



链滴

# JS Emoji 过滤

作者: [Vanessa](#)

原文链接: <https://ld246.com/article/1489758460114>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

想用的直接看最后一行就好

## 基础知识

不得不承认阮蜀黍的文笔通俗易懂

- [字符编码笔记: ASCII, Unicode和UTF-8](#)
- [Unicode与JavaScript详解](#)

## Emoji Unicode Tables

1. Emoticons ( 1F601 - 1F64F )
2. Dingbats ( 2702 - 27B0 )
3. Transport and map symbols ( 1F680 - 1F6C0 )
4. Enclosed characters ( 24C2 - 1F251 )
5. Uncategorized
6. Additional
  1. Additional emoticons ( 1F600 - 1F636 )
  2. Additional transport and map symbols ( 1F681 - 1F6C5 )
  3. Other additional symbols ( 1F30D - 1F567 )

## 转换?

### 只怪 JavaScript 你诞生的太早

JavaScript 只识别 UCS-2, 对于 UTF-16 的 4 个编码来说, 只能识别为 2 个单独的字符

### 友好的 ES6

ES6 对 Unicode 做了大幅增强支持

```
'\u{1f604}' === '\ud83d\ude04' // true
```

### 笨笨的我

然而我并不知道 `'\u{1f604}'.replace(/\u{1f604}/g, "")` 这个的正则怎么写, So, 只能把带有辅助平面字符的 unicode 转换为 2 个单独的字符

```
var unicode2UTF16 = function (unicode) {  
  const h = Math.floor((unicode - 0x10000) / 0x400) + 0xD800;  
  const l = (unicode - 0x10000) % 0x400 + 0xDC00;  
  return '\\u' + h.toString(16) + '\\u' + l.toString(16);};
```

### map

细细看下来, Emoji Unicode Tables 给的范围不按顺序出牌。Uncategorized 直接乱了爸妈都不认, 瞎眼整理:

```
\ud83d[\udc00-\udec5]
```

- 0x1F601(\ud83d\ude01) - 0x1F64F(\ud83d\ude4f)
- 0x1F600(\ud83d\ude00) - 0x1F636(\ud83d\ude36)
- 0x1F680(\ud83d\ude80) - 0x1F6C0(\ud83d\udec0)
- 0x1F681(\ud83d\ude81) - 0x1F6C5(\ud83d\udec5)
- 0x1F400(\ud83d\udc00) - 0x1F567(\ud83d\udd67)

```
\ud83c[\udd70-\udf4e]
```

- 0x1F30D(\ud83c\udf0d) - 0x1F3E4(\ud83c\udfe4)
- 0x1F170(\ud83c\udd70) - 0x1F251(\ud83c\ude51)

```
XXXXYYYYOOOO
```

## 结果

如果要弄太花费精力, 还是网上搜了下, 选了个大的(最后一个)。。。

- \ud83d[\udc00-\udec5]\ud83c[\udd70-\udf4e]
- \ud83c[\udf00-\udfff]\ud83d[\udc00-\ude4f]
- \ud83c[\udc00-\udfff]\ud83d[\udc00-\udfff][\u2000-\u2fff]

```
return str.replace(/\ud83c[\udc00-\udfff]\ud83d[\udc00-\udfff][\u2000-\u2fff]/g, "")
```

## 881 天后更新

使用上速正则及 google 到的正则都无法对 <https://github.com/b3log/vditor/blob/master/src/ts/moji/allEmoji.ts> 中的 1503 个 emoji 进行完全匹配, 因此重写如下:

- 1446 个 emoji 分别为 2、3、4、5、6、8、11 个字符, 正则如下:

```
/([\u{1F300}-\u{1F5FF}][\u{2000}-\u{206F}][\u{2700}-\u{27BF}]|([\u{1F900}-\u{1F9FF}][\u{1F300}-\u{1F5FF}][\u{1F680}-\u{1F6FF}][\u{1F600}-\u{1F64F}])|[\u{2000}-\u{206F}][\u{2600}-\u{26FF}][\u{1F300}-\u{1F5FF}][\u{1F100}-\u{1F1FF}][\u{1F600}-\u{1F64F}][\u{1F680}-\u{1F6FF}][\u{1F200}-\u{1F2FF}][\u{1F900}-\u{1F9FF}][\u{1F000}-\u{1F02F}][\u{FE00}-\u{FE0F}][\u{1F0A0}-\u{1FFF}][\u{0000}-\u{007F}][\u{20D0}-\u{20FF}][\u{0000}-\u{007F}][\u{FE00}-\u{FE0F}][\u{20D0}-\u{20FF}])$/u)
```

- 44 个 emoji 为 1 个字符, 正则如下:

```
/[\u{2100}-\u{214F}][\u{2000}-\u{206F}][\u{2460}-\u{24FF}][\u{0080}-\u{00FF}][\u{2700}-\u{27BF}][\u{2600}-\u{26FF}][\u{2B00}-\u{2BFF}][\u{2300}-\u{23FF}][\u{25A0}-\u{25FF}][\u{3000}-\u{303F}][\u{3200}-\u{32FF}][\u{2900}-\u{297F}][\u{2190}-\u{21FF}]/u
```

- 剩余 13 个为图片