

## 5. QUALIFICATIONS

### Using the Question Mark

#### Introduction

We are now moving onto looking at “qualification metacharacters” which are metacharacters that allow us to specify the number of instances of a single character or a grouping of characters in a Regular Expression. Our first qualification metacharacter is the question mark (“?”). **The question mark metacharacter is matched if there is zero or one instances of the preceding character (or grouping) in a Regular Expression.**

#### Why Use the Question Mark?

If we were searching for the word “colour”, but we are dealing with a collection of documents some of which are from the USA and some from Europe; we know that in the USA we will be looking for the word “color”, so we can create a Regular Expression to describe this as follows:

```
Regex_Pattern = "colour|color"
```

But a more compact way of saying the same thing is to state that every time we find the word “colour”, the “u” may appear either zero or one times in that word. So, if it appears zero times, then the word is “color” and if it appears one time, then the word is “colour”. This can be stated using the Question Mark as follows:

```
Regex_Pattern = "colou?r"
```

So it’s a more compact way of matching with either “colour” and “color”, as the character preceding the Question Mark (“u”) can appear zero or one times.

#### Matching with Multiple Characters

If we were looking for documents about Regular Expressions, and we were looking for the phrase “Regular Expression” or “RegEx” in them, we could do it as follows:

```
Regex_Pattern = "Regular Expression|RegEx"
```

We can achieve the same result using the Question Mark metacharacter in conjunction with the round brackets as follows:

```
Regex_Pattern = "Reg(ular)?Ex(pression)?"
```

So it will match with either “Regular Expression” or “RegEx”, as the two strings preceding the question marks (“ular “ and “pression”) can appear zero or one times.

#### Matching the Question Mark Character

If we are searching for the actual question mark character (not the metacharacter), then mathematically we represent it as follows: \?

However, most programming languages prefer we state the Question Mark as:

```
Regex_Pattern = "\\?"
```

And this will allow us to locate the question mark character in a String.

#RegExThursday © Damian Gordon