-This document will illustrate the power of 25 July regular expressions in a Microsoft Word document. With regular expressions, one can detect dates such as 02/10/2006 or 07/30/2002 with no problem. Sequences of the same format which do not represent dates such as 08/32/2008 will not be detected. 4 April Through the use of assemblies to validate regular expressions we can reject sequences that many simple implementations of regular expressions will miss, such as 02/30/2001 or 02/29/2003, since there was no leap year in 2003. But 02/29/2004 and 02/29/2008 will be detected. Dates in the form 12-22-1977 are also detected as well as dates with an En Dash 03–27–1967.

If the date is in the European format, (day-month-year), such as 17/08/2001, the US date Regular Expression will not detect it, but the European expression will. And the same holds vice-versa. Dates which could be interpreted either way, such as 04/09/1956 will be detected by both, therefore one needs to know the origin of the document and choose the corresponding Regular Expression.

Social security numbers are also ideal for Regular Expressions. A typical SSN number would be 552-57-1234. The first 3 numbers can be any value from 001- 672. Therefore 000-33-8888 is invalid and so is 777-33-9876. One can also search through tables as shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Birth Date | SSN | Email Address |
| John Smith | 04/12/1981 | 552-76-8765 | J.Smith@gmail.com |
| Carol E. Jones | 12/08/1966 | 217-50-2565 | CJones@ucc.edu |
| Tom S. Green | 07-11-1958 | 477-34-0986 | Tom.Green@alpha-7.net |
| Earl Jefferson | 11–04–1986 | 300-45-9876 | EarlJ@RockRoll.com |
|  |  |  |  |

Phone numbers work similar to social security numbers. A typical number (818)766-1449 has the area code in parenthesis followed by the 7-digit number. Many times there is no parenthesis at all such as 212 646-1876. Sometimes the number 1 precedes the number like 1-(213)-853-1212. It is possible to just have spaces between numbers like 1 203 889 7272. In many instances, the dash may be an En dash as in (714)–888–7654.

URLs can be a bit tricky. There are the common URL such as if you want to find out how the .NET Split Container works <http://msdn.microsoft.com/en-us/library/system.windows.forms.splitcontainer.aspx> . Of course, simple URLS such as <http://www.yahoo.com> or Google advanced search, <http://www.google.com/advanced_search?hl=en> are all detected. There are also ftp sites: <ftp://63.229.95.49/> . Note that the numeric part of the ftp address will also be detected as an IP address. If we did not want this we could have modified the regular expression for the IP address to begin and end on a word boundary (\b).