## 通过javascript进行UTF-8编码的实现方法

javascript的字符集：

javascript程序是使用Unicode字符集编写的。Unicode是ASCII和Latin-1的超集，并支持地球上几乎所有的语言。ECMAScript3要求JavaScript必须支持Unicode2.1及后续版本，ECMAScript5则要求支持Unicode3及后续版本。所以，我们编写出来的

javascript程序，都是使用Unicode编码的。

UTF-8

UTF-8（UTF8-bit Unicode Transformation Format）是一种针对Unicode的可变长度字符编码，也是一种前缀码。

它可以用来表示Unicode标准中的任何字符，且其编码中的第一个字节仍与ASCII兼容，这使得原来处理ASCII字符的软件无须或只须做少部分修改，即可继续使用。因此，它逐渐成为电子邮件、网页及其他存储或发送文字的应用中，优先采用的编码。

目前大部分的网站，都是使用的UTF-8编码。

将javascript生成的Unicode编码字符串转为UTF-8编码的字符串

如标题所说的应用场景十分常见，例如发送一段二进制到服务器时，服务器规定该二进制内容的编码必须为UTF-8。这种情况下，我们必须就要通过程序将javascript的Unicode字符串转为UTF-8编码的字符串。

转换方法

转换之前我们必须了解Unicode的编码结构是固定的。

不信可以试一试 String 的 charCodeAt 这个方法，看看返回的 charCode 占几个字节。

•英文占1个字符，汉字占2个字符

然而，UTF-8的编码结构长度是根据某单个字符的大小来决定长度有多少。

下面为单个字符的大小占用几个字节。单个unicode字符编码之后的最大长度为6个字节。

•1个字节：Unicode码为0 - 127

•2个字节：Unicode码为128 - 2047

•3个字节：Unicode码为2048 - 0xFFFF

•4个字节：Unicode码为65536 - 0x1FFFFF

•5个字节：Unicode码为0x200000 - 0x3FFFFFF

•6个字节：Unicode码为0x4000000 - 0x7FFFFFFF

Js的Unicode编码转换方法

charCodeAt() 方法可返回指定位置的字符的 Unicode 编码。这个返回值是 0 - 65535 之间的整数。

方法 charCodeAt() 与 charAt() 方法执行的操作相似，只不过前者返回的是位于指定位置的字符的编码，而后者返回的是字符子串。

语法：

stringObject.charCodeAt(index)

index 必需。表示字符串中某个位置的数字，即字符在字符串中的下标。

fromCharCode() 可接受一个指定的 Unicode 值，然后返回一个字符串。

String.fromCharCode(numX,numX,...,numX)

参数 描述

numX 必需。一个或多个 Unicode 值，即要创建的字符串中的字符的 Unicode 编码。

提示和注释

注释：该方法是 String 的静态方法，字符串中的每个字符都由单独的数字 Unicode 编码指定。

它不能作为您已创建的 String 对象的方法来使用。因此它的语法应该是 String.fromCharCode()，而不是 myStringObject.fromCharCode()。

网址链接：http://www.jb51.net/article/87418.htm