Bug 710 (Priority: High)

Summary:

StringEscapeUtils.escapeXML() can't process UTF-16 supplementary characters

Description:

What steps will reproduce the problem?

Supplementary characters in UTF-16 are those whose code points are above 0xffff, that is, require more than 1 Java char to be encoded, as explained here: http://java.sun.com/developer/technicalArticles/Intl/Supplementary/

Additional Information

Currently, StringEscapeUtils.escapeXML() isn't aware of this coding scheme and treats each char as one character, which is not always right.

Low

Bug 710 (Priority: Low)

Summary:

StringIndexOutOfBoundsException when calling unescapeHtml4("&#03")

Description:

What steps will reproduce the problem?

When calling unescapeHtml4() on the String "&#03" (or any String that contains these characters) an Exception is thrown:

Exception in thread "main" java.lang.StringIndexOutOfBoundsException: String index out of range: 4

at java.lang.String.charAt(String.java:686)

at org.apache.commons.lang3.text.translate.NumericEntityUnescaper.translate(NumericEntityUnescaper.java:49)

at org.apache.commons.lang3.text.translate.AggregateTranslator.translate(AggregateTranslator.java:53)

at org.apache.commons.lang3.text.translate.CharSequenceTranslator.translate(CharSequenceTranslator.java:88)

at org.apache.commons.lang3.text.translate.CharSequenceTranslator.translate(CharSequenceTranslator.java:60)

at org.apache.commons.lang3.StringEscapeUtils.unescapeHtml4(StringEscapeUtils.java:351)