the “File” to open the drop menu and then select the “Select Folder..” option. Browse to the folder containing one or more mbox and/or eml mail archive files and select it. All valid mail archive files will appear in the Mail Archive Tree window.

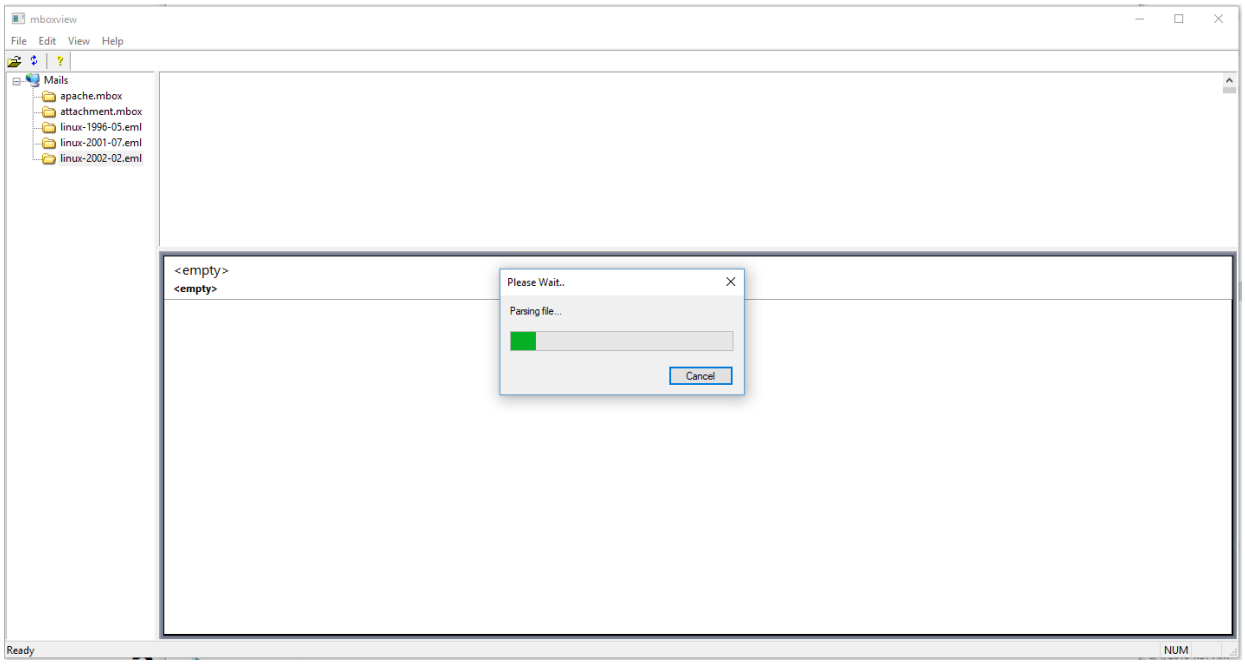


2. Click on one of the archive files to load all mails present in the archive. Progress bar will appear and automatically close after the selected archive is fully processed. Header information of each email will appear in the Summary window.

   Note that parsing of very large archive file may take some time since the mail archives are

text files and every character has to be examine one by one. However, subsequent loading of mails is done from the index file created by the mboxview during the initial parsing of the archive file. The created index file contains the header information and content meta data of each mail in the archive file and the position of each mail within the mail file for quick access to the mail message/body. Loading from the index file is much faster. The index files have the .mboxview suffix.

Note that all mails in the Summary window are marked in bold to indicate loading from the archive file while mails loaded from the index file are not marked in bold.



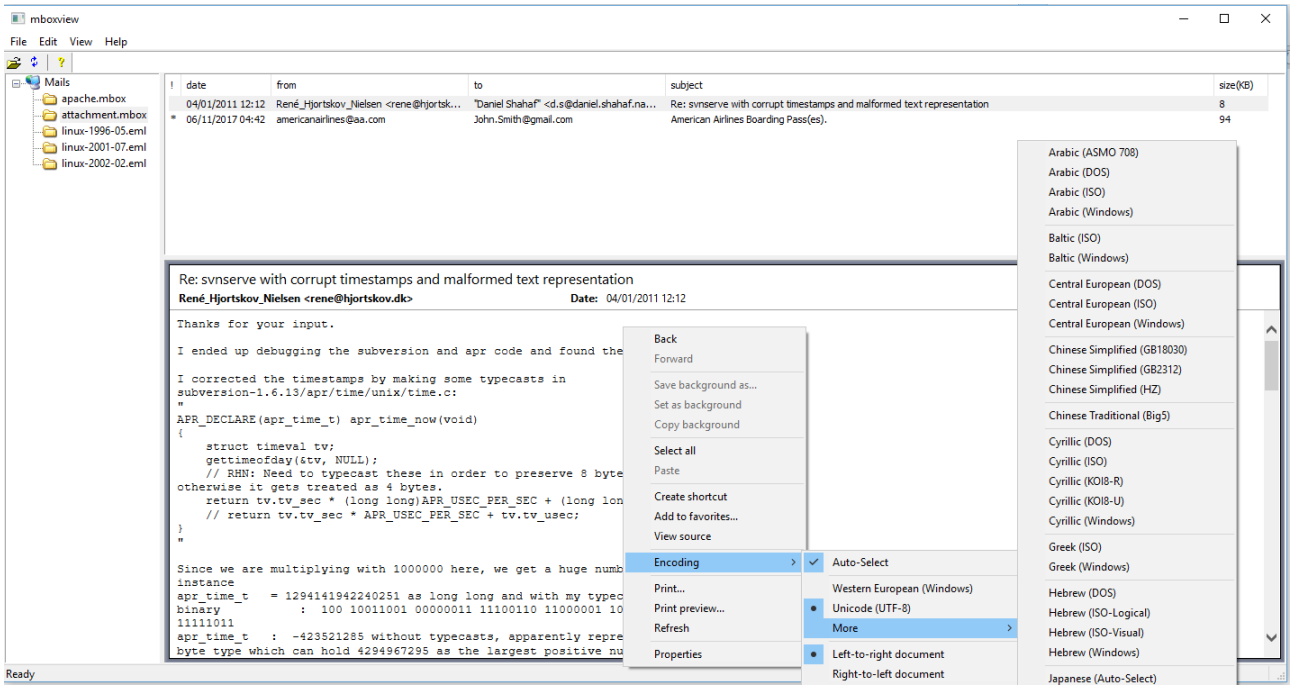
3. Click on one of the mails in the Summary window to show the Message/Body of that email in the Message window.

5.4 Language Support

5.4.1 Message Window

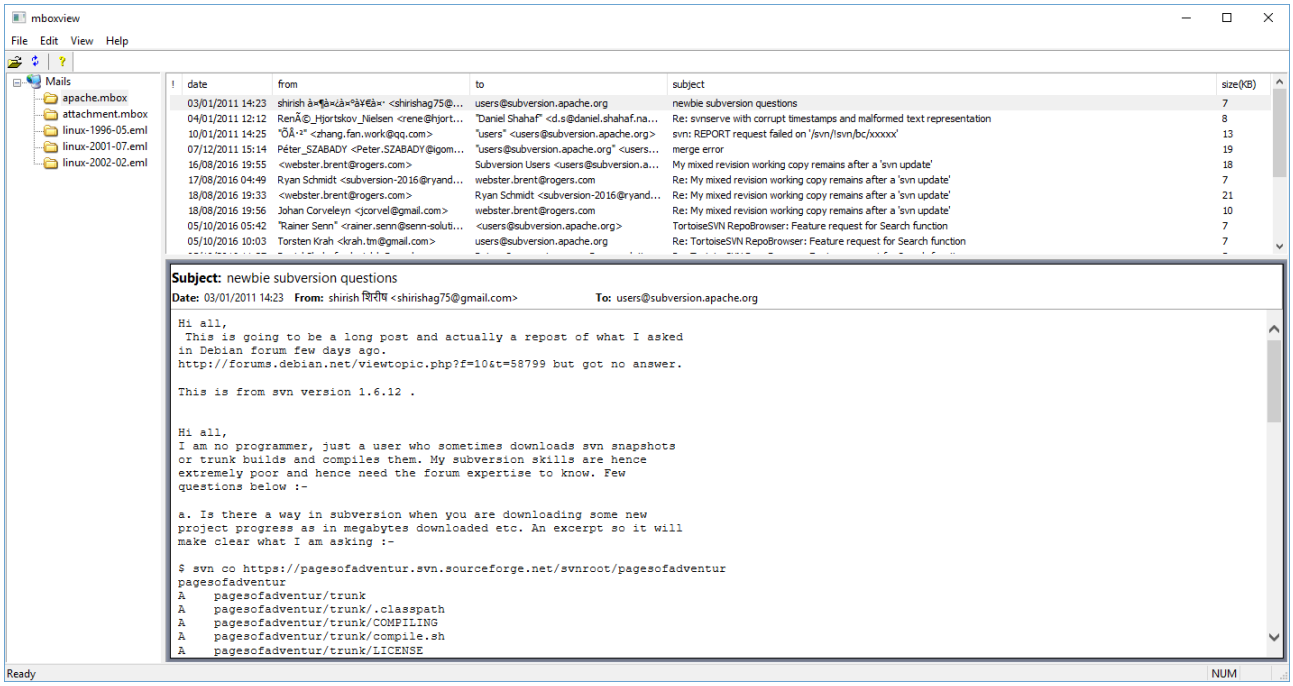
Different languages (or character sets) are quite well supported in the Message Window. This is due to the Microsoft Web Browser emulation/control class integrated into the mboxview. The mboxview generates the htm file from the mail's body and Web Browser displays the generated file in the Message window.

Display may not be correct if the character set for the mail's body is not defined in the mail. You can try to apply different encoding by right clicking within the window, selecting "encoding" and "more" options to apply different language.



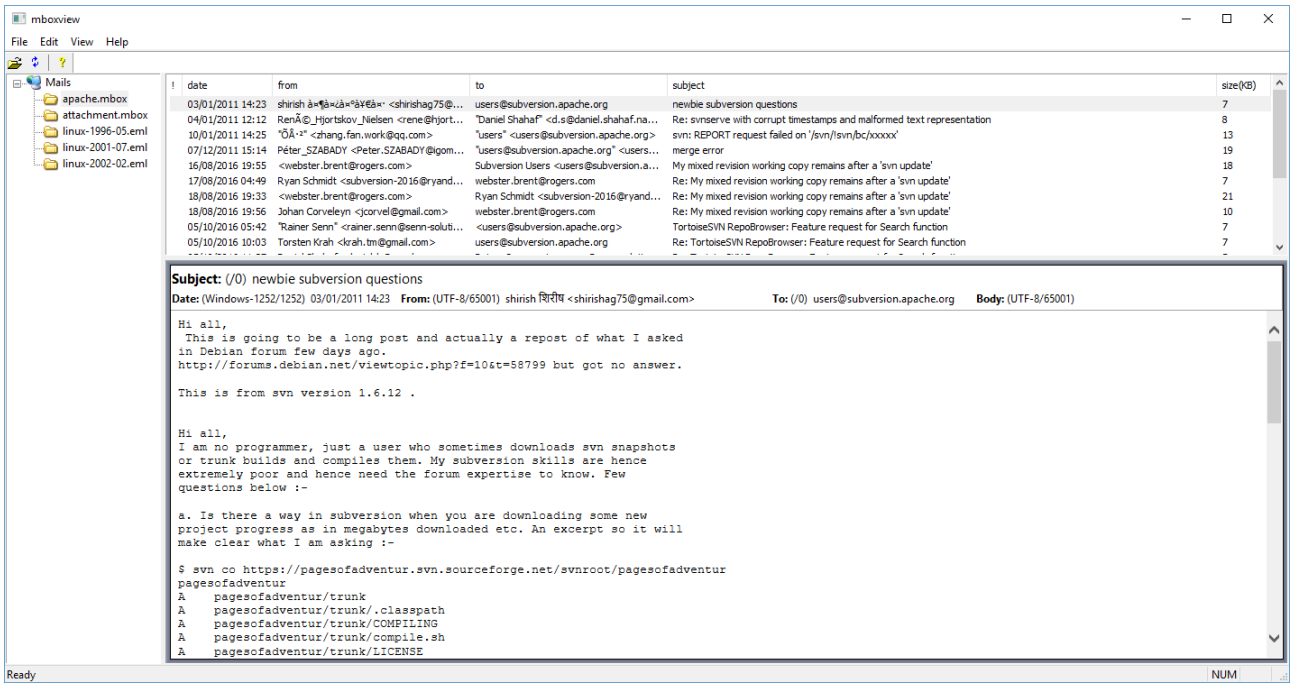
5.4.1.1 *Header of Message Window*

Header layout was changed in the 1.0.1.13 release. At the same time, implementation was enhanced to display mail header fields correctly as long as the corresponding field encoding type is present in the email. Note that the same “from” field below is presented correctly in Header of the Message Window but not in the Summary Window. This is because the non trivial effort might be required to support proper presentation of fields in the Summary Window. Looking to address this issue in the future.



Two global options can be used to address incorrect field presentation in the Message Header in case the field encoding is not present in the email. Missing field encoding type may or may not be a problem. It should not be a problem if the default/local character set installed on your computer matches the original encoding of the header fields. It might be an issue if you are receiving emails from different countries.

As shown below, you can see all encoding/character sets applied to the header fields by enabling “Show Charsets” option. To enable, select “Options” from the “File” drop menu and enable “Show Charsets”.



If the field text presentation is not correct, you can try to configure character set (or so called code page according to Microsoft terminology) for selected header fields and see if that helps. Per field code pages can be configured via Options dialog box. Note that the user provided encoding doesn't override encoding in the email, if present, which is denoted by the code page greater than zero. Code Page set to zero indicates missing or unknown character set.

If you change the code page value(s) in the Options Dialog box, you need to refresh the selected email by going up and down (or down and up).

All Windows supported code pages are listed in chapter 6.

The character set information appears after the field title and is formatted as “(character set name/code page)”. The ‘*” star character is appended after the code page if the user specified code is applied). Example “(UTF-8/65001*)”.

The character set information after the “Date:’ title shows the local/default character set installed on your computer.

5.4.2 Summary Window

Current Windows MFC window object to support Summary window has limited ability to support different character sets and therefore presentation of the to, from and subject fields is not always entirely correct, except perhaps for ASCII characters.

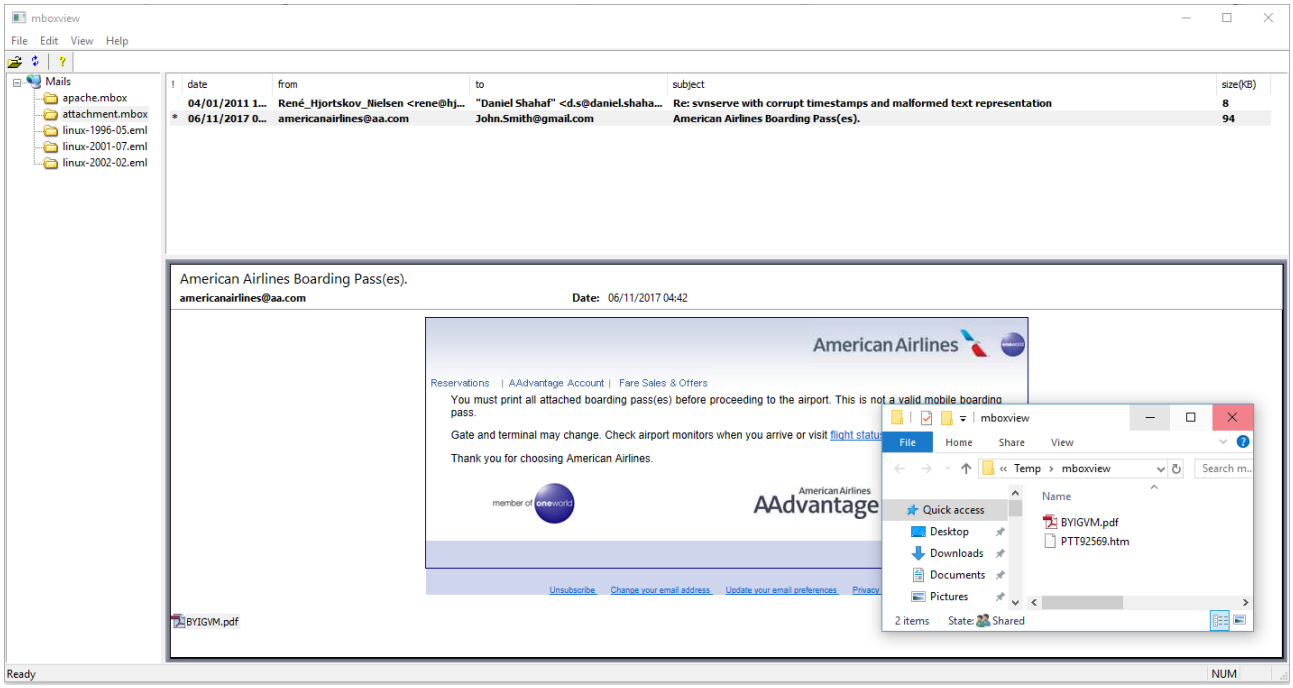
In 1.0.1.13 release, the field text is no longer re-encoded to the default Windows ANSI code page/character set which can differ on different computers. Instead, the field text is displayed as is/received. The approach should benefit searching and matching.

Note that once the mail is selected in the Summary Window, the header fields should be properly presented in the header of the Message Window.

Better support/implementation needs to be investigated but will likely require non trivial code changes.

5.5 Mail Attachments

Mail attachments are shown at the bottom of the Message window. Also, the “*’ character appears in the first column for the associated mail. To view attachments, click on the attachment icon or double click on the mail in the Summary window. Temporary folder will open containing all attachments for the email. Use standard Windows applications to view attachments.



5.6 Mail Navigation

Click on a specific mail in the Summary window and use **UP, DOWN** keys to move to the next or previous mail respectively.

Use **HOME** key to jump to the first mail. Use **END** key to jump to the last mail.

5.7 Mail Sorting

The Summary window is divided into 6 columns: attachment indicator, date, from, to, subject and mail size. Each column, except the first column, can be sorted by clicking on the column title. Subsequent click on the same column title will reverse the sorting rule.

5.8 Dynamic Column Resize

Columns in the Summary windows are automatically resized when the window is resized.

However, if columns are resized manually, some of the columns may no longer be present in visible area/window.

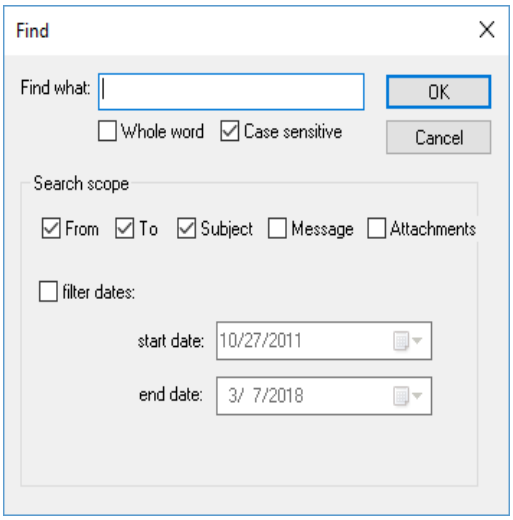
You may need to use the scroll bar to see missing columns, otherwise maximize and then restore the window to see all columns within the Summary window again.

5.9 Mail Search

The mboxview allows users to configure the search scope by selecting all or subset of the from, to, subject, message and text attachments options.

Click the “Edit” drop menu and select “Find” option to configure and start the search. Alternatively the CTRL+F will also present the search dialog.

The search dialog box allows user to specify the search string, whether to match the whole word, whether the search is case sensitive, configure the search scope and the start and end dates to consider for filtering.



To repeat the search in order to find the next matching mail, click the “Edit” drop menu and select “Find again” option, or select F3 key.

If “Find again” or F3 key is selected without prior “Find” or CTRL+F, the search dialog box will be presented to a user.

NOTE: Currently the “date” and “to” header fields are not consider for searching.

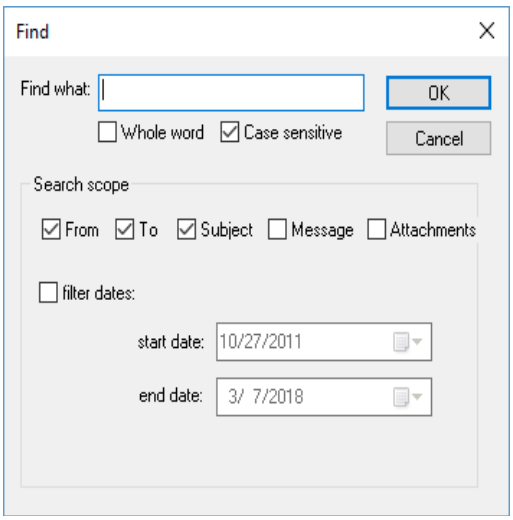
5.9.1 Search Reliability

Searching may not work as expected in all cases due to limited support of different character sets on input and in the mails.

Base64 or “quoted-printable” text content blocks and text attachments are decoded first before search to eliminate false negative results.

False positive results may happen when searching text/html content blocks since the html tags are not deleted from the content. The search string may match one of the html tags and result in false positive match.

If possible, use ASCII characters only to compose the search string.



5.9.2 Special Search Key

The single character '*', i.e. the star character can be specified as the search string to enable traversing the subset of mails between the start and end dates.

5.9.3 Search Progress Bar

Search progress bar displays the progress along the configured search string and enables user to cancel the search is so desired. The progress bar is not created unless a given search duration lasts longer than the delay time controlled by -PROGRESS_BAR_DELAY command line option or specified via GUI. Default value is 1 second.

5.9.4 Search Performance

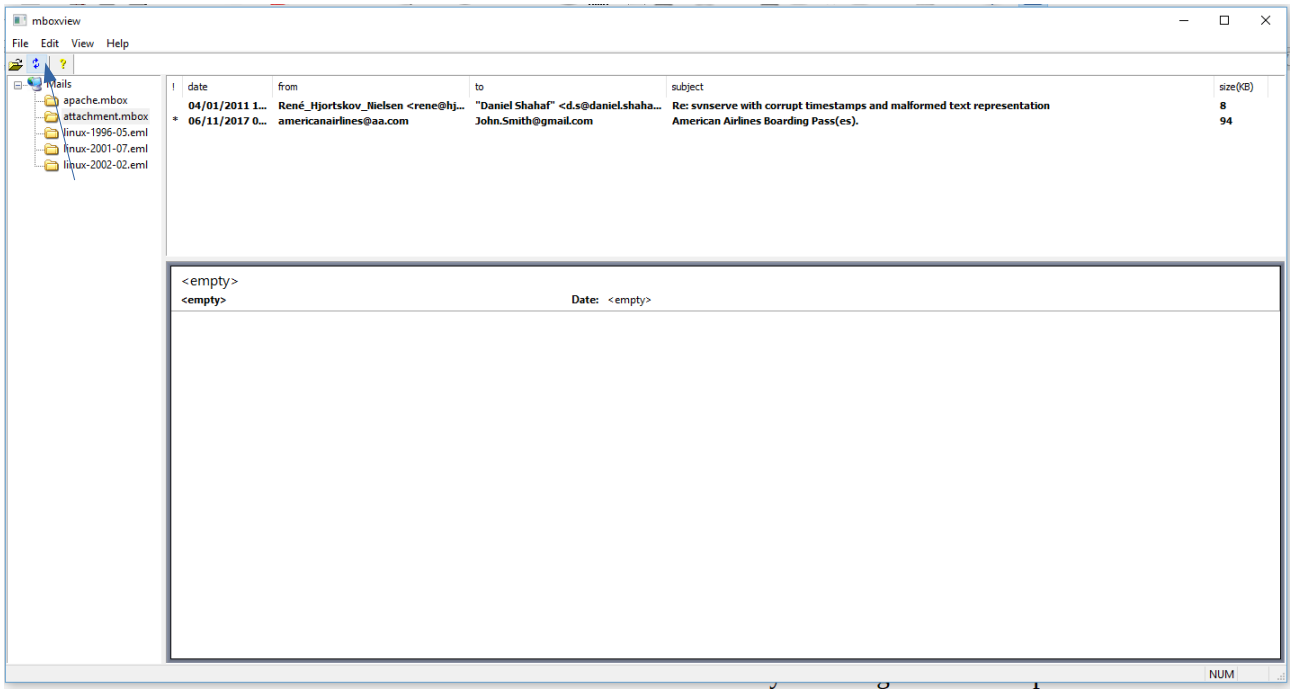
Searching of header fields performs best since all fields reside in the memory.

Searching of messages and text attachments may take some time for very large files, say larger than 10-20 Gbytes. Search requires to read the content blocks from the email file and decode the content when required. Worse case scenario is when the specified search string is not present in the mail archive. When activated, the search progress bar shows the configured search string and enables user to cancel the search is so desired.

Note that when searching of message and text attachments is configured, all mails will be sorted first as they appear in the mail file to maximize the search performance.

5.10 Refreshing Archive Tree Window

If an additional mbox or eml file is placed in the already selected folder, the Archive Tree window needs to be refreshed to see the new additions. This can be done by selecting "Refresh" option under the "File" drop menu or by clicking on the button showed below.



6 Windows Code Page Identifiers

The table below shows the supported character sets/code pages by Windows platform. The list is based on the following link.

https://msdn.microsoft.com/en-us/library/windows/desktop/dd317756(v=vs.85).aspx

CodePage	Name	Display Name
37	IBM037	IBM EBCDIC (US-Canada)
437	IBM437	OEM United States
500	IBM500	IBM EBCDIC (International)
708	ASMO-708	Arabic (ASMO 708)
720	DOS-720	Arabic (DOS)
737	ibm737	Greek (DOS)
775	ibm775	Baltic (DOS)
850	ibm850	Western European (DOS)
852	ibm852	Central European (DOS)
855	IBM855	OEM Cyrillic
857	ibm857	Turkish (DOS)
858	IBM00858	OEM Multilingual Latin I
860	IBM860	Portuguese (DOS)
861	ibm861	Icelandic (DOS)
862	DOS-862	Hebrew (DOS)
863	IBM863	French Canadian (DOS)
864	IBM864	Arabic (864)

CodePage	Name	Display Name
865	IBM865	Nordic (DOS)
866	cp866	Cyrillic (DOS)
869	ibm869	Greek , Modern (DOS)
870	IBM870	IBM EBCDIC (Multilingual Latin-2)
874	windows-874	Thai (Windows)
875	cp875	IBM EBCDIC (Greek Modern)
932	shift_jis	Japanese (Shift-JIS)
936	gb2312	Chinese Simplified (GB2312)
949	ks_c_5601-1987	Korean
950	big5	Chinese Traditional (Big5)
1026	IBM1026	IBM EBCDIC (Turkish Latin-5)
1047	IBM01047	IBM Latin-1
1140	IBM01140	IBM EBCDIC (US-Canada-Euro)
1141	IBM01141	IBM EBCDIC (Germany-Euro)
1142	IBM01142	IBM EBCDIC (Denmark-Norway-Euro)
1143	IBM01143	IBM EBCDIC (Finland-Sweden-Euro)
1144	IBM01144	IBM EBCDIC (Italy-Euro)
1145	IBM01145	IBM EBCDIC (Spain-Euro)
1146	IBM01146	IBM EBCDIC (UK-Euro)
1147	IBM01147	IBM EBCDIC (France-Euro)
1148	IBM01148	IBM EBCDIC (International-Euro)
1149	IBM01149	IBM EBCDIC (Icelandic-Euro)
1200	utf-16	Unicode
1201	unicodeFFFE	Unicode (Big-Endian)
1250	windows-1250	Central European (Windows)
1251	windows-1251	Cyrillic (Windows)
1252	Windows-1252	Western European (Windows)
1253	windows-1253	Greek (Windows)
1254	windows-1254	Turkish (Windows)
1255	windows-1255	Hebrew (Windows)
1256	windows-1256	Arabic (Windows)
1257	windows-1257	Baltic (Windows)
1258	windows-1258	Vietnamese (Windows)
1361	Johab	Korean (Johab)
10000	macintosh	Western European (Mac)
10001	x-mac-japanese	Japanese (Mac)
10002	x-mac-chinesetrad	Chinese Traditional (Mac)
10003	x-mac-korean	Korean (Mac)
10004	x-mac-arabic	Arabic (Mac)

CodePage	Name	Display Name
10005	x-mac-hebrew	Hebrew (Mac)
10006	x-mac-greek	Greek (Mac)
10007	x-mac-cyrillic	Cyrillic (Mac)
10008	x-mac-chinesesimp	Chinese Simplified (Mac)
10010	x-mac-romanian	Romanian (Mac)
10017	x-mac-ukrainian	Ukrainian (Mac)
10021	x-mac-thai	Thai (Mac)
10029	x-mac-ce	Central European (Mac)
10079	x-mac-icelandic	Icelandic (Mac)
10081	x-mac-turkish	Turkish (Mac)
10082	x-mac-croatian	Croatian (Mac)
12000	utf-32	Unicode (UTF-32)
12001	utf-32BE	Unicode (UTF-32 Big-Endian)
20000	x-Chinese-CNS	Chinese Traditional (CNS)
20001	x-cp20001	TCA Taiwan
20002	x-Chinese-Eten	Chinese Traditional (Eten)
20003	x-cp20003	IBM5550 Taiwan
20004	x-cp20004	TeleText Taiwan
20005	x-cp20005	Wang Taiwan
20105	x-IA5	Western European (IA5)
20106	x-IA5-German	German (IA5)
20107	x-IA5-Swedish	Swedish (IA5)
20108	x-IA5-Norwegian	Norwegian (IA5)
20127	us-ascii	US-ASCII
20261	x-cp20261	T.61
20269	x-cp20269	ISO-6937
20273	IBM273	IBM EBCDIC (Germany)
20277	IBM277	IBM EBCDIC (Denmark-Norway)
20278	IBM278	IBM EBCDIC (Finland-Sweden)
20280	IBM280	IBM EBCDIC (Italy)
20284	IBM284	IBM EBCDIC (Spain)
20285	IBM285	IBM EBCDIC (UK)
20290	IBM290	IBM EBCDIC (Japanese katakana)
20297	IBM297	IBM EBCDIC (France)
20420	IBM420	IBM EBCDIC (Arabic)
20423	IBM423	IBM EBCDIC (Greek)
20424	IBM424	IBM EBCDIC (Hebrew)
20833	x-EBCDIC-KoreanExtended	IBM EBCDIC (Korean Extended)
20838	IBM-Thai	IBM EBCDIC (Thai)

CodePage	Name	Display Name
20866	koi8-r	Cyrillic (KOI8-R)
20871	IBM871	IBM EBCDIC (Icelandic)
20880	IBM880	IBM EBCDIC (Cyrillic Russian)
20905	IBM905	IBM EBCDIC (Turkish)
20924	IBM00924	IBM Latin-1
20932	EUC-JP	Japanese (JIS 0208-1990 and 0212-1990)
20936	x-cp20936	Chinese Simplified (GB2312-80)
20949	x-cp20949	Korean Wansung
21025	cp1025	IBM EBCDIC (Cyrillic Serbian-Bulgarian)
21866	koi8-u	Cyrillic (KOI8-U)
28591	iso-8859-1	Western European (ISO)
28592	iso-8859-2	Central European (ISO)
28593	iso-8859-3	Latin 3 (ISO)
28594	iso-8859-4	Baltic (ISO)
28595	iso-8859-5	Cyrillic (ISO)
28596	iso-8859-6	Arabic (ISO)
28597	iso-8859-7	Greek (ISO)
28598	iso-8859-8	Hebrew (ISO-Visual)
28599	iso-8859-9	Turkish (ISO)
28603	iso-8859-13	Estonian (ISO)
28605	iso-8859-15	Latin 9 (ISO)
29001	x-Europa	Europa
38598	iso-8859-8-i	Hebrew (ISO-Logical)
50220	iso-2022-jp	Japanese (JIS)
50221	csISO2022JP	Japanese (JIS-Allow 1 byte Kana)
50222	iso-2022-jp	Japanese (JIS-Allow 1 byte Kana - SO/SI)
50225	iso-2022-kr	Korean (ISO)
50227	x-cp50227	Chinese Simplified (ISO-2022)
51932	euc-jp	Japanese (EUC)
51936	EUC-CN	Chinese Simplified (EUC)
51949	euc-kr	Korean (EUC)
52936	hz-gb-2312	Chinese Simplified (HZ)
54936	GB18030	Chinese Simplified (GB18030)
57002	x-iscii-de	ISCII Devanagari
57003	x-iscii-be	ISCII Bengali
57004	x-iscii-ta	ISCII Tamil
57005	x-iscii-te	ISCII Telugu
57006	x-iscii-as	ISCII Assamese
57007	x-iscii-or	ISCII Oriya

CodePage	Name	Display Name
57008	x-iscii-ka	ISCII Kannada
57009	x-iscii-ma	ISCII Malayalam
57010	x-iscii-gu	ISCII Gujarati
57011	x-iscii-pa	ISCII Punjabi
65000	utf-7	Unicode (UTF-7)
65001	utf-8	Unicode (UTF-8)