



XML Schema Regular Expressions

Overview

A Regular Expression (also called RegEx) is the name given to a sequence of special characters used to describe a range of text which is permissible in an XML document constrained by an XML Schema. RegEx is very flexible: the range of text referred to may be a specific character or a specific String (series of characters); a language Character set; a subset such as Letters, Numbers, Whitespace or Upper and/or Lower case. Wildcard symbols allow the author a degree of flexibility within the target phrase. The choice is infinite. RegEx adds an extremely powerful, if somewhat complicated for beginners, means of describing the permitted content of a document. TelFormFactory aims to facilitate the use of RegEx by validating the syntax and allowing the testing of sample data against the expression. Future versions of TelFormFactory in preparation will provide for the progressive building and testing of RegEx by choosing RegEx atoms (smallest RegEx components) to form RegEx branches (RegEx sub-expressions) which make up the final expression.

Background to Regular Expressions

RegEx is a significant component of the Perl programming language, and some elements appear in Unix-like shell scripting. However there are important differences and the W3C XML Schema Definition Language (part 2 Datatypes) description of its version of Regular Expressions is in the process of being re-drafted. TelFormFactory references W3C Working Draft 3 December 2009 Appendix G of for the specification. However version 2.10.0 of Xerces from the Apache Software Foundation is used to implement RegEx parsing and validation and results are subject to the interpretation and implementation thereof. As this sub-application matures - which could almost be an application in its own right, other APIs will be reviewed.

XML Schema validation and Client validation

The 'legalese' nature of Schema validation does not lend itself to easy interpretation of errors by the average TelForm user - another layer of error handling should be provided by the client. Current versions of TelForm clients do not support local schema validation but hopefully the client does give a reasonable indication of where errors lie. Future versions of PC-based clients will do both. However no account of RegEx errors is made by TelForm Clients at this time. RegEx are stored in the TelFormFactory document and are recovered for XML Schema and Client revision.

TelFormFactory RegEx Editor

Some time has been spent developing an integral Regular Expression Editor. This has now been suspended following the 'discovery' of **RegexBuddy** from 'Just Great Software'

<http://www.regexbuddy.com>

The RegexBuddy web-site contains a lot of help regarding Regular Expressions generally which may be used to produce working expressions for your TelForm, and has a lot of everyday and esoteric ready-made examples. RegexBuddy is very easy to use and is very good value. RegexBuddy caters for many 'flavours' of RegEx so be sure to select **XML Schema**. RegexBuddy is an **exe** application but runs very well under Linux and Wine.

Note: The output of RegexBuddy may still output **\b** indicating **word boundary**. This is not acceptable in XML Schema RegEx syntax and will be stripped by TelFormFactory.

also has 'RegexMagic' which takes a slightly different approach and may also be useful.

Copyright © terry-comms 2003-2011 version-20120511 : 2035 |