

Unicode 字符集

SourceCharacter ::

702

any Unicode code point

<https://www.fileformat.info/info/unicode/>
<https://home.unicode.org/> 官网

<https://www.cnblogs.com/leesf456/p/5317574.html>

ASCII 字符集 0 - 128

U+0007 BELL - 没响?
U+000A 回车
U+000C 换行
U+0020 SPACE

CJK 中日韩
https://www.fileformat.info/info/unicode/block/cjk_unified_ideographs/index.htm
中文字符匹配

BMP - 4 个十六进制位表示，基本字符平面，兼容性好

超出 ASCII 字符集会涉及文件编码问题？

```
'厉害'.codePointAt(0).toString(16);    // 5389
'厉害'.codePointAt(1).toString(16);    // 5bb3

var \u5389\u5bb3 = 1;
console.log(厉害);
```

最佳实践：js 限制在 **ascii**/BMP 内，最好不要超

InputElement

- WhiteSpace 空格
 - <https://www.fileformat.info/info/unicode/category/Zs/index.htm>
 - ZWSP 不在 空格 分组里，\uFEFF
 - TAB, \u0009
 - VT
 - FF, \u000c
 - **SP, \u0020**
 - NBSP, 分词不换行，\u00a0
 - ZWNBSP, \uFEFF, 零宽空格, BOM, bit order mask
 - USP, Unicode Space
- LineTerminator
 - **U+000A, \n**
 - U+000D, \r
- Comment
 - 单行, //
 - 多行, /* ... */, *不可以用\u 转码，不可嵌套
- Token, 一切有效的
 - Punctuator, 符号
 - IdentifierName
 - Keywords, 关键字
 - Identifier, 标识符
 - 变量名，不能与关键字重合
 - 属性，可以和关键字重合
 - Future Reserved Keywords: enum
 - Literal, 直接量
 - NumericLiteral
 - StringLiteral
 - Template

Types

- Number
 - IEEE 754 Double Float
 - Sign(1)
 - Exponent(11)
 - Fraction(52)
 - Grammar, parseInt
 - DecimalLiteral
 - 0
 - 0.
 - .2
 - 1e3
 - 97 .toString, **97.toString**词法上优先原则。空格（无意义的字符）的分割使得语法成立和美观
 - BinaryIntegerLiteral
 - 0b111
 - OctalIntegerLiteral
 - 0o11
 - HexIntegaerLiteral
 - 0xFF|
 - **写一个正则，匹配 Number 所有直接量(11.8.3 Numeric Literals)**
 - Practice
 - Safe Integar
 - Number.MAX_SAFE_INTEGER
 - Float Compare
 - 0.1 + 0.2 - 0.3 <= Number.EPSILON
- String
 - Character
 - Code Point
 - Encoding

字符集：

- ASCII
- Unicode
- UCS => BMP, U+0000 - U+FFFF
- GB, ASCII+中文
 - GB2312
 - GBK(GB13000)
 - GB18030
- ISO-8859
- BIG5

每种字符集都可以对应多种编码方式，码点定了，如何存（编码）？
UTF-8 存储: 01100001, 01100010
UTF-16: 00000000 01100001, 00000000 01100010

js, 基本UTF-16，在内存里存储，一般不承认是 1 个字符

charCodeAt/FromCharCode
codePointAt/FromCodePoint

Homework

写一个 **UTF-8 Encoding** 的函数

```
function UTF8_Encoding (string) {
  // ...return new Buffer();
}
```

Grammar

- ''
- ""
- template

\b \f \n \r \t \v

写一个正则表达式，匹配所有的字符串直接量，单引号和双引号

- Boolean
- Null
- Undefined
- Object
- Symbol

正则直接量
//

<https://regexpr.com/>