User Guide for Log Analysis Tool and Engineering [LATE]

Author: Kiran Puranik (GkiranP@gmail.com)

Version: 1.0

Date: 10th Feb 2016

License Statements: Log Analysis Tool and Engineering [LATE] is an experimental tool designed by keeping in mind it shall provide positive benefits to its user. This tool is not tested properly and hence, surely will contain unattended bugs. Author will not be responsible for any harmful action happened by installing or using this tool or its any of the components in any of the user machine. So, please use this tool by your own risk.

This tool is designed using Qt open-source framework, version number 5.4.2. All the license terms and conditions applicable to Qt framework will be applicable to this tool also. Hence, author will automatically assume that, user has agreed for all the license terms and conditions of Qt framework will be applicable to him/her by installing or using this tool by any means. To read more about Qt framework and its licensing terms and conditions, please visit http://www.qt.io/licensing/.

This tool has also used freely available icons from https://www.iconfinder.com website. Hence author will assume that user has agreed for all the license terms and conditions applicable for Iconfinder.com, will be automatically applicable to him/her who installs this tool or uses this tool by any means.

If you are not ready to agree upon any of the above mentioned license terms and conditions, please do not use this tool or any components bundled with this tool.

Index

1 About LATE

LATE stands for Log analysis Tool and Engineering. The tool is designed to help its user in proper formatting and filtering any kind of text based logs generated by systems. This tool is still in its experimental stage, and not tested properly. Hence it is advisable to user that they can use this tool by their own risk.

2 Usage Guide

LATE has beautiful tool box at the top, which provides wide range of formatting and filtering options to its user. Below figure 2.1 shows tool bar displayed on LATE.

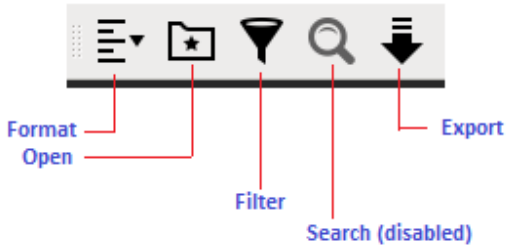

Fig 2.1

There are 5 tools operational in LATE,

Format – used to format and save large set of log files into small section of interested file contents.

Open – allows user to open log files into workspace.

Filter – used to Highlight or Delete set of regular expression based contents loaded into workspace of LATE.

Search – option is disabled. However, in future, I have plans to provide pattern search utility with this option.

Export – allows user to export contents of workspace into HTML page.

Let's have a look into each tool in detail, in following sections.

2.1 Format

Formatting option allows user to specifically format huge set of log files and its contents into small set of interested contents by extensively providing string pattern search and save method. Usage of format option will be very much helpful, when user has huge set of distributed log files and they have few common contents, which need to be extracted and saved into single file for operations to carry out.

Below figure 2.1.1 shows, how Format option works,

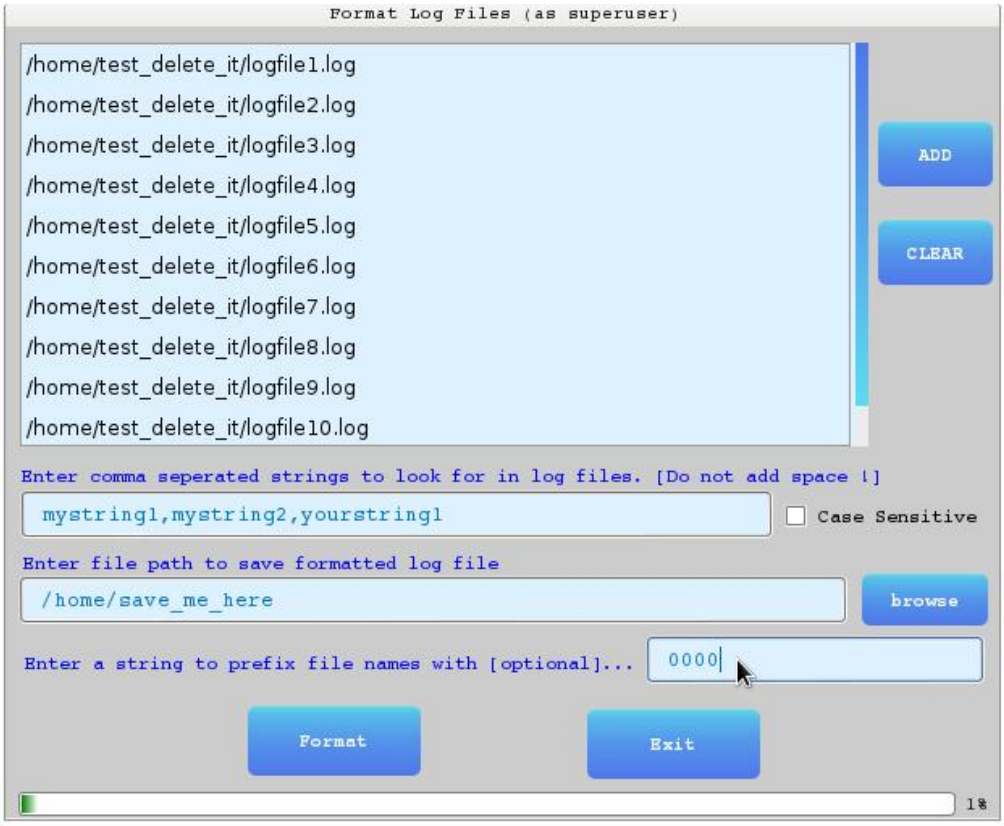

Fig 2.1.1

User has below controls to operate on Format window,

- Press “ADD” button to add list of raw log files from single directory at a time. If user needs to add files from multiple directories, he/she has to ADD them multiple times.

- Press “CLEAR” button to clear all the files from list of added files. Please be careful to operate this button, because user will not be prompted with any message before removing files from list. However, files will be just removed from format window list, while actual files will not be altered or modified by any means.

- Enter interested strings, each comma separated, into the text section provided. The strings user has entered will be searched into files contents and if matching is found they will be stored into corresponding formatted files. Please keep in mind, as many numbers of files will be created as many strings are added in this text section. It means if user has entered 3 search strings, for example, then 3 new files will be created into user specified directory containing matching contents from all the above listed log files. The newly created files will be having file names as that of strings; for example in above figure 2.1.1, user has entered 3 strings namely mystring1, mystring2 and yourstring1; and hence tool will look for all three string patterns into the log files listed above one by one and if the pattern matches, then 3 new files will be created by name mystring1.log, mystring2.log and yourstring1.log containing log information related to these 3 strings. Operation happening here is, just string comparison and there are no

effects of regular expression will be carried out. Please also note the warning and do not add spaces between strings after comma.

- Enter or browse for the path location to store all newly generated log files to store into.

- User has optional parameter to pass for file names in a text section. This string will be appended along with finally generated new log files names. For example, in above figure 2.1.1, user has entered “000” in text section, hence finally generated new 3 log files names will be mystring1_000.log, mystring2_000.log and yourstring1_000.log. This option will be helpful to user who wants to pre-append current date or time or any string into log files generated as output for future reference.

- You have 2 important buttons, “Format”, which will going to do all the formatting operation for you when you sit back and relax; and “Exit”, button to close the formatting window. Please keep in mind, no warnings/user messages will be displayed if user has missed any of the above parameter before pressing this button.

- A beautiful progress bar will display percentage of operation done and the format window will be in disabled state until 100% progress is completed. Until then user will not be allowed to do any operation, once formatting is started.

2.2 Open

Open is a tool to open log files into LATE workspace. Please keep in mind you can open only one log files into workspace once. Previous log contents will be removed before opening new one.

2.3 Filter

This is the most interesting and important option in LATE. This option will provide user to analyze workspace contents using set of Regular Expressions, which user can apply to do either highlight or delete wanted or unwanted contents, respectively. Below figure 2.3.1 shows filter window,

Fig 2.3.1

User has below controls to operate on Filter window,

- Select operation you would like to carry out, either HIGHLIGHT or DELETE the content from drop down menu.

- Select the regular expression type to apply for from second drop down menu options. Below are Regular Expressions provided for user benefits,

ROW – operates on particular row of string

COLUMN – operates on particular column of string in a comma separated strings

DATE – looks for matching date regular expression to operate on

TIME – looks for matching time regular expression to operate on

STRING – operates on particular string type regular expression

CUSTOM REG EXP – user can provide his/her own regular expression to operate on. To know more about supported regular expressions, please refer “Appendix A: Regular Expression Syntax” of this document.

- As user selects different regular expressions from above drop down lists, its corresponding regular expression string values will be displayed in text section

provided. User has all the controls to edit this text section to enter row/column numbers and enter his/her own expression type strings.

- Press “ADD” button to add above selected filter pattern into list. User can add multiple filters into this list one by one and apply all of them from top-to-bottom order at once for workspace contents.

- Select any filter in list, and press “REMOVE” button to particularly remove off selected filter component if it unwanted for user.

- “SAVE” button is very interesting and useful tool for widely distributed teams to sit and work at different places. User can save all filter components added to list into a CSV file format and carry this file to anywhere with them. This gives a flexibility of re-usability and distribution of operation over widely spread team over the globe to carry same operation on one interested file. User can also edit manually the filter file saved in CSV to add/remove/edit huge components for later usage. Please remember not to change the order of CSV file contents they saved in.

- “LOAD” button is, as discussed above, will be used to load back the filter file previously saved in CSV format.

- Press “Apply” button to apply all the listed filters into contents of workspace. Please be informed that you may face slight dragging in applying filters depending on the size of the contents in workspace and number of filters you are applying. Please also keep in mind, all the filters listed in list will be applied in top-to-down order; that means, output of top filter will be fed as an input to the next immediate filter. Hence, while deciding your filter please be careful.

- Press “cancel” button to exit from Filter window. By cancelling, it doesn’t remove filters added into the list and they will load back when you re-open Filter window in later stage.

I insist you please play with this filter option to get more information and what exactly it can do to help you get better analysis of logs. As I said, since it is a experimental tool, I cannot guarantee you all the filters works as expected. But I am sure I have designed them as enough as to exceed your expectations.

2.4 Search

Search option is disabled with this version of LATE. But in near future I have plans to provide the user pattern search options with this.

2.5 Export

Export is really powerful tool to export all your workspace contents into HTML page. These will sure short help user who has applied HIGHLIGHT filter into his contents and wants to export the contents along with highlighted important message.

2.6 Workspace

Last but not the least feature of LATE is its own workspace. User can load any text contents into these workspace area and/or type in his own contents along with.

3 Bottom lines

LATE is designed to help people all around the globe, keeping positive attitude that it will be helpful to address their painful needs. All the suggestions and comments are welcome. If anyone is ready to test it and provide me feedback with test result, I will be gladder.

Appendix A: Regular Expression Syntax

[Ref: https://www.ics.com/designpatterns/book/regexsyntax.html]

Much like a string, a regular expression is a series of characters; however, not all characters are taken literally. For example, while an "a" in a regular expression will match an "a" in the target string, the character "." will match any character. Here "." is called a meta-character. Another common meta-character is "*", which is used to indicate that zero or more of the preceding character may exist in the target string. For example, "a*" would match any number of "a"s (including zero) in a row. There are many different kinds of meta-characters, as illustrated below.

Following are some of the most commonly used meta-characters.

Special characters

. (the dot matches any character)

\n (matches the newline character)

\f (matches the form feed character)

\t (matches the tab character)

\xhhhh (matches the Unicode character whose code is the hexadecimal number hhhh in the range 0x0000 to 0xFFFF)

Quantifiers – Modifiers that specify the number of occurrences of the preceding character (or group) that may appear in the matching expression.

+ (1 or more occurrences)

? (0 or 1 occurrences)

* (0 or more occurrences)

{i,j} (at least i but not more than j occurrences)

Character Sets – Sets of allowable values for the character in the specified position of the matching expression. Several character sets are predefined:

\s (matches any whitespace character)

\S (matches any non-whitespace character)

\d (matches any digit character: '0' to '9')

\D (matches any non-digit character)

\w (matches any "word" character; i.e., any letter or digit or the underscore '_')

\W (matches any non-word character)

Character sets can also be specified in square brackets:

[AEIOU] (matches any of the characters 'A', 'E', 'I', 'O', or 'U')

[a-g] (the dash makes this a range from 'a' to 'g')

[^xyz] (matches any character except for 'x', 'y', and 'z')

Grouping and Capturing Characters – (round parentheses) can be used to form a group. Groups can be back-referenced, meaning that if there is a match, the grouped values can be captured and accessed in various ways.

For convenience, up to 9 groups can be referenced within the regular expression by using the identifiers \1 thru \9.

There is also a QRegExp member function cap(int nth), which returns the nth group (as a QString).

Anchoring Characters – Assertions that specify the boundaries of a matching effort.

The caret (^), if it is the first character in the regex, indicates that the match starts at the beginning of the string.

The dollar sign ($), when it is the last character in the regex, means that the effort to match must continue to the end of the string.

In addition, there are word boundary (\b) or non-word boundary (\B) assertions that help to focus the attention of the regex.

Table A.1. Examples of Regular Expressions

Pattern	**Meaning**
hello	Matches the literal string, hello
c*at	Quantifier: zero or more occurrences of c, followed by at: at, cat, ccat, etc.
c?at	Matches zero or 1 occurrences of c, followed by at: at or cat only.
c.t	Matches c followed by any character, followed by t: cat, cot, c3t, c%t, etc.
c.*t	Matches c followed by 0 or more characters, followed by t: ct, caaatttt, carsdf$#S8ft, etc.
ca+t	+ means 1 or more of the preceding "thing", so this matches cat, caat, caaaat, etc., but not ct.
c\.*t	Backslashes precede special characters to "escape them" so this matches only the string c.*t
c\\\.t	Matches only the string, c\.t
c[0-9a-c]+z	Between the 'c' and the 'z' one or more of the chars in the set [0-9a-c] – matches strings like c312abbaz and"caa211bac2z"

Pattern	Meaning
the (cat|dog) ate the (fish|mouse)	(Alternation) the cat ate the fish or the dog ate the mouse or the dog ate the fish, or the cat ate the mouse
\w+	A sequence of one or more alphanumerics (word chars), same as [a-zA-Z0-9]+
\W	A character which is not part of a word (punctuation, whitespace, etc)
\s{5}	Exactly 5 whitespace chars (tabs, spaces, or newlines)
^\s+	Matches one or more white space at the beginning of the string.
\s+$	Matches one or more white space at the end of the string.
^Help	Matches Help if it occurs at the beginning of the string.
[^Help]	Matches any single char except one of the letters in the word Help, anywhere in the string. (a different meaning for the metacharacter ^)
\S{1,5}	At least 1, at most 5 non-whitespace (printable characters)
\d	A digit [0-9] (and \D is a non-digit, i.e., [^0-9])
\d{3}-\d{4}	7-digit phone numbers: 555-1234
\bm[A-Z]\w+	\b means word boundary: matches mBuffer but not StreamBuffer