张东

zhangdong@tedu.cn

3样东西:

1. 书: 犀牛书

2. 上届笔记和思维导图:

上届笔记: 预习

思维导图: 备查

3. 公众号: 前端大全

谷歌小恐龙:

断网，打开谷歌浏览器，访问任何网址，看到下恐龙点空格

正课:

1. String

2. 正则表达式

1. String

什么是字符串: 由多个字符组成的只读数组

vs 数组:

相同: 1. 下标, 2. length, 3. for遍历, 4. slice

不同: 类型不同! API不通用

API: 别人写好的，咱们用现成的函数

特点: 所有字符串API都无权修改原字符串

只能返回新字符串

大小写转换:

str.toUpperCase()

str.toLowerCase()

何时: 不区分大小写时

比如: 验证码, 用户名, 电子邮件

获取指定位置的字符:

str.charAt(i) <=> str[i]

获取指定位置字符的unicode号

var unicode="字".charCodeAt()如果不写参数，默认为0

var unicode="xxx".charCodeAt(i)获得i位置的字符的unicode号

从unicode号，逆向转回字:

var char=String.fromCharCode(u);

选取子字符串:

str.substring(starti,endi+1) <==>str.slice(starti,endi+1)

差别: substring不支持负数参数！

str.substr(starti,n) 从starti开始选取n个

强调: n不用+1

共同特点: 省略第二参数，都表示到结尾！

查找关键词: 4种:

1. 查找一个固定的关键词出现的位置:

var i=str.indexOf("关键词"[,开始位置])

在str中指定"开始位置"之后，查找下一个"关键词"出现的位置下标i

如果没找到，就返回-1

查找最后一个关键词的位置:

var i=str.lastIndexOf("关键词");

问题: 一次只能查找一种关键词，只要关键词变化，就找不到

解决: 正则表达式

2. 判断是否包含敏感词(支持正则)

var i=str.search(/正则表达式/)

问题1: 没有第二个参数，不能设置开始位置，永远只能找第一个

问题2: 只能获得位置，不能获得内容

3. 查找所有敏感词的内容:

var kwords=str.match(/正则表达式/ig);

返回值: 包含多个敏感词的数组

如果找不到，返回null!

强调: 如果一个函数返回null！必须先验证，再使用！

问题1: 默认正则只匹配第一个敏感词就退出！

解决: 在第二个/后加g (global)

问题2: 正则表达式，默认区分大小写!

解决: 在第二个/后加i (ignore)

问题: 无法获得每个关键词的位置:

4. 即找所有关键词的内容，又找所有关键词的位置

regexp.exec()

替换: 将找到的敏感词替换为指定的新关键词

2种:

简单替换: 将所有敏感词都替换为统一的新内容

str=str.replace(/正则/ig, "新内容");

高级替换: 根据每个敏感词的不同，自动选择对应的新内容替换。

str=str.replace(/正则/ig, function(kw){

//kw: 自动获得当前找到的敏感词

return 根据kw的不同，返回对应新值

})

衍生: 删除: 将找到的敏感词替换为""

切割:

2. \*\*\*正则表达式:

什么是: 描述一类字符串中字符出现规律的规则

何时: 2种:

1. 用规则模糊查找多种敏感词时

2. 验证字符串的格式

语法规则:

1. 最简单的正则表达式: 就是关键词原文

2. 字符集:

什么是: 定义一位字符上的备选字符列表的集合

何时: 只要一位字符上有多种备选字时

如何: [备选字符列表]

简写: 如果备选字符集中部分字符连续，可用-省略中间字符

包括: 一位数字: [0-9]

一位小写字母:[a-z]

一位大写字母:[A-Z]

一位字母: [A-Za-z]

一位字母或数字: [0-9A-Za-z]

一位汉字: [\u4e00-\u9fa5]

特殊: 除了: [^xxx]

3. 预定义字符集:

一位数字: \d =>[0-9]

一位字母，数字或下划线: \w => [0-9A-Za-z\_]

一位空字符: \s => 一切空字符: 空格 Tab 换行...

通配符: .

问题: 预定义字符集只能定义一位字符的内容，无法灵活定义字符出现的次数

解决:

4. 量词:

什么是: 专门定义一个字符集出现次数的规则

何时: 只要定义字符集出现的次数

如何: 字符集量词

默认: 量词只修饰它紧邻的前一个字符集

包含2大类:

1. 有明确数量边界: 3种:

{m,n} 至少m次，最多n次

{m,} 至少m次，多了不限

{m} 必须m次

2. 没有明确数量边界: 3种:

\* 可有可无，多了不限

+ 至少一次，多了不限

? 可有可无，最多一次

问题: 量词默认只修饰相邻的一个字符集

解决:

5. 选择和分组:

分组: 将多个字符集用()包裹在一起成为一组

放在()后的量词，修饰整个分组

选择: 或 规则1|规则2

两个规则匹配其一即可

中国大陆手机号:

+86或0086 可有可无,最多一次 (+86|0086)?

至少一个空字符 \s+

1

3,4,5,7,8中选一个 [34578]

9位数字 \d{9}

(+86|0086)?\s+1[34578]\d{9}

身份证号:

15位数字 2位数字 数字或X或x

后三位整体可有可无，最多一次

\d{15}(\d{2}[0-9Xx])?

微信 weixin wx

(微|w(ei)?)\s\*(信|x(in)?)

6. 匹配特殊位置: 3个位置:

1. 字符串开头: ^

何时: 只要仅匹配字符串开头的内容时

比如: 开头的空字符: ^\s+

2. 字符串结尾: $

何时: 只要仅匹配字符串结尾的内容时

比如: 结尾的空字符: \s+$

再比如: 同时匹配开头或结尾的空字符

错误: ^\s+$ 从头到尾必须只能是\s

正确: ^\s+|\s+$

正则中的|优先级最低,将整个正则一分为2，前后分别匹配，毫无关系

3. 单词边界: \b 空格，标点，字符串开头和结尾

何时: 如果选择一个单词时