

Using “Regular Expressions” when importing from Word or Excel

Gregory Furter
[EMLYON Business School](#)
Lyon, France
furter@em-lyon.com

2011 European Users Conference ◆ **Brussels**



Content

- ▶ Introduction
- ▶ Questionmark at EMLYON Business School
- ▶ What are “Regular Expressions”
 - Syntax examples
- ▶ Tools and methods
 - Detailed examples
- ▶ References
 - Web
 - Books
 - Apps

Introduction

Gregory Furter

- ▶ Work for over 10 years in IT support
 - Still do, being helpdesk supervisor for EMLYON IT Department
- ▶ Work for 1.5 years with Questionmark
 - The only one person in charge of QMP at EMLYON

EMLYON Business School

- ▶ Founded in 1872
- ▶ Small business school
 - 3200 students (40% non-French)
 - 5000 participants in executive education programs
 - 110 faculty members
- ▶ Nice rankings
 - Financial Times
 - 10th Business School in Europe
 - 3rd MBA in France
 - 13th MBA in Europe

Questionmark at EMLYON

Facts

- ▶ 30 assessments
- ▶ 3000 students evaluated in 1 year
- ▶ 11 professors using Questionmark
- ▶ 2000 questions created last year

My goal : developing QMP usage

- ▶ Because it is not widely used
 - Only 1 out of 10 professors actually use it
- ▶ Because it is not promoted
 - Neither by the school
 - Nor by the faculty

My means :

Offering the best possible experience

- ▶ Introducing the least changes in professors' habits
 - They are familiar with Word and Excel
- ▶ Not adding new tasks to already hard working assistants
- ▶ Providing the best feedback and advice
 - So the idea is worth spreading

What are Regexp ?

Developer toolbox

- ▶ Complicated syntax
 - "QuestionMark" will match "QuestionMark"
 - "[Qq]uestion[Mm]ark" will match :
 - QuestionMark
 - Questionmark
 - questionMark
 - Questionmark
 - "[0-9]" or "[AaBb]" indicate character sets
 - "^" indicates the beginning of a line
 - "\$" indicates the end of a line
 - "." indicates any character except line break

Developer toolbox

- ▶ Complicated syntax
 - "*" indicates that the preceding character may repeat zero to infinite times.
 - "+" indicates that the preceding character may repeat one or more times.
 - "?" indicates that the preceding character is optional.
 - "\" is used to escape special characters (like those above).
 - "()" are used to define groups
 - ... (see references for more)

Search and replace on steroids

- ▶ Fine tuning the search conditions
 - In one "sentence"
- ▶ Replacing depending on the match

Regex and QMP

- ▶ Unrelated
 - No regex implementation into Authoring Manager as far as I know
- ▶ May be helpful
 - As a preparation tool

Tools and methods

Creating the process basics

- ▶ Analyze how the writer issues their questions so far.
- ▶ Define a way to customize the original Word document in order to ease the import process so that :
 - Nobody has to copy/paste the questions into another software (Authoring Manager, QM Live...), you don't have to find out which are the correct answers or reformat the document yourself (see example).

Word example

▶ Before formatting

1. Regular expression stands for:
 - a. A form of language activism?
 - b. A text query syntax?
 - c. Search and Replace on steroids?
 - d. A programming language?
 - e. An authoring Manager option for Questionmark ?

▶ After formatting

1. MR Regular expression stands for:
 - a. -A form of language activism?
 - b. +A text query syntax?
 - c. +Search and Replace on steroids?
 - d. A programming language?
 - e. -An authoring Manager option for Questionmark ?

Here you want to make sure that the question writer formats (or re-formats) the Word document to ease the import process. In this example, you want to know the question type and the scoring for each answer. I prefer to put these information at the beginning of each line rather than at the end.

Using the right tool

- ▶ Notepad++
 - <http://notepad-plus-plus.org/>
 - Free source code editor and Notepad replacement.
 - Regexp enabled find/replace

Formatting

- ▶ Importing into Notepad++
 - Open the Word document with MS Word
 - Select All (CTRL+A)
 - Open Notepad ++
 - Create a new document (make sure it is UTF-8)
 - Paste into Notepad++ (CTRL+V)

```
1 1. MR Regular expression stands for:  
2 a. -A form of language activism?  
3 b. +A text query syntax?  
4 c. +Search and Replace on steroids?  
5 d. A programming language?  
6 e. -An authoring Manager option for Questionmark ?
```

Formatting

- ▶ Search for texts to be replaced
 - ▶ Questions wording
 - ▶ Open the "Find" window (CTRL+F)
 - ▶ Look for any line beginning with a number followed by a dot and a tab then a "MC" or "MR" and a space...
 - ▶ The regexp will be : `"^[0-9]+\.\s(M[CR]) "`

You may use the "Mark" tab from the find/replace window to highlight the result of your regexp and make sure it matches what you want to replace. Sometimes you'll have to make some trials and errors to find the exact syntax.

The round brackets "()" will be used as a back reference when replacing.

The screenshot shows a text editor window with three tabs: 'remplacement de textes.txt', 'procédure de modification des questions.txt', and 'new 2'. The main text area contains a list of questions:

- 1 1. MR Regular expression stands for:
- 2 a. -A form of language activism?
- 3 b. +A text query syntax?
- 4 c. +Search and Replace on steroids?
- 5 d. A programming language?
- 6 e. -An authoring Manager option for Questionmark ?
- 7
- 8 199. MC Regular expression stands for:
- 9 a. -A form of language activism?
- 10 b. +A text query syntax?
- 11 c. +Search and Replace on steroids?
- 12 d. A programming language?
- 13 e. -An authoring Manager option for Questionmark ?

The 'Mark' dialog box is open, showing the following settings:

- Find what: `^[0-9]+\.\s(M|CR)?`
- Buttons: Find All, Clear, Close
- Mark line
- Purge for each search
- Match whole word only
- Match case
- Wrap around
- In selection
- Search mode:
 - Normal
 - Extended (\n, \r, \t, \0, \x...)
 - Regular expression
- Transparency:
 - Transparency
 - On losing focus
 - Always

Formatting

- ▶ Replace with ASCII tags
 - Question wording
 - Go to the "Replace" tab
 - Replace with appropriate tag for question types and question wordings
 - Replace string will be : "<T>\1\n<Q>"
 - Hit "Replace All"

"\1" is a back reference to either "MC" or "MR" depending on the search result.

"\n" inserts a line feed, could be replace by "\r"

```

remplacement de textes.txt  procédure de modification des questions.txt  new 2
1 <T>MR
2 <Q>Regular expression stands for:
3 a. -A form of language activism?
4 b. +A text query syntax?
5 c. +Search and Replace on steroids?
6 d. A programming language?
7 e. -An authoring Manager option for Questionmark ?
8
9 <T>MC
10 <Q>Regular expression stands for:
11 a. -A form of language activism?
12 b. +A text query syntax?
13 c. +Search and Replace on steroids?
14 d. A programming language?
15 e. -An authoring Manager option for Questionmark ?

```

Replace

Find what: Find Next

Replace with: Replace

In selection Replace All

Match whole word only

Match case

Wrap around

Replace All in All Opened Documents

Search mode: Normal Extended (\n, \r, \t, \0, \x...)

Regular expression

Direction: Up Down

Transparency: On losing focus Always

Replace All 2 occurrences were replaced.

Formatting

- ▶ Search for texts to be replaced
 - ▶ Answers
 - ▶ Back to the "Mark" tab to test our regexp
 - ▶ Look for any line beginning with a letter followed by a dot and a tab then a "+" or "-" (+/- being optional)
 - ▶ The regexp will be : `"^[a-z]+\.\s([\+-]?)"`


```

remplacement de textes.txt  procédure de modification des questions.txt  new 2
1 <T>MR
2 <Q>Regular expression stands for:
3 a. -A form of language activism?
4 b. +A text query syntax?
5 c. +Search and Replace on steroids?
6 d. A programming language?
7 e. -An authoring Manager option for Questionmark ?
8
9 <T>MC
10 <Q>Regular expression stands for:
11 a. -A form of language activism?
12 b. +A text query syntax?
13 c. +Search and Replace on steroids?
14 d. A programming language?
15 e. -An authoring Manager option for Questionmark ?

```

Mark

Find | Replace | Find in Files | Mark

Find what:

Find All | Clear | Close

Mark line In selection

Purge for each search

Match whole word only

Match case

Wrap around

Search mode

Normal

Extended (\n, \r, \t, \0, \x...)

Regular expression

Transparency

On losing focus

Always

—|—

Formatting

▶ Replace with ASCII tags

▪ Answers

- Go to the "Replace" tab
- Replace with appropriate tag for question types and answers and scoring
- Replace string will be : "<C\1>"
- Hit "Replace All"

"\1" is a back reference to the "([\+-]?)" in the search expression and places a "+" or a "-" where it found one, and places nothing where it found nothing.

remplacement de textes.txt | procédure de modification des questions.txt | new 2

```

1 <T>MR
2 <Q>Regular expression stands for:
3 <C->A form of language activism?
4 <C+>A text query syntax?
5 <C+>Search and Replace on steroids?
6 <C>A programming language?
7 <C->An authoring Manager option for Questionmark ?
8
9 <T>MC
10 <Q>Regular expression stands for:
11 <C->A form of language activism?
12 <C+>A text query syntax?
13 <C+>Search and Replace on steroids?
14 <C>A programming language?
15 <C->An authoring Manager option for Questionmark ?

```

Replace

Find what:

Replace with:

In selection

Match whole word only

Match case

Wrap around

Search mode: Normal, Extended (\\n, \\r, \\t, \\0, \\x...), Regular expression

Direction: Up, Down

Transparency: On losing focus, Always

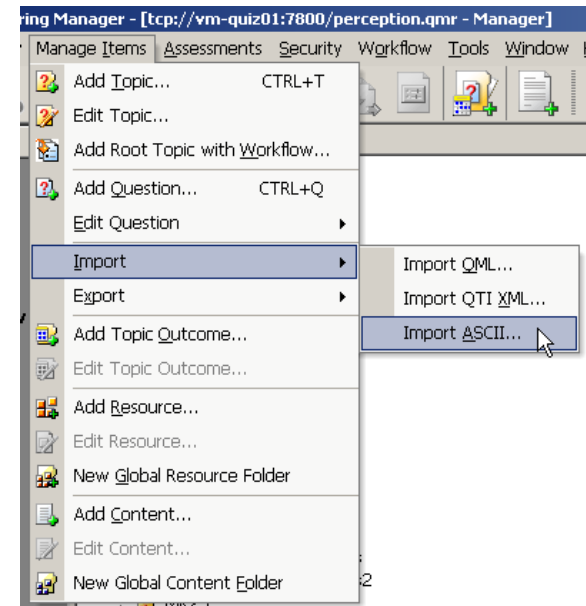
Formatting

- ▶ Make sure there is no untagged line remaining
 - Look for any line not beginning with a "<"
 - Regexp will be : "`^[^<]`"
 - For remaining comments, you may want to use a back reference, searching for "`^([^\<])`" and replacing with "`<#\>\1`"
- ▶ Save the file.

Importing into your repository

- ▶ In Authoring Manager
 - Select a question topic
 - Go to “Manage Items|Import|Import ASCII...”
 - Select the text file to be imported

Et voilà !



From an Excel document

- ▶ This is very unusual
- ▶ You may first reorganize the layout to match something similar to :

Question type ; Question wording ; Answer A ; Answer B ; Answer C ; ...

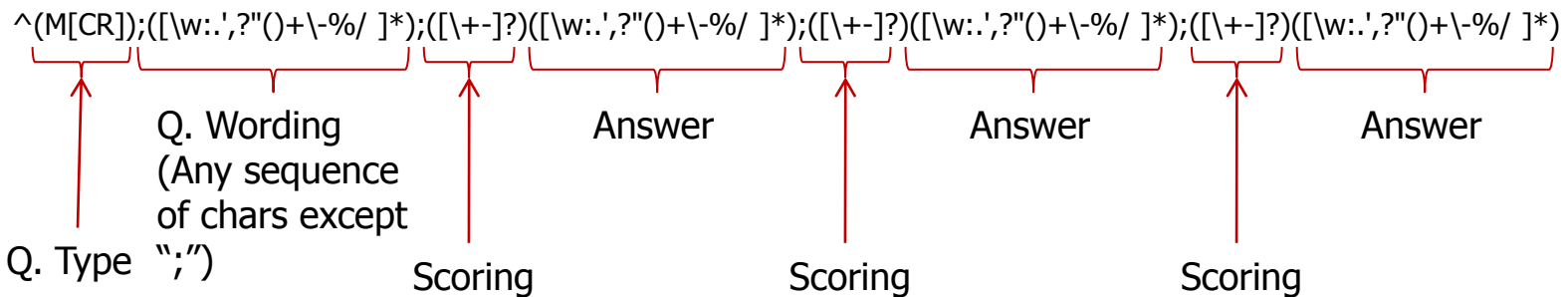
- ▶ Place the scoring indicator at the beginning of each answer (see example next page)
- ▶ Also make sure there is no “;” anywhere in the document.
- ▶ Save as CSV file and open with NotePad++

It will sometimes require a bit of excel formulas to get to the following result.

	A	B	C	D	E
1	Question Type	Question Wording	Answer A	Answer B	Answer C
2	MR	En Europe, un déposant ayant de l'argent sur un compte bancaire:	+court le risque de perdre ses fonds si la banque fait faillite.	-est protégé par le système de garantie des dépôts quels que soient les montants déposés.	+est protégé par le système de garantie des dépôts à hauteur de EUR 20 000.
3	MC	En France, la garantie bancaire prend la forme:	-d'une adhésion au Fonds de garantie des dépôts français, uniquement réservé aux ressortissants de nationalité française.	+d'une adhésion au Fonds de garantie des dépôts français, ouvert à tous les déposants, quels que soient leur nationalité ou leur lieu de	-d'une adhésion au Fonds de garantie européen.

From an Excel document

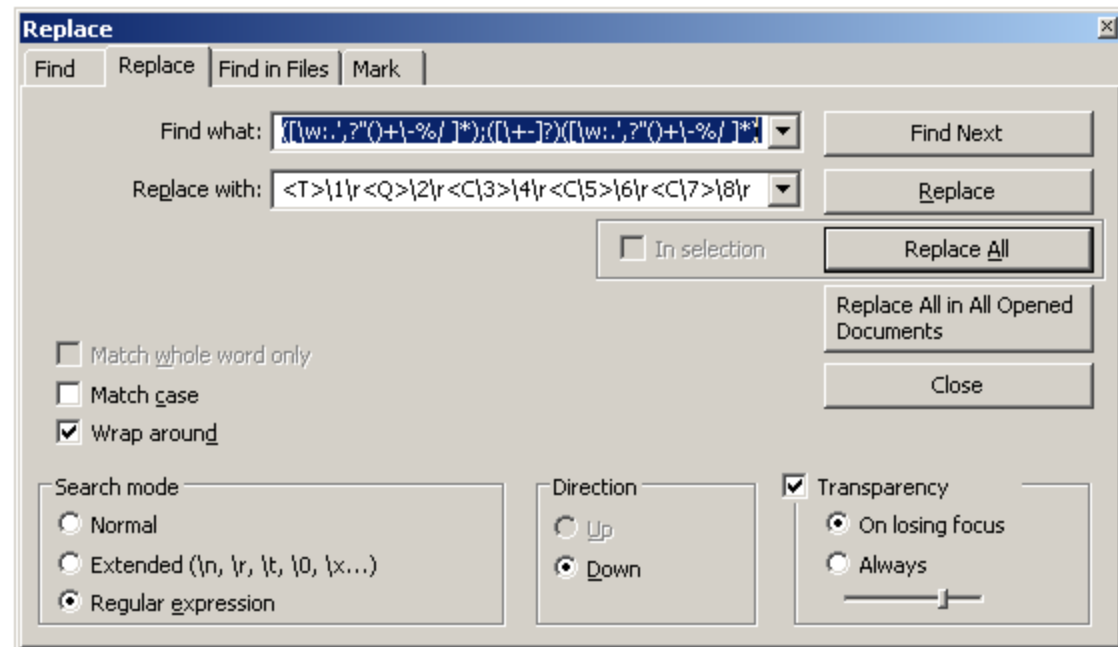
► Search for :



► Replace with :

<T>\1\r<Q>\2\r<C\3>\4\r<C\5>\6\r<C\7>\8\r

1 Question Type;Question Wording;Answer A;Answer B;Answer C;Answer D
 2 <T>MR
 3 <Q>En Europe, un déposant ayant de l'argent sur un compte bancaire:
 4 <C+>court le risque de perdre ses fonds si la banque fait faillite.
 5 <C->est protégé par le système de garantie des dépôts quels que soient les montants déposés.
 6 <C+>est protégé par le système de garantie des dépôts à hauteur de EUR 20 000.
 7
 8 <T>MC
 9 <Q>En France, la garantie bancaire prend la forme:
 10 <C->d'une adhésion au Fonds de garantie des dépôts français, uniquement réservé aux ressortissan
 11 <C+>d'une adhésion au Fonds de garantie des dépôts français, ouvert à tous les déposants, quels
 12 <C->d'une adhésion au Fonds de garantie européen.
 13
 14



Conclusion

▶ Regexp

- Is a handy solution for Saving time preparing ASCII Import
- Doesn't require any training for questions writers
- Is great when dealing with large amount of questions
- Is best if questions are mostly MC, MR and YN
- Is also best if scoring is 1, 0, -1 (otherwise you need to tweak scoring within Authoring Manager)

References

▶ Online resources

- <http://www.regular-expressions.info/>
- Regular Expression [article on Wikipedia](#)

▶ E-book

- [A Guide to Using Regular Expressions in PHP 5 \[Kindle Edition\]](#)
 - Although the book is about PHP, the first part is a reference guide to be referred to (and it is only \$6.89).

▶ Text editors

- [Notepad++](#)
- [EditPad Pro](#)

References

▶ iOS Apps

- [Regular Expression](#) (Universal, 0.79€)
- [Regexecution](#) (iPad, Free)
- [Regular Expression Pocket Reference](#) (O'Reilly e-book, iPhone, 2.99€)

▶ Android Apps

- [Regular Expression Pocket Reference](#) (O'Reilly e-book, 2.09€)
- [RegEx Cheatsheet](#) (Free)
- [Regex Dictionary](#) (1.14€)

Contact

- ▶ Should you have any question or information request, feel free to drop me an e-mail : furter@em-lyon.com