

极客大学前端进阶训练营

程劭非 (winter)

前手机淘宝前端负责人

有限状态机处理字符串

有限状态机

- 每一个状态都是一个机器
 - 在每一个机器里，我们可以做计算、存储、输出.....
 - 所有的这些机器接受的输入是一致的
 - 状态机的每一个机器本身没有状态，如果我们用函数来表示的话，它应该是纯函数（无副作用）
- 每一个机器知道下一个状态
 - 每个机器都有确定的下一个状态（Moore）
 - 每个机器根据输入决定下一个状态（Mealy）

使用有限状态机处理字符串

- 在一个字符串中，找到字符“a”

使用有限状态机处理字符串

- 在一个字符串中，找到字符 “ab”

使用有限状态机处理字符串

- 在一个字符串中，找到字符 “abcdef”

JS中的有限状态机 (Mealy)

```
//每个函数是一个状态
function state(input) //函数参数就是输入
{
    //在函数中，可以自由地编写代码，处理每个状态的逻辑
    return next;//返回值作为下一个状态
}

//////////以下是调用//////////
while(input) {
    //获取输入
    state = state(input); //把状态机的返回值作为下一个状态
}
```

额外内容

- 我们如何用状态机处理诸如 “abcbabx” 这样的字符串？
- 作业：使用状态机完成 “abababx” 的处理。
- 可选作业：我们如何用状态机处理完全未知的pattern？
 - 参考资料：字符串KMP算法
https://en.wikipedia.org/wiki/Knuth%E2%80%93Morris%E2%80%93Pratt_algorithm

HTML的解析

浏览器



第一步——拆分文件

第一步总结

- 为了方便文件管理，我们把parser单独拆到文件中
- parser接受HTML文本作为参数，返回一颗DOM树

第二步——创建状态机

<https://html.spec.whatwg.org/multipage/parsing.html#before-attribute-name-state>

第二步总结

- 我们用FSM来实现HTML的分析
- 在HTML标准中，已经规定了HTML的状态
- Toy-Browser只挑选其中一部分状态，完成一个最简版本

第三步——解析标签

第三步总结

- 主要的标签有：开始标签，结束标签和自封闭标签
- 在这一步我们暂时忽略属性

第四步——创建元素

第四步总结

- 在状态机中，除了状态迁移，我们还会要加入业务逻辑
- 我们在标签结束状态提交标签token

第五步——处理属性

第五步总结

- 属性值分为单引号、双引号、无引号三种写法，因此需要较多状态处理
- 处理属性的方式跟标签类似
- 属性结束时，我们把属性加到标签Token上

第六步——构建DOM树

第六步总结

- 从标签构建DOM树的基本技巧是使用栈
- 遇到开始标签时创建元素并入栈，遇到结束标签时出栈
- 自封闭节点可视为入栈后立刻出栈
- 任何元素的父元素是它入栈前的栈顶

第七步——文本节点

第七步总结

- 文本节点与自封闭标签处理类似
- 多个文本节点需要合并

作业：跟上课堂内容，完成DOM树构建

THANKS! |  极客大学