

# Primer #00 – Using “build-regex-for-junkmail-filters”

## User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 1

For things here to make better sense, I’m you might skim through the RegEx concepts and conventions presented in Primer #01 and Primer #02.

**NOTE: This document has been renamed to Primer #00**

### Release 4.17

**Note: Cumulative Release History has been moved to the end of this User Guide.**  
**The most current Release History is also on the following page.**

**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.

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 so your suggestions will be appreciated. The spreadsheet has been enabled to execute Visual Basic macros – which saves a significant amount of time.

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 2

#### New in release Cumulative (4.17, including 4.15-4.16): Enhancements

- Renamed this User Manual to Primer #00 so that it is first in the series
- Corrected a few minor VBA Errors from previous versions
  - **Corrected the way “optimization” is implemented**
- New with v4.17 - More practical use of the ‘Controls’ Tab, moving towards making the application fully parameterized and table driven and expandable.
  - **Not there yet, but started the process**
  - **Does not affect current user operations**
- Corrections and optimizations to some of the regular expression examples in the tabs
  - **Fixed an obscure “Dr Oz” problem**
  - **Fixed an obscure RegEx problem on Women tab**
- Made additions to the test data included as further testing has shown the need for additional test cases
- **Added an experimental 3<sup>rd</sup> RegEx field in “Women’ tab.**
  - **Col H is a “prefix” RegEx string that is appended to the front of the concatenated adjusted RegEx strings in Col B and Col E**
  - **So the resultant output Cell J1 is the contatenation of Col H, Col B and Col E in that order**
  - **I was testing some RegEx strings with data that I felt needed a 3<sup>rd</sup> field to be added as a conditional prefix.**
  - **Later testing showed that a small change in the RegEx in Cols B and E would trap the data so there might not be a need for a prefix or a third field, but since the code had been de-bugged and added to the page, I decided to leave it in for now.**
  - **Cols A and D can be easily copied onto a generic 2-field Tab**
  - **Some user may find a practical use for the third field**
- Fix launch to RegEx test programs
- Fix copy of Disease1 Tab test data (was incorrectly copying vend1 test data)
- Updated Primers 01, 02 and 04 to release 4.17

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 3

#### **Why do any of this?**

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 RegEx strings. Secondly, I realized I could use EXCEL to transform the simple RegEx strings into error-free more complex 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 manually.
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

#### **This version of the Spreadsheet contains twenty (20) relevant TABS:**

**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 by MWP) 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 RegEx and Test Strings (colored Orange) is converted to lower case to make testing easier.

**NOTE3:** For all TABs (except for the special TAB) automatic character Substitution has been implemented for the following RegEx control characters: (‘\$’, ‘[’, ‘]’, ‘/’).

- There is room in substitution table to add up to ten (10) additional control characters that might need to be tested in in RegEx strings at some point in the future. This can be expanded further if necessary.
- Character Cross-Reference Table (Cols N-R, Rows 2-96) is setup in the **controls Tab**
- Character Substitution Table (Cols T-Z, Rows 2-18) is setup in the **controls TAB**.
- A **special TAB** Character Substitution Table (Cols T-Z, Rows 24-46) is setup in the **controls TAB**.

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 4

- For the **non-special 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
    - / substituted with \x2F
    - This substitution is implemented for all TABs (except for special TAB).
- **New with v4.15 Character Substitution Table for special TAB**
  - I’ve had to re-think the character substitution issue for a problem I faced in the special TAB. I created a RegEx string that was testing for a right parenthesis and that caused an unbalanced parenthesis error.
  - I needed a way in the special TAB to convert the stray parenthesis to its hex equivalent so I came up with the idea of a second character substitution table.
  - Until the issue of whether a character is literal data character or a controls character becomes clearer to me, I’ll use this second character substitution table only for the special TAB.
  - The special TAB data does not undergo the same string transformations done to the other TABS – where delimiters (.-\_) are converted to blanks and then all blanks are transformed to the sub-expression [-\_]\*.
  - If a character on this second table is enclosed in single quotes, then it will be converted to its data character equivalent in hex format – so that a slash in single quotes ‘/’ will be converted to \x2F as above.
  - See the special TAB substitution table (Cols T-Z, Rows 24-46) for the conversion specifications.

**NOTE4:** When you click the **Append** or **Full Copy** button, the RegEx string undergoes minimal syntax checking.

- **A Double Pipe ‘| |’ was found in a RegEx string**
  - Two pip characters in succession (a double pipe ‘| |’) should not be in a RegEx string as it would signify that everything would pass what that RegEx expression is testing for.
  - An error message warns user to check input columns (green cells) for stray data.
    - Because of the method used to create complex RegEx from simple RegEx, stray data somewhere in the input columns (green cells) is the most likely cause of a double pipe

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 5

- The unintentional entry of a double pipe in a RegEx string could also trigger this error
- **Unbalanced Parentheses in RegEx String**
  - **RegEx expressions must have equal numbers of Left and Right Parentheses**
  - **An error message warns user to check input columns (green cells) unbalanced parentheses.**
    - **The error message returns the counts of Left and Right Parentheses in the String.**
    - **The most likely culprit for this is RegEx data that you most recently added or changed.**
    - **Using the MS-Excel find command (ctrl-f) command to search through the input columns (green cells) may be useful in determining where you may have entered or omitted a left or right parenthesis.**
- **Optimizing RegEx Strings**
  - **The Optimize Button now allows you to combine (merge) selected RegEx expressions on a TAB by putting an ‘X’ in the select column for the expressions to be combined.**
    - **This rationale for this – allows user:**
      - **To combine RegEx having common element(s) of your choosing (such as alpha, meaning, similarity, etc.), or**
      - **To combine RegEx to conserve space on the TAB as the current maximum number of rows allowed is 70.**
  - **The select column(s) are the salmon colored columns to the right of the input RegEx strings (green cells) and the right of the converted RegEx strings (yellow cells)**
    - **Column C for generic TABS**
    - **Columns C & F for equation TABS**
  - **The Optimize routine combines the selected expressions on a single line, separated by pipes and enclosed in parentheses.**
  - **The routine also deletes the components selected for the merged expression and resorts the TAB to eliminate blank rows.**
  - **NOTE: The user can also make manual optimizations to the RegEx.**
    - **The automation of combining (merging) RegEx reduces the probability of committing errors.**
  - **An error message (Invalid Selection(s) for Optimization ) is displayed if the user:**
    - **has selected a blank row for optimization**
    - **has entered a character other than ‘x’ or null in the select column(s) (salmon colored – Column C or Column F)**

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 6

#### **THE TABS**

<b>1. <u>head-generic TAB</u></b>	Special handling for RegEx strings found in email headers
<b>2. <u>body-generic TAB</u></b>	Plain-vanilla RegEx handling for strings found in email body
<b>3. <u>special TAB</u></b>	Special case RegEx strings for header and/or body. Some of these may merit promotion to a separate TAB
<b>4. <u>vend1 TAB</u></b>	Special “equation” for vendor/provider entities (on the left) that are selling or offering products/services (on the right)
<b>5. <u>vend2 TAB</u></b>	Special “equation” for vendor/provider products/services (on the left) with characteristics of these products/services (on the right)
<b>6. <u>disease1 TAB</u></b>	Special “equation” for spammers posing as offering legitimate methods, techniques, etc. for curing illnesses, diseases, conditions, etc.
<b>7. <u>disease2 TAB</u></b>	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 <b>disease1 TAB</b>
<b>8. <u>credit TAB</u></b>	This is to detect spammers posing as legitimate businesses offering incredible products and services dealing with credit cards, debit cards, debt, etc.
<b>9. <u>photo TAB</u></b>	This is to detect spammers posing as legitimate businesses offering incredible deals to preserve family photos, convert movies to DVD, etc.
<b>10. <u>women TAB</u></b>	This is to detect spammers offering various ethnic women etc., as brides, etc.
<b>11. <u>military TAB</u></b>	This is to detect spammers offering various military/ruggedized equipment, for survivalists, etc.
<b>12. <u>head-saved TAB</u></b>	This is to archive data captured in the <b>head-generic TAB</b> .
<b>13. <u>body-saved TAB</u></b>	This is to archive data captured in the <b>body-generic TAB</b> .
<b>14. <u>head-blank TAB</u></b>	This is a blank version of the <b>head-generic TAB</b> .
<b>15. <u>body-blank TAB</u></b>	This is a blank version of the <b>body-generic TAB</b> .
<b>16. <u>equation-blank TAB</u></b>	This is blank version of the <b>equation TAB</b> .
<b>17. <u>controls TAB</u></b>	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.
<b>18. <u>notes TAB</u></b>	This TAB contains notes regarding current & future developments.
<b>19. <u>regex testers TAB</u></b>	Comparison of RegEx Test Programs I’ve used
<b>20. <u>regex test data</u></b>	An archived collection of various RegEx test data

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 7

#### **1. head-generic TAB**

- 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.)**

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 8

- 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 the data is unchanged, and contain a **C** when the data has changed, and contain an **A** when the changed data has been archived. The TAB also changes color when data changes. To Reset the status and TAB color, Click the RESET Button in the **head-generic TAB**
  
- 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
  - v. **Once a RESET has been performed, the application is automatically saved**
  
- 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



## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page 9

- 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. **HeadSave Button**
  - i. Clicking this button will take you to the head-saved TAB
  
- m. **BodySave Button**
  - i. Clicking this button will take you to the body-saved TAB
  
- n. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
  
- o. **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.
  
- p. **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.
  
- q. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String
  
- r. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.
  
- s. **Char Subst Button**
  - i. Clicking this button takes you to the Character Table and Character Substitution Table in the controls TAB.
  
- t. **Test Setup Button**
  - i. Clicking this button takes you to the Setup RegEx Test Program Table in the controls TAB.
  
- u. **Save Button**
  - i. Clicking this button will save the application in its current state.
  
- v. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 10

#### **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 the data is unchanged, and contain a **C** when the data has changed, and contain an **A** when the changed data has been archived. The TAB also changes color when data changes. To Reset the status and TAB color, Click the RESET Button in the **head-generic TAB**
- h. **In conjunction with Cells F1 & F3 – TABs change color when RegEx data has changed**
  - i. head-generic & body generic TABs change to red when **Modified**
  - ii. head-generic & body generic TABs change to green when **Archived**
  - iii. All other TABs change to green when **Modified**
  - iv. RESET can be run only when **all** modified TABs are green

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 11

v. Tabs set to default following RESET

i. **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 ')’.

j. **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.

k. **Sort Button**

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

l. **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.

m. **BodySave Button**

- i. Clicking this button will take you to the body-saved TAB

n. **Home Button**

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

o. **Optimize Button**

- i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.

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

- i. 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)**

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 12

- 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 into the clipboard, where it can be pasted into a RegEx Test Program.
  
- s. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.
  
- t. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 13

#### **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. With v4.15, there is a special TAB character substitution table that is performed along with lower case conversion. While any special RegEx controls must be embedded in the string, there is no conversion of delimiters to blanks. However any character enclosed in single quotes (') that is found in the special TAB table will be converted to its hex equivalent. 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. **Cells in Column D** (colored yellow) are blank when the data is unchanged, and contain a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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)

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 14

- 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.

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 15

#### **n. SPECIAL7 - (Suspect Domain Suffix)**

- 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. SPECIAL8 – (Emails with High Spam Scoring Messages)**

- 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)

#### **p. Full Copy8 Button**

- i. Click this button to copy **Cell B23** (protected, Orange cell) contents (RegEx) to the clipboard.

#### **q. Home Button**

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

#### **r. Enter Test Data (Column F)**

- i. When you enter various iterations of test data in Column F, the spreadsheet creates an ordered list in Column G

#### **s. Paste Test Data (Column G)**

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

#### **t. Test RegEx Button**

- i. Clicking this button launches the RegEx Test Program specified in the controls TAB.

- u. The **Numbers** in Column C are the number of characters in each Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 16

#### **4. vend1 TAB**

- 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) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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 [-\_]\*



## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 17

- v. this replacement & concatenation sets up a regular expression to test for the presence of all possible combinations of these separators between the both sides of the equation whenever there is a blank.
  - a. Examples: **(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 absence of separators between both sides of the equation.
  - a. Example: **(leftright)**
- h. **Cell G1** contains concatenated **Cell J1** & **Cell L1** enclosed in parentheses
- i. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, Click the RESET Button in the head-generic TAB
- 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. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- n. **Enter Test Data (Column F)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column G
- o. **Paste Test Data (Column G)**
  - i. When you enter various iterations of test data in Column F, the spreadsheet creates an ordered list in Column G and converts it to lower case.
- p. **Test Copy Button** (for test data)
  - i. Click this button to copy the ordered test data contents from **Column G** into the clipboard, where it can be pasted into a RegEx Test Program.
- q. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.

**Primer #00 – Using “build-regex-for-junkmail-filters”**

**User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page**

**18**

- r. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 19

#### **5. vend2 TAB**

- 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; 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) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- j. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
- k. **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.
- l. **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.
- m. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.

**Primer #00 – Using “build-regex-for-junkmail-filters”**

**User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page  
20**

- n. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 21

#### 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; cures diabetes
- b. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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..
- 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. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- h. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
- i. **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.
- j. **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.
- k. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.
- l. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 22

#### 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; diabetes cured
- b. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes (‘|’) and enclosed in parentheses.
- h. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
- i. **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.
- j. **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.
- k. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.
- l. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 23

#### **8. credit TAB**

- 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; debt counseling
- b. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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)
- u. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- g. **Enter Test Data (Column G)**
  - i. 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.
- j. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.

**Primer #00 – Using “build-regex-for-junkmail-filters”**

**User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page**

**24**

- k. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String



## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 25

#### 9. **photo TAB**

- 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; preserve tapes
- b. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- h. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
- i. **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.
- j. **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.
- k. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.
- l. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 26

#### **10. women TAB**

- 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; Asian Escorts
- b. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, 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. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- h. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
- i. **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.
- j. **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.
- k. **Test RegEx Button**
  - i. Clicking this button launches the RegEx Test Program specified in the controls TAB.

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 27

- I. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

#### **11. military TAB**

- m. 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; surveillance camera
- n. **Cell G2** (colored yellow) is blank when the data is unchanged, and contains a **C** when the data has changed. The TAB also changes color when data has changed. To Reset the status and TAB color, Click the RESET Button in the **head-generic TAB**
- o. In **Column A, Column D, Column B, Column E, Cell J1, Cell, L1, Cell G1, Cell G2**, same as in **vend1 TAB**.
- p. **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.
- q. **Sort Button**
  - i. Click this button to sort the data in **Column A** and **Column D**. Makes for easier reading of the RegEx components
- r. **Home Button**
  - i. Clicking this button to return you to the first TAB (head-generic)
- s. **Optimize Button**
  - i. Clicking this button combines selected expressions in to a single expression separated by pipes ('|') and enclosed in parentheses.
- t. **Enter Test Data (Column G)**
  - i. When you enter various iterations of test data in Column G, the spreadsheet creates an ordered list in Column H
- u. **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.
- v. **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.
- w. **Test RegEx Button**

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 28

- i. Clicking this button launches the RegEx Test Program specified in the controls TAB.
- x. The **Number** in between the **Test Copy** Button and the **Test RegEx** Button is the number of characters in the Full Copy RegEx String

#### **12. head-save TAB**

- a. This contains all of the data from the head-generic TAB that has been Archived
- b. Archiving this data gives you the ability to:
  - i. Look for and recognize patterns
  - ii. Optimize simple RegEx expressions
  - iii. Find and correct RegEx errors
  - iv. Reconstruct rules/filters you’ve inadvertently deleted
  - v. Search for a string that is trapping a valid email
- c. **Sort Button**
  - i. Click this button to sort the data in **Column B**. Makes for easier reading of the RegEx components
- d. **Home Button**
  - i. Clicking this button to return you to the first TAB (head-generic)
- e. **Last Row Button**
  - i. Moves the focus to the last row used in this archive

#### **13. body-save TAB**

- f. This contains all of the data from the body-generic TAB that has been Archived
- g. Archiving this data gives you the ability to:
  - i. Look for and recognize patterns
  - ii. Optimize simple RegEx expressions
  - iii. Find and correct RegEx errors
  - iv. Reconstruct rules/filters you’ve inadvertently deleted
  - v. Search for a string that is trapping a valid email
- h. **Sort Button**
  - i. Click this button to sort the data in **Column B**. Makes for easier reading of the RegEx components
- i. **Home Button**

**Primer #00 – Using “build-regex-for-junkmail-filters”**  
**User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page**  
**29**

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

**j. Last Row Button**

i. Moves the focus to the last row used in this archive

### Release History

- Renamed this User Manual to Primer #00 so that it is first in the series
- Corrected a few minor VBA Errors from previous versions
  - **Corrected the way “optimization” is implemented**
- New with v4.17 - More practical use of the ‘Controls’ Tab, moving towards making the application fully parameterized and table driven and expandable.
  - **Not there yet, but started the process**
  - **Does not affect current user operations**
- Corrections and optimizations to some of the regular expression examples in the tabs
  - **Fixed an obscure “Dr Oz” problem**
  - **Fixed an obscure RegEx problem on Women tab**
- Made additions to the test data included as further testing has shown the need for additional test cases
- **Added an experimental 3<sup>rd</sup> RegEx field in “Women’ tab.**
  - **Col H is a “prefix” RegEx string that is appended to the front of the concatenated adjusted RegEx strings in Col B and Col E**
  - **So the resultant output Cell J1 is the contatenation of Col H, Col B and Col E in that order**
  - **I was testing some RegEx strings with data that I felt needed a 3<sup>rd</sup> field to be added as a conditional prefix.**
  - **Later testing showed that a small change in the RegEx in Cols B and E would trap the data so there might not be a need for a prefix or a third field, but since the code had been de-bugged and added to the page, I decided to leave it in for now.**
  - **Cols A and D can be easily copied onto a generic 2-field Tab**
  - **Some user may find a practical use for the third field**
- Fix launch to RegEx test programs
- Fix copy of Disease1 Tab test data (was incorrectly copying vend1 test data)
- Updated Primers 01, 02 and 04 to release 4.17

**New in release 4.15: (Enhancement(s))**

- **Correct a VBA Error from v 4.14**
- **New with v4.15 Character Substitution Table for special TAB**
  - I’ve had to re-think the character substitution issue for a problem I faced in the special TAB. I created a RegEx string that was testing for a right parenthesis and that caused an unbalanced parenthesis error.
  - I needed a way in the special TAB to convert the stray parenthesis to its hex equivalent so I came up with the idea of a second character substitution table.
  - Until the issue of whether a character is literal data character or a controls character becomes clearer to me, I’ll use this second character substitution table only for the special TAB.
  - The special TAB data does not undergo the same string transformations done to the other TABS – where delimiters (.-\_) are converted to blanks and then all blanks are transformed to the sub-expression [-\_]\*.
  - If a character on this second table is enclosed in single quotes, then it will be converted to its data character equivalent in hex format – so that a slash in single quotes ‘/’ will be converted to \x2F as above.
  - See the special TAB substitution table (Cols T-Z, Rows 24-46) for the conversion specifications.

**New in release 4.14: (Enhancement(s))**

- **Optimizing RegEx Strings**
  - **The Optimize Button now allows you to combine (merge) selected RegEx expressions on a TAB by putting an ‘X’ in the select column for the expressions to be combined.**
    - **This rationale for this – allows user:**
      - **To combine RegEx having common element(s) of your choosing (such as alpha, meaning, similarity, etc.), or**
      - **To combine RegEx to conserve space on the TAB as the current maximum number of rows allowed is 70.**
  - **The select column(s) are the salmon colored columns to the right of the input RegEx strings (green cells) and the right of the converted RegEx strings (yellow cells)**
    - **Column C for generic TABS**
    - **Columns C & F for equation TABS**
  - **The Optimize routine combines the selected expressions on a single line, separated by pipes and enclosed in parentheses.**
  - **The routine also deletes the components selected for the merged expression and resorts the TAB to eliminate blank rows.**
  - **NOTE: The user can also make manual optimizations to the RegEx.**
    - **The automation of combining (merging) RegEx reduces the probability of committing errors.**



**Primer #00 – Using “build-regex-for-junkmail-filters”**  
**User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page**  
**33**

**New in release 4.13: (Enhancement(s))**

- **Check RegEx String for unbalanced parentheses before copying RegEx to clipboard**
  - **Unbalanced Parentheses in RegEx String**
    - **RegEx expressions must have equal numbers of Left and Right Parentheses**
    - **An error message warns user to check input columns (green cells) unbalanced parentheses.**
      - **The error message returns the counts of Left and Right Parentheses in the String.**
      - **The most likely culprit for this is RegEx data that you most recently added or changed.**
      - **Using the MS-Excel find command (ctrl-f) command to search through the input columns (green cells) may be useful in determining where you may have entered or omitted a left or right parenthesis.**
  - **Could there be other syntax checking done to the RegEx string?**

**New in release 4.12: (Several Enhancements)**

- **Added automatic Application Save after RESET**
  - No need to do ctrl-s, etc.
- **Added SaveApp Button to Main TAB (head-generic)**
  - So you can manually save without needing to do ctrl-s, etc.
- **Check RegEx String for double pipe sequence before copying RegEx to clipboard**
  - **A Double Pipe ‘| |’ was found in a RegEx string**
    - **Two pip characters in succession (a double pipe ‘| |’) should not be in a RegEx string as it would signify that everything would pass what that RegEx expression is testing for.**
    - **An error message warns user to check input columns (green cells) for stray data.**
      - **Because of the method used to create complex RegEx from simple RegEx, stray data somewhere in the input columns (green cells) is the most likely cause of a double pipe**
      - **The unintentional entry of a double pipe in a RegEx string could also trigger this error**
  - **No need to check test data before clipboard copy as anything can be in test data strings.**
  - **Could there be other syntax checking done to the RegEx string?**

**New in release 4.11: (Several Enhancements)**

- **Excel Spreadsheet Application is renamed as:**
  - build-regex-for-junkmail-filters\_v4-11.xlsm
  - renamed all primers to lower case for web downloads & set to v4-11
  
- **Expanded Character Substitution Table in the controls TAB**
  - Table now provides for a maximum of seventeen (17) characters.
  - Seven (7) Characters are currently defined in this Table:
    - **Underscore ‘\_’**
    - **Dash ‘-’**
    - **Period ‘.’**
    - **Dollar ‘\$’**
    - **Left Bracket ‘[’**
    - **Right Bracket ‘]’**
    - **Slash ‘/’**
  - In the data entry conversion, blanks are substituted for the first 3 Substitution Table characters in all TABs except for the body-generic TAB and the special TAB
  - **Do not modify the first 3 Substitution Table Characters because the above substitution is essential for converting simple RegEx into complex RegEx.**
  - The character definitions for rows 9 through 18 remain open
  - NOTE: Any characters defined for substitution cannot be used as RegEx control characters when writing simple RegEx
  
- **Resolved issues with body-generic TAB.**
  - The body-generic TAB originally performed only the same character substitution as is done for the header-generic TAB and the equation TABS – substituting blanks for dashes (-), periods (.) and ( ) underscores. The substitution for these 3 characters in the body-generic TAB has been stopped.
  - The substitutions made in the body-generic TAB are now for the characters defined in rows 4 through 7 of the Substitution Table in the controls TAB:
    - Dollar ‘\$’ converted to ‘\x24’
    - Left Bracket ‘[’ converted to ‘\x5b’
    - Right Bracket ‘]’ converted to ‘\x5d’
    - Slash ‘/’ converted to ‘\x2f’

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 36

- **New Navigation Buttons**
  - **Button in header-saved TAB allows for jump to last used Row**
  - **Button in body-saved TAB allows for jump to last used Row**
  
- **Limit/Range Enforcement**
  - **Prevent copy of empty RegEx Data cells to clipboard**
  - **Prevent copy of empty Test Data columns to clipboard**
  - **Prevent Sort when number of rows available is 0 or 1**
  
- **Added Procedures for data archive TABS**
  - **Added documentation for head-save and body-save TABS**
  - **Added Navigation Buttons for head-save and body-save TABS**
  
- **Correct protection status in TAB cells that should have been unprotected**
  
- **Cleaned up some Visual Basic (VBA) code issues**

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 37

#### **New in release 4.10: (Many Enhancements)**

- **Added Test RegEx Button that replaces hyperlink**
  - Button launches the program that is specified in the controls TAB (AA1:AH7)
  
- **Added section of controls TAB (AA1:AH7) to allow user to specify up to five (5) RegEx Test Programs.**
  - Can be either windows program (local) or web-based program
  - User specifies the full path name or url name and application determines how to launch the program.
  
- **Added the RegEx control character '/' to character substitution table in controls TAB (T1:Y11), leaving 3 open placeholders for additional characters. '/' is concerted to '\x2f' so that '/' can be tested as data.**
  
- **Added Buttons on Home Page (header-generic TAB) to jump to the following control TAB sections:**
  - **Character Table & Character Substitution Tables**
    - Added Home Button on this page to return to header-generic TAB
  - **Select RegEx Test Program Table**
    - Added Home Button on this page to return to header-generic TAB

**Primer #00 – Using “build-regex-for-junkmail-filters”**  
**User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page**  
**38**

**New in release 4.09: (Many Enhancements)**

- **Added Tab Color Change whenever RegEx on any TAB has changed**
  - **TABs change color when RegEx data has changed**
    - head-generic & body generic TABs change to red when **Modified**
    - head-generic & body generic TABs change to green when **Archived**
    - All other TABs change to green when **Modified**
    - RESET can be run only when **all** modified TABs are green
    - Tabs set to default following RESET
- **Added New TAB navigation Buttons in several TABs**
- **Added hyperlink to <http://www.regex tester.com> to all Test Data Tables**
- **Standardized the protection of TABS**
- **Changed color of Test Data fields to be copied to clipboard to orange.**

**New in release 4.08: (Many 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 additional characters in the character 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.
- **Special TAB - Compound TAB renamed Special TAB and has been revised and enhanced.**
  - Special TAB now has a Test Data Table

## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 40

- 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 additional characters in the character 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



## Primer #00 – Using “build-regex-for-junkmail-filters”

### User Manual: Excel Spreadsheet To Create MailWasherPro Filters/Rules – page

#### 41

##### **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