

--	--	--

Regular Expressions used in Notepad++ to clean up 035 strings to find OCLC numbers

Accompanies:

Roberts, C. and Corby, M. (2023, November 2-3) *Going through withdrawal (without too much suffering)* [Conference presentation]. ELSUG 2023 Virtual Conference.

To clean up 035 strings.

Example:

(OCoLC)24434924;522850;(KMK)522850-ksudb

turns into

*24434924

To be entered into the Find/Replace window in Notepad++ in the following order.

The "Regular expression" radio button in the Find/Replace window needs to be selected for these to work.

Step 1: Turn instances of (OCoLC)ocm#### and (OCoLC)ocn#### into (OCoLC)####

Find:

[^\\(OCoLC\\)]oc(m|n)

Replace:

\\(OCoLC\\)

Step 2: Remove instances of "ocm" and "ocn"

Find:

oc(m|n)

Replace:

leave field blank

Step 3: Remove any characters preceding (OCoLC) in a line

Find:

^\\.+(OCoLC\\)

Replace:

\\(OCoLC\\)

Step 4

Find:

^\\(OCoLC\\)(\\d+);(.+)\$

Replace:

*\\1

--	--	--

Step 5: Scan through list and deal with any anomalies individually.

Example: (OCoLC)on#### needs to be (OCLC)####

Step 6: Replace (OCoLC) at the beginning of a line with a *

Find:

^\(OCoLC\)

Replace:

*

Step 7: Remove lines that do not begin with a *

Find:

^[^*].+

Replace:

leave field blank

Step 8: Edit > Line Operations > Remove empty lines