Subject: A new function to Web Package Unicode-Escape-Javascript -> Unicode
Posted by sergeynikitin on Wed, 20 Jul 2011 07:28:31 GMT
View Forum Message <> Reply to Message

I propose to include in the package a new feature WEB Unicode-Escape-Javascript -> Unicode.

For international characters in Javascript is used for special encoding non-Latin characters. Looks like: \ u0410 \ u0422 \ u0417 ....
For converting this encoding to Unicode, I needed a new function. I propose that its AEs in the package a new feature WEB Unicode-Escape-Javascript -> Unicode.

For international characters in Javascript is used for special encoding non-Latin characters. Looks like: \ u0410 \ u0422 \ u0417 ....
For converting this encoding to Unicode, I needed a new function. I have not found it, so I wrote in haste.

Maybe someone else will need.

I also want to note that some of the operations inside the function can be optimized. For example the plural multiplying by 16 can be replaced by bit shift.

```
String Javascript2Unicode(String s) {
 String res ;
 RegExp reg("\\\\u([0-9a-f]{4})",RegExp::MULTILINE);

 reg.Clear();

 int i_start,i_end,i_old=0;
 while (reg.GlobalMatch(s)) {
  reg.GetMatchPos(0,i_start,i_end);
  res << s.Mid(i_old,i_start-i_old-2);
  wchar wc[2] = {0,0};
  WString ws;
  String m = (String)reg[0];
  wc[0] = StrInt(m.Mid(0,1))*16*16*16
    +StrInt(m.Mid(1,1))*16*16
    +StrInt(m.Mid(2,1))*16
    +StrInt(m.Mid(3,1));

  ws = wc;
  String ss = ws.ToString();
  res << ss;
  i_old = i_end;
 }
 return res;
}
```

PS

Maybe there is already something like that? And I wasted 4 hours? Advise me if you know.

Subject: Re: A new function to Web Package Unicode-Escape-Javascript -> Unicode
Posted by sergeynikitin on Wed, 20 Jul 2011 17:23:45 GMT
View Forum Message <> Reply to Message

Sender Ghost wrote PM on Wed, 20 July 2011 19:34Hello, Sergey.

I used the function with the same behaviour. It is called String UrlDecode(const String& s) from Web package.

The following string

%u0410%u0422%u0417

converts to

Unfortunately, there are characters within the string like "\" and "%". Therefore, simple replacement of "\" to "%" will not work. Will either have to have two different functions, or to make an extra parameter to the function UrlDecode.

In any case, thank you!

Subject: Re: A new function to Web Package Unicode-Escape-Javascript -> Unicode
Posted by sergeynikitin on Wed, 20 Jul 2011 19:01:55 GMT
View Forum Message <> Reply to Message

I propose to extend function UrlDecode with form \uXXXX additionally to %uXXXX enciding. Form \uXXXX is reserved for old browsers and it is standard like form %uXXXX.

Interesting Online Encodings converter:
http://rishida.net/tools/conversion/

pls apply patch to Function:

```
String UrlDecode(const char *b, const char *e)
{
	StringBuffer out;
	byte d1, d2, d3, d4;
	for(const char *p = b; p < e; p++)
```

```
            if(*p == '+')
                out.Cat(' ');
            else if(*p == '%' && (d1 = ctoi(p[1])) < 16 && (d2 = ctoi(p[2])) < 16) {
                out.Cat(d1 * 16 + d2);
                p += 2;
            }
            else if(*p == '%' && (p[1] == 'u' || p[1] == 'U')
            && (d1 = ctoi(p[2])) < 16 && (d2 = ctoi(p[3])) < 16
            && (d3 = ctoi(p[4])) < 16 && (d4 = ctoi(p[5])) < 16) {
                out.Cat(WString((d1 << 12) | (d2 << 8) | (d3 << 4) | d4, 1).ToString());
                p += 5;
            }
            else
                out.Cat(*p);
        return out;
}
```

I propose change like this:

```
String UrlDecode(const char *b, const char *e)
{
        StringBuffer out;
        byte d1, d2, d3, d4;
        for(const char *p = b; p < e; p++)
            if(*p == '+')
                out.Cat(' ');
            else if(*p == '%' && (d1 = ctoi(p[1])) < 16 && (d2 = ctoi(p[2])) < 16) {
                out.Cat(d1 * 16 + d2);
                p += 2;
            }
            else if((*p == '%' || *p == '\') && (p[1] == 'u' || p[1] == 'U') // <-This line changed
            && (d1 = ctoi(p[2])) < 16 && (d2 = ctoi(p[3])) < 16
            && (d3 = ctoi(p[4])) < 16 && (d4 = ctoi(p[5])) < 16) {
                out.Cat(WString((d1 << 12) | (d2 << 8) | (d3 << 4) | d4, 1).ToString());
                p += 5;
            }
            else
                out.Cat(*p);
        return out;
}
```

Subject: Re: A new function to Web Package Unicode-Escape-Javascript -> Unicode
Posted by sergeynikitin on Wed, 20 Jul 2011 19:13:01 GMT
View Forum Message <> Reply to Message

This function have name UnEscape in Javascript.

Therefore maybe We must make different functions UrlDecode and UnEscape?
Because logic of Javascript's Unescape a bit different: It don't replace "+" with "space".

Subject: Re: A new function to Web Package Unicode-Escape-Javascript -> Unicode
Posted by Sender Ghost on Wed, 20 Jul 2011 20:22:15 GMT
View Forum Message <> Reply to Message

sergeynikitin wrote on Wed, 20 July 2011 21:13
Because logic of Javascript's Unescape a bit different: It don't replace "+" with "space".
Because UrlEncode/UrlDecode functions (also, from the meaning of function names) used for URL(s). I also think, different functions needed for content. May be, with general implementation.

From UrlDecode you could see, as you said, function optimizations and shifts of bits, instead of using regular expressions.

References:
- Some implementation of UrlEncode/UrlDecode.
- About UrlEncode on Wikipedia and why " " (space) converted to "+" instead of "%20" on early stage(s).
