Primer #03 on Creating Regular Expressions (RegEx) with Excel Spreadsheets for MailWasherPro Filters and Rules – page 1

For things here to make sense, I'm assuming you're already familiar with the RegEx concepts and conventions presented in Primer #01 and Primer #02.

NOTE: The goal here is to allow the user with minimal Knowledge of Regular Expressions to craft relatively simple RegEx and to use the spreadsheet application to generate more complex RegEx.

New in release 4.08: (Many Useful Enhancements)

- **Established character substitution method for RegEx control characters & lower case conversion for alpha characters.**
    - Automatic character Substitution implemented for RegEx control characters (**'$', '[', ']'**).
        - Room in substitution table to add up to four (4) additional control characters that might need to be tested in in RegEx strings at some point in the future.
        - Character Cross-Reference Table (Cols N-Q) and Character Substitution Table (Cols S-Y) are setup **in controls TAB**.
    - User may enter any of these listed control characters in the string to be tested for and application will substitute the hex equivalent character in RegEx strings.
        - Examples:
            - $ substituted with \x24
            - [substituted with \x5B
            -] substituted with \x5D
    - This substitution is implemented for **all TABs (except for special TAB).**

- **For All TABS, all data input in Green cells is converted to lower case.**
    - All RegEx data strings and test data strings are input into Green cells.
    - Lower case conversion facilitates testing of resultant rules/filters with RegEx testing programs.
    - Hyperlink to web-based RegEx test program from all Test Data Tables

- **For All TABS, (except Special TAB) character substitution is implemented.**
    - All RegEx data strings and test data strings are input into Green cells.
    - Character substitution process is implemented for all Green input **(except Special TAB)**
        - Special Characters setup in Substitution Table in Controls TAB.
        - Room for 4 additional characters in substitution table.
        - Table & Substitution test can be expanded if necessary.
    - Characters are converted to their hex equivalents to facilitate testing for control characters when they occur as text.
    - All RegEx data strings and all Test Data strings are converted to lower case to facilitate testing.
    - All RegEx data copied to the clipboard comes from Orange cells.

Primer #03 on Creating Regular Expressions (RegEx) with Excel Spreadsheets for MailWasherPro Filters and Rules – page 2

- **Special TAB - Compound TAB renamed Special TAB and has been revised and enhanced.**
  - Special TAB now has a Test Data Table
  - All RegEx data strings and test data strings are input into Green cells.
  - All RegEx data strings and all Test Data strings are converted to lower case to facilitate testing.
  - Buttons renamed for clarity
  - Screen Layout modified for clarity
  - All RegEx data copied to the clipboard comes from Orange cells.
  - HOME button added to allow return to first TAB

- **Head-Generic, Body-Generic and all "Equation" TABs enhanced.**
  - Character substitution process is implemented for all Green input **(except Special TAB)**
    - Special Characters setup in Substitution Table in Controls TAB.
    - Room for 4 additional characters in substitution table.
    - Table & Substitution test can be expanded if necessary.
  - All RegEx data strings and all Test Data strings are converted to lower case to facilitate testing.
  - HOME button added to allow return to first TAB
  - Buttons renamed for clarity
  - A NEW blank Equation TAB is replaces the old format for this TAB

New in release 4.06: (Status Flags & Reset Button)

- **STATUS FLAGS** indicate when you've changed RegEx data in a TAB
  - All TABs show a **STATUS FLAG** of C (meaning Changed, red code, yellow background) whenever the RegEx data has been changed.
  - The **head-generic** and **body-generic** TABs show a **STATUS FLAG** of A (red code, yellow background) when RegEx data in these TABs has been archived.
- **RESET** Status Button to clear the **STATUS FLAGS**
  - The **head-generic TAB** contains a **RESET** button that resets (clears) the **STATUS FLAGS**
  - **RESET** Button work **only** when:
    - **head-generic** and/or **body-generic** TABs have been archived after changing, **or**
    - **head-generic** and/or **body-generic** TABs have not changed since the last RESET
  - This means you cannot reset the STATUS FLAGS until you have first archived any changes made to the **head-generic** and/or **body-generic** TABs

Again, I maintain the majority my MailWasherPro filters/rules in an Excel spreadsheet. The older filters/rules were entered and maintained in MailWasherPro manually. I have examined some of these the older filters and have determined that it is possible to optimize some of these rules and to consolidate and remove some of the older filters and rules.

The question boils down to how you invest your time and is it worth it. For me, it's worth some investment of time because I get a LOT of SPAM, I have a lot of filters and that high number of filters can be reduced and that reduction will undoubtedly improve MailWasherPro performance.

A few of things that Firetrust could do that would make maintaining filters for MWP easier would be to:

1. Allow for the importing of rules/filters from a file while MWP is running, and/or
2. Allow for the specification of file(s) that contains rules/filters that can be dynamically updated while MWP is running, and/or
3. Provide statistics and timestamps on filters/rules so that less active and/or older filters/rules could be aged out in a similar way that old white/black list entries are aged out.

Since I started maintaining newer filters/rules in an Excel spreadsheet, just the act of building the filters had forced me to learn more of the rules. I probably know enough to be dangerous and I realize that there is a lot more to learn about Regular Expressions.

I'm happy to share what I've learned with the MailWasherPro community and I'm happy to receive feedback on what I'm sharing. What I'm doing can always be improved. I have included two (2) spreadsheets with this document. The spreadsheets are identical; however one of them has been enabled to execute Visual Basic macros – which saves a lot of time.

<mark>Why do any of this</mark>?

1. The more time I spend fighting SPAM, the less time I'm doing productive work or having fun. That makes spammers content, but it drives me crazy. So I examined what I could do that's productive and creative and that at the same time would help fight the spammers. Firstly, I created the spreadsheets to save the search strings. Secondly, I realized I could use EXCEL to transform the search strings into error-free Regular Expressions that could be copied/pasted into MWP filters and rules. So then, I created the VBA macros to help me deal with the repetitive and time consuming "stuff" I was doing within EXCEL.
2. NOTE: All TABs on the spreadsheet are protected - Green cells on the TABs are unprotected data entry cells
3. **Why save spam data at all?** (couple of reasons)
   a. In case you need to reconstruct rules/filters you've inadvertently deleted.
   b. In case you need to search for a string that is trapping a valid email.
   c. In case you want to look for new patterns in spam to enhance/optimize rules

<mark>This version of the Spreadsheet contains eighteen (18) relevant TABS:</mark>

NOTE 1: Invest a little time to skim (quickly) through the rejected emails MWP will trap using these regular expressions. This is a most important step (especially) when you roll out a new filter or make major changes to an existing filter. Your goal is to scan for any good emails (usually colored Green) that have been trapped by the filters, and then determine whether it makes more sense to:

- **adjust the filter(s) to remove or modify the string(s) that have trapped valid emails, or**
- **place the trapped email address on an override filter that designates such emails as legitimate before the RegEx filter tests for bad guys**

NOTE 2: Now all data in the <u>Orange RegEx output strings</u> is converted to lower case to make testing easier. Data in the <u>Green RegEx output strings</u> is <u>not</u> converted to lower case.

THE TABS

1. **head-generic TAB**	Special handling for RegEx strings found in email headers
2. **body-generic TAB**	Plain-vanilla RegEx handling for strings found in email body
3. **special TAB**	Special case RegEx strings for header and/or body. Some of these may merit promotion to a separate TAB
4. **vend1 TAB**	Special "equation" for vendor/provider entities (on the left) that are selling or offering products/services (on the right)
5. **vend2 TAB**	Special "equation" for vendor/provider products/services (on the left) with characteristics of these products/services (on the right)
6. **disease1 TAB**	Special "equation" for spammers posing as offering legitimate methods, techniques, etc. for curing illnesses, diseases, conditions, etc.
7. **disease2 TAB**	Special "equation" for spammers posing as offering legitimate methods, techniques, etc. for curing illnesses, diseases, conditions, etc. This is pretty much an inverse world from **disease1** TAB
8. **credit TAB**	This is to detect spammers posing as legitimate businesses offering incredible products and services dealing with credit cards, debit cards, debt, etc.
9. **photo TAB**	This is to detect spammers posing as legitimate businesses offering incredible deals to preserve family photos, convert movies to DVD, etc.
10. **women TAB**	This is to detect spammers offering various ethnic women etc., as brides, etc.
11. **military TAB**	This is to detect spammers offering various military/ruggedized equipment, for survivalists, etc.
12. **head-saved TAB**	This is to archive data captured in the **head-generic** TAB.
13. **body-saved TAB**	This is to archive data captured in the **body-generic** TAB.
14. **head-blank TAB**	This is a blank version of the **head-generic** TAB.
15. **body-blank TAB**	This is a blank version of the **body-generic** TAB.
16. **Equation-blank TAB**	This is blank version of the **equation** TAB.
17. **controls TAB**	This TAB contains a table of values that determines if a TAB's data has changed; a table of valid domain suffixes for testing on the Special Tab; a table of character equivalent values (ASCII-Hex); substitution table values.
18. **notes TAB**	This TAB contains notes regarding current & future developments.

1. <mark>**head-generic TAB**</mark>
   a. This TAB contains what I will call "generic" regular expressions that are random collections of data strings that spammers put in their email headers (from, to, subject, etc.). Sometimes these "data strings" occur only once. However, once they start to look familiar, they are suspects likely to return, so I form RegEx statements and archive them.

   b. In **Column A** (un-protected Green cells), **enter or paste** the strings you are searching for as plain text or in RegEx form.
      i. You may add simple RegEx controls to optimize entries
         1. Example
            a. unable to (display|show) full message

   c. In **Column B** (protected, Yellow cells) the following operations occur:
      i. lower case conversion
      ii. substitution for RegEx control characters
      iii. substitution of blanks for non-blank control characters ('-'   '_'   '.')
      iv. leading & trailing blanks removed

   d. In **Cell D1** (protected), the following operations occur:
      i. spreadsheet formulas replace blanks with the RegEx [-_.]*
      ii. leading & trailing blanks removed
      iii. spreadsheet formulas concatenate the RegEx from each row used (up to 70) from **Column B**, separated by pipes '|' into **Cell D1**
      iv. this replacement & concatenation sets up a regular expression to test for the presence of all possible combinations of these separators between the components whenever there is a blank.
         1. Examples:
            a. **(a-b-c, a_b_c, a.b.c, a b c, a-b.c, a-b_c, a b_c, etc.)**

      v. The asterisk (*) at following the right bracket sets up the regular expression to also test for the absence of separators between the components.
         1. Examples:
            a. **(abc, a-bc, ab.c, a_bc, etc.)**

e. **Cells H1 & H2** contain variants of **Column D** and are the selected cells for copy/paste operations

   i. **Cell H1** (protected, Orange cell) is the same as RegEx in Column D preceded by a pipe '|' character, meaning that it should be pasted to the end of an existing RegEx, just before the closing right parentheses ')' at the end of a RegEx that starts with a left parentheses '('.

   ii. **CELL H2** is the same RegEx with the leading pipe '|' removed and enclosed in parentheses so it can be used as a stand-alone rule/filter.

f. **Cells F1 & F3** (colored yellow) are blank when there has been no change to the data, contain a C when the data has been changed and contain an A when the changed data has been archived.

g. **Reset Button (only in this tab)**

   i. **RESET** Button works **only** when:

   ii. **head-generic** and/or **body-generic** TABs have been archived after changing, **or**

   iii. **head-generic** and/or **body-generic** TABs have not changed since the last RESET

   iv. This means you cannot reset the STATUS FLAGS until you have first archived any changes made to the **head-generic** and/or **body-generic** TABs

h. **Append Button**

   i. Click this button to copy the Orange cell **(H1)** contents (RegEx) to the clipboard, where it can be appended to an existing MWP rule/filter..

   ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

   iii. You can then paste the clipboard contents of this Orange cell at the end of a Rule in an MWP Filter - again, before the closing right parentheses ')'.

i. **Full Copy Button**

   i. Click this button to copy the Orange cell **(H2)** contents (RegEx) to the clipboard, where it can be pasted into an MWP rule/filter or into a RegEx test program.

   ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

   iii. You can then paste the clipboard contents of this Orange cell as a stand-alone rule in a filter.

j. **Sort Button**

   i. Click this button to sort the data in Column A. Makes for easier reading of the RegEx components

k. **Archive Button**

   i. Clicking this button will:

   1. Copy data from Column B to next available row in the **head-saved** TAB.

2. Clear the data from Column A of the **head-generic** TAB.

l. **Enter Test Data (Column G)**

    i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

m. **Paste Test Data (Column H)**

    i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

n. **Test Copy Button** (for test data)

    i. Click this button to copy the ordered test data contents into the clipboard, where it can be pasted into a RegEx test program.

2. **body-generic TAB**
   a. This TAB contains what I will call "generic" regular expressions that are random collections of data strings that spammers put in the body of their emails.
   b. Sometimes these "data strings" occur only once. However, once they start to look familiar, they are suspects likely to return, so I form RegEx statements and archive them.

   c. In **Column A** of the **body-generic** TAB (un-protected Green cells), **enter or paste** the strings you are searching for as plain text or in RegEx form.

   d. In **Column B** (protected, Yellow cells), the data is converted to lower case and substitution for RegEx control characters is done.

   e. In **Column D** (protected), spreadsheet formulas will concatenate these strings with pipe '|' separators. In addition, leading and trailing blanks will be removed from the string.

      i. You may add simple RegEx controls to optimize entries
         1. Example
            a. unable to (display|show) full message
      ii. In **Column D** (protected), up to 70 simple expressions (Column B) are concatenated into one big regular expression.

   f. **Cells H1 & H2** contain variants of Column D and are the selected cells for copy/paste operations
      i. **Cell H1** (protected, Orange cell) is the same as RegEX in Column D preceded by a pipe '|' character, meaning that it should be pasted to the end of an existing RegEx, just before the closing right parentheses ')' at the end of a RegEx that starts with a left parentheses '('.
      ii. **CELL H2** is the same RegEx with the leading pipe '|' removed and enclosed in parentheses so it can be used as a stand-alone rule/filter.

   g. **Cells F1 & F3** (colored yellow) are blank when there has been no change to the data, contain a C when the data has been changed and contain an A when the changed data has been archived.

h. **Append Button**
   i. Click this button to copy the Orange cell **(H1)** contents (RegEx) to the clipboard, where it can be appended to an existing MWP rule/filter..
   ii. This is the equivalent of clicking on the cell and then typing CTRL-C.
   iii. You can then paste the clipboard contents of this Orange cell at the end of a Rule in an MWP Filter - again, before the closing right parentheses ')'.

i. **Full Copy Button**
   i. Click this button to copy the Orange cell **(H2)** contents (RegEx) to the clipboard, where it can be pasted into an MWP rule/filter or into a RegEx test program.
   ii. This is the equivalent of clicking on the cell and then typing CTRL-C.
   iii. You can then paste the clipboard contents of this Orange cell as a stand-alone rule in a filter.

j. **Sort Button**
   i. Click this button to sort the data in Column A. Makes for easier reading of the RegEx components

k. **Archive Button**
   i. Clicking this button will:
      1. Copy data from Column B to next available row in the **head-saved** TAB.
      2. Clear the data from Column A of the **head-generic** TAB.

l. **Home Button**
   i. Clicking this button to return you to the first TAB (head-generic)

m. **Enter Test Data (Column G)**
   i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

n. **Paste Test Data (Column H)**
   i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

o. **Test Copy Button** (for test data)
   i. Click this button to copy the ordered test data contents into the clipboard, where it can be pasted into a RegEx test program.

3. **special TAB**

    a. This TAB contains what I call "simple but somewhat complex" regular expressions that are pretty specific in looking for types of spam. At the time of design, these subject areas did not seem large enough to warrant a special entry procedure as in the case for the **vend1** and **disease1, etc.** TABs.

    b. There are multiple subject areas on this single TAB.

    c. Here again, the data entry cells (Green) are un-protected.

    d. NOTE: The subject areas of these RegEx could be anything. These subject areas are what needed to search for.

    e. Other than lower case conversion, there is no transformation of the data so any special RegEx code must be embedded in the string. So the key principle here is to work within the specific structure of the code. If something cannot fit in this structure, then it is either a candidate for a more complex TAB of its own or it is more generic data that goes in the **head-generic** TAB.

    f. **Column C Cells** (colored yellow) are blank when there has been no change to the data, and contain a **C** when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

        THESE SPECIAL CASE RegEx CAN BE CUSTOMIZED TO SUIT YOUR OWN SITUATIONS

    g. **Check For Domain Suffix Status**

        i. In **Cell D2** (unprotected, green), enter a domain suffix (with or without the period). You'll get the status in **Cell E2**.

    h. **SPECIAL1 - (Bogus Domain Suffix)**

        i. Spammers have started sending out thousands of emails using a bogus domain suffix. When/if spammers add new names, I just add to this list.

        ii. Enter RegEx data in **Cell B1** (unprotected, Green Cell)

        iii. **Full Copy1 Button** – click this button to copy **Cell B2** (protected, Orange cell) contents (RegEx) to the clipboard.

    i. **SPECIAL2 – (Fake Email Addresses Spoofing my Domains - Using Functional Names)**

        i. Spammers are sending out emails using several of my domains with "fake" functional names (admin, accounting, copier, etc.). When/if spammers start a new campaign of adding new functional names, I just add to this list.

        ii. Enter RegEx data in **Cell B4** (unprotected, Green Cell)

        iii. **Full Copy2 Button** – click this button to copy **Cell B5** (protected, Orange cell) contents (RegEx) to the clipboard.

j. **SPECIAL3 – (Vend3 – a Vendor Equation that did not warrant an equation TAB)**

    i.   This RegEx creates is for vendor offerings (products, services, etc.) on the left and a generic-like RegEx for vendor actions and adjectives (quotes, ranking, etc.) on the right. This was the **old vend2 TAB** but there was as much activity with this rule. So there is now a **new vend2 TAB**

    ii.   Enter RegEx data in **Cell B7** (unprotected, Green Cell)

    iii.   **Full Copy3 Button** – click this button to copy **Cell B8** (protected, Orange cell) contents (RegEx) to the clipboard.

k. **SPECIAL4 – (Open for custom use)**

    i.   Enter RegEx data in **Cell B10** (unprotected, Green Cell)

    ii.   **Full Copy4 Button** – click this button to copy **Cell B11** (protected, Orange cell) contents (RegEx) to the clipboard.

l. **SPECIAL5 – (Doctor Oz)**

    i.   Spammers think it's really cool to promote stuff from Dr Oz using many different character substitutions. I've not encountered anything like this except for Dr Oz.

    ii.   Enter RegEx data in **Cell B13** (unprotected, Green Cell)

    iii.   **Full Copy5 Button** – click this button to copy **Cell B14** (protected, Orange cell) contents (RegEx) to the clipboard.

m. **SPECIAL6 – (Fake Emails form My Domains - Using Randomly Generated Names)**

    i.   This spoof was consistently using fake names ending in a number so it was pretty easy to design a filter that would trap such an animal.

    ii.   Enter RegEx data in **Cell B16** (unprotected, Green Cell)

    iii.   **Full Copy6 Button** – click this button to copy **Cell B17** (protected, Orange cell) contents (RegEx) to the clipboard.

n. <mark>**SPECIAL7 - (Suspect Domain Suffix)**</mark>

  i.   Spammers are sending out emails with legitimate suffixes that are notorious for spamming such as: .ru, .vn, etc.   When/if spammers add new names, I just add to this list.

  ii.   Enter RegEx data in **Cell B19** (unprotected, Green Cell)

  iii.   **Full Copy7 Button** – click this button to copy **Cell B20** (protected, Orange cell) contents (RegEx) to the clipboard.

o. <mark>**SPECIAL8 – (Emails with High Spam Scoring Messages)**</mark>

  i.   Email messages often contain scores from ISP SPAM FILTERS. In some cases you will find that the message was given an unusual flag or an unusually high score.

   a.   Examples:
    i.   F62_Suspicious_Sender
    ii.   F28_Unwanted_Text_In_Message_Body

  ii.   Add these strings to this filter by entering RegEx data in **Cell B22** (unprotected, Green Cell)

  iii.   **Full Copy8 Button** – Click this button to copy **Cell B23** (protected, Orange cell) contents (RegEx) to the clipboard.

   a. Spam patterns detected were the left-hand side of an equation (representing vendors offering stuff) and the right-hand side of the equation (representing stuff vendors offered).

      i. Examples:

         i. **Left:** kmart, Walmart, sears, kroger, etc.

         ii. **Right:** e-cards, bonus points. Rewards, etc.

   b. This TAB creates two equations that are concatenated into a single RegEx in **Cell G1** – colored Orange.

   c. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

   d. In **Column A** of the **vend1** TAB (un-protected Green cells), **enter or paste** the "left-hand" strings you are searching for as plain text or in RegEx form.

      i. You may add simple RegEx controls to optimize entries

         1. Example:

            2. unable to (display|show) full message

   e. In **Column D** of the **vend1** TAB (un-protected Green cells), **enter or paste** the "right-hand" strings you are searching for as plain text or in RegEx form.

      i. You may add simple RegEx controls to optimize entries as in the Example above.

   f. In **Column B & Column E** (protected, Yellow cells), the data is converted to lower case and substitution for RegEx control characters is done.

g. In **Cell J1** & **Cell L1** (protected, off-screen) & **Cell G1** (protected Orange Cell), spreadsheet formulas calculate:

  i. In **Cell J1** (protected, off-screen) & **Cell L1** (protected, off-screen), the following operations occur:

    i. spreadsheet formulas replace blanks with the RegEx [-_.]*
    ii. leading & trailing blanks removed
    iii. spreadsheet formulas concatenate the RegEx from each row used (up to 70) as follows:

      1. from **Column B**, separated by pipes '|' into **Cell J1** and
      2. from **Column E**, separated by pipes '|' into **Cell L1**

    iv. the resultant **Cell J1** and **Cell L1** are then concatenated into **Cell G1** and separated by the RegEx [-_.]*

    v. this replacement & concatenation sets up a regular expression to test for the presence of <u>all possible combinations</u> of these separators between the both sides of the equation whenever there is a blank.

      1. Examples:
        a. **(left-right, left_right, left.right, left right)**

    vi. The asterisk (*) at following the right bracket sets up the regular expression to also test for the <u>absence</u> of separators between both sides of the equation.

      1. Examples:
        a. **(leftright)**

h. **Cell G1** contains concatenated **Cell** J1 & **Cell L1** enclosed in parentheses

i. **Cell G2** (colored yellow) is blank when there has been no change to the data, contain a <mark>C</mark> when the data has been changed.

j. **Full Copy**
  i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..
  ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

k. **Sort Button**
  i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

l. **Home Button**
  i. Clicking this button to return you to the first TAB (head-generic)

m. <u>Enter Test Data (Column G)</u>

  i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

n. <u>Paste Test Data (Column H)</u>

  i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

o. <u>Test Copy Button</u> (for test data)

  i. Click this button to copy the ordered test data contents from **<u>Column H</u>** into the clipboard, where it can be pasted into a RegEx test program.

5. <mark>**vend2 TAB**</mark>
   a. Spam patterns detected were the left-hand side of an equation (using verbs and prepositions) and the right-hand side of the equation (representing the vendors).
   b. Examples:
      i. Examples:
         i. **Left-Right:** shop sears
         ii. **Left-Right:** at walmart
   c. This TAB creates two equations that are concatenated into a single RegEx in **Cell G1** – colored Orange.
   d. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

   e. In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

   f. **Full Copy**
      i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..
      ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

   g. **Sort Button**
      i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

   h. **Home Button**
      i. Clicking this button to return you to the first TAB (head-generic)

   i. **Enter Test Data (Column G)**
      i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

   j. **Paste Test Data (Column H)**
      i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

   k. **Test Copy Button** (for test data)
      i. Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.

6. **disease1 TAB**
    a. Spam patterns detected were the left-hand side of an equation (representing remedies, cures, etc.) and the right-hand side of the equation (representing the diseases, conditions, illnesses, etc.).
        i. Examples:
            i. **Left-Right:** prevents memory loss
            ii. **Left-Right**: cures diabetes
    b. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

    c. In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

    d. **Full Copy**
        i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..
        ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

    e. **Sort Button**
        i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

    f. **Home Button**
        i. Clicking this button to return you to the first TAB (head-generic)

    g. **Enter Test Data (Column G)**
        i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

    h. **Paste Test Data (Column H)**
        i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

    i. **Test Copy Button** (for test data)
        i. Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.

7. **disease2 TAB**

   a. Spam patterns detected were the left-hand side of an equation (representing diseases, illnesses, etc.) and the right-hand side of the equation (representing the remedies and cures, etc.).

      i. Examples:

         i. **Left-Right:** memory loss reversed

         ii. **Left-Right**: diabetes cured

   b. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

   c. In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

   d. **Full Copy**

      i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..

      ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

   e. **Sort Button**

      i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

   f. **Home Button**

      i. Clicking this button to return you to the first TAB (head-generic)

   g. **Enter Test Data (Column G)**

      i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

   h. **Paste Test Data (Column H)**

      i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

   i. **Test Copy Button** (for test data)

      i. Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.

8. <mark>**credit TAB**</mark>

   a. Spam patterns detected were the left-hand side of an equation (representing credit, debit, debt) and the right-hand side of the equation (financial products & services, etc.).
      i. Examples:
         i. **Left-Right:** credit relief
         ii. **Left-Right**: debt counseling

   b. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

   c. In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

   d. **Full Copy**
      i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..
      ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

   e. **Sort Button**
      i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

   f. **Home Button**
      i. Clicking this button to return you to the first TAB (head-generic)

   g. **Enter Test Data (Column G)**
      i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

   h. **Paste Test Data (Column H)**
      i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

   i. **Test Copy Button** (for test data)
      i. Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.

9. <mark>**photo TAB**</mark>

    a.   Spam patterns detected were the left-hand side of an equation (representing actions to preserve or save something) and the right-hand side of the equation (objects to be saved, etc.). **NOTE: Some entities appear on both sides of the equation.**

        i.   Examples:

            i.   **Left-Right:** digitize films

            ii.   **Left-Right**: preserve tapes

    b.   **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

    c.   In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

    d.   **Full Copy**

        i.   Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..

        ii.   This is the equivalent of clicking on the cell and then typing CTRL-C.

    e.   **Sort Button**

        i.   Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

    f.   **Home Button**

        i.   Clicking this button to return you to the first TAB (head-generic)

    g.   **Enter Test Data (Column G)**

        i.   I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

    h.   **Paste Test Data (Column H)**

        i.   When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

    i.   **Test Copy Button** (for test data)

        i.   Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.

10. <mark>**women TAB**</mark>
    a. Spam patterns detected were the left-hand side of an equation (representing ethnicities such as Korean, Russian, Filipina, Cambodian, Baltic, etc.) and the right-hand side of the equation (representing women, girls, brides, sluts, escorts, etc.). However, some entities appear on both sides of the equation.
        i. Examples:
            i. **Left-Right:** Ukrainian Brides
            ii. **Left-Right**: Asian Escorts
    b. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a <mark>C</mark> when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

    c. In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

    d. **Full Copy**
        i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..
        ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

    e. **Sort Button**
        i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

    f. **Home Button**
        i. Clicking this button to return you to the first TAB (head-generic)

    g. **Enter Test Data (Column G)**
        i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

    h. **Paste Test Data (Column H)**
        i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

    i. **Test Copy Button** (for test data)
        i. Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.

11. military TAB

    j. Spam patterns detected were the left-hand side of an equation (representing entities such as survivalists, Navy Seals, Special Ops Groups, etc.) and the right-hand side of the equation (representing objects such as flashlights, lanterns, lamps, lasers, etc.). **NOTE: Some entities appear on both sides of the equation.**

        i. Examples:

            i. **Left-Right:** military grade flashlights

            ii. **Left-Right**: surveillance camera

    k. **Cell G2** (colored yellow) are blank when there has been no change to the data, contains a **C** when the data has been changed. To Reset this status, Click the RESET Button in the **head-generic TAB**

    l. In **Column A**, **Column D**, **Column B**, **Column E**, **Cell J1**, **Cell, L1**, **Cell G1**, **Cell G2**, same as in **vend1 TAB**.

    m. **Full Copy**

        i. Click this button to copy the Orange cell **(G1)** contents (RegEx) to the clipboard, where it can be pasted as a stand-alone MWP rule/filter..

        ii. This is the equivalent of clicking on the cell and then typing CTRL-C.

    n. **Sort Button**

        i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components

    o. **Home Button**

        i. Clicking this button to return you to the first TAB (head-generic)

    p. **Enter Test Data (Column G)**

        i. I found that these complex RegEx were hard to test without having some test data. Then when analyzing the test data is not easy to identify problems unless the data is sequenced. So when you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H

    q. **Paste Test Data (Column H)**

        i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H and converts it to lower case.

    r. **Test Copy Button** (for test data)

        i. Click this button to copy the ordered test data contents from **Column H** into the clipboard, where it can be pasted into a RegEx test program.