

2. 词法和类型

1. javascript支持的字符集

通识理解：

javascript中，支持以unicode的字符集。

在U+FFFF以内的为bmp，即基本平面的字符集，一般使用String.fromCharCode(码值) 可以获取到对应的字符（或者使用String.fromCodePoint(码值)）

如果超出了基本平面的符合集，我们应该使用String.fromCodePoint(码值)来获取对应的字符集。

javascript明面上是说支持所有的unicode集合，但是是实践证明，最好是使用在基本平面范围内的字符集的兼容性能是最好的。

如果超出了基本平面的字符集合，我们可以使用\u转义来获取对应字符。使用\u转义可以定义变量或者在字符串中转化为对应的字符集。

词法分析

WhiteSpace 空白符，用来分隔token或者代码锁进等操作。

```
<TAB> -> "\t" 或者 "\u0009" (默认是占4为单位做制表操作)
<VT> -> "\v" 或者 "\u0011"
<FF> -> "\f" 或者 "\u0010"
<SP> ->直接空格 或者 "\u0020"
<NBSP> ->&nbsp; (是一个实体)代表不分词的空格，在使用的过程中如果在两个词之间使用&nbsp;连接，则这两个词不会换行，一定会连在一起。
<ZWNBSP> ->零宽空格，虽然是空格，但是他不形成间隙。
<USP>->表示unicode中的空格，我们可以去
(https://www.fileformat.info/info/unicode/category) 查找。
```

LineTerminator

```
<LF> -> "\n" 或者 "\u000A"
```

<LF> -> \n 或者 \0010
<CR> -> 回车(\r)
<LS>
<PS>
最好选用\r或者\n

Comment

```
//  
  
/*  
  
*/
```

Token

IdentifierName: (在进行词法分析的时候, 首先验证是否是标识符, 因为都是以字符、\$、_等作为第一个字符的开始)

Keyword

keyword

Identifier:

变量 (不可以和关键字名相同)

属性 (可以和关键字同名)

Future reserved Keywords:enum等

Punctuator(符号)

Literal

Number :

DecimalLiteral: 0、0.、.2、1e3

BinaryIntegerLiteral : 0b1111

OctalIntegerLiteral :0o10

HexIntegerLiteral:0xff

最佳实践: 浮点比较:Math.abs(xxx - yyy) <= Number.EPSILON

String

双引号“”

单引号’

里面可以包含的字符:

1. 单双引号中可以包含除去"、'、\、回车换行之外的单一源字符
2. \ + 转义字符, 比如n、b、r、f、0等
3. \u+4位或者\x+2位

反引号`

注意: 一般使用utf8、utf16进行解码操作, 在js中默认是使用utf16进行编码

Boolean

true

false

Null:

null