

上海交通大学

编译原理与技术 lab2 文档

学生姓名：_____ 杨景凯 _____

学 号：_____ 520021910550 _____

2022 年 09 月 28 日

目录

1	处理方式	3
1.1	对注释的处理	3
1.2	对字符串的处理	3
1.3	对错误的处理	4
1.4	对末尾的处理	4
2	其他特色	4

1 处理方式

1.1 对注释的处理

使用 Start condition, 当匹配到 `/*` 时, 开始 *COMMENT*。在开始后, 首先令注释层级为 1。

- 如果遇到 `/*`, 那么注释层级加 1。
- 如果遇到 `*/`, 那么注释层级减 1。如果此时注释层级为 0, 那么就退出 *COMMENT*。
- 其他情况则简单跳过。

1.2 对字符串的处理

使用 Start condition, 当匹配到 `"` 时, 开始 *STR*。在开始后, 首先清空 *string_buf_*。

- 如果遇到 `\`, 且后面跟着一个或多个 `\n` 或 `\t`, 且最后有一个 `\`, 那么表示在 *STR* 内的下行拼接, 简单跳过即可。
- 如果遇到 `\`, 且后面跟着三个数字, 那么表示 ASCII 编码, 将其翻译成单个的 char, 并加入到 *string_buf_* 的末尾。
- 如果遇到 `\^`, 且后面跟着大写字母、`@`、`[`、`]`、`\`、`-` 或 `6`, 那么表示 ASCII 控制字符, 将其翻译成单个的 char, 并加入到 *string_buf_* 的末尾。
- 如果遇到 `\`, 那么将 matched 字符串设置为 *string_buf_*, 退出 *STR*。
- 如果遇到 `\`, 且不在上述情况内, 那么表示特殊字符, 如果后面是 `n` 或者 `t`, 那么将其翻译成单个字符 `\n` 或 `\t`, 其余情况则直接翻译成 `\` 后的字符, 并加入到 *string_buf_* 的末尾。
- 其他情况则直接将其加入到 *string_buf_* 的末尾。

1.3 对错误的处理

- 如果遇到其他任何不能识别的匹配模式，那么报错“illegal token”。
- 在 *COMMENT* 中, 如果未匹配到结束而到达末尾, 那么报错“unclosed comment”。
- 在 *STR* 中, 如果未匹配到结束而到达末尾, 那么报错“unclosed string”。

1.4 对末尾的处理

对末尾的处理已经在上节对错误的处理中提到。

2 其他特色

具有很多注释，体现了良好的代码风格。