# daily-code

主要完成C语言的预处理功能

C语言预处理主要针对如下关键字进行特定的动作

主要完成对源文件的扩展，#include #if 递归包含

|  |  |
| --- | --- |
| command | explain |
| #define | 宏定义 |
| #error | 用于调试，输出错误信息 |
| #include | 包含其他文件（递归包含） |
| #if | 条件编译 |
| #else |  |
| #elif |  |
| #endif |  |
| #ifdef |  |
| #ifudef |  |
| #undef | 取消宏定义 |
| #line | 改变预先定义的符号，\_\_LINE\_\_ , \_\_FILE\_\_ |
| #pragma | 设定编译器状态 |

词法分析过程:

字符分类：

字母(letter)： a-z, A-Z

数字(digit)： 0-9

空格(ws)： tab space

其他字符(other): 其他非以上三种字符

所有字符(all)： 表示所有字符

有限自动机定义

state，遇见的字符，next\_state

通过逐步定义所有state形成一个状态表，其中START表示起始状态，END表示终止状态，任何非START状态的第一条规则都是all字符，表示一个默认状态，然后对该状态可接受的字符定义。

策略是：当前状态下查看字符，如果非END状态，则”eat”该字符，进入下一个状态，

如果是END状态，则直接返回该字符的内部编码。字符有可能会向前查看来达到最佳匹配，比如 “==”比”=”更匹配（贪婪匹配），所以需要多向前查看一个字符来达到是否已经终结，用END\_FORWARD状态表示遇到字符后，直接”eat”然后返回一个内部编码

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 所有状态 | START , ID1,NUM1,NUM2,NUM3,DOT1,STR1,STR2,STR3,CC1,CC2,WS1 | | | | |
| 符号含义 | 起始状态 | 遇见字符 | 下一个状态 | END返回token类别 | 解释 |
| ID | START | letter | \_ | ID1 |  |  |
| ID1 | all | END | NAME | 标识符 |
| ID1 | letter| \_ |digit | ID1 |  |  |
| num | START | digit | NUM1 |  |  |
| START | ‘.’ | NUM2 |  | 浮点型 |
| NUM1 | all | END | NUMBER |  |
| NUM1 | digit | letter | ‘.’ | NUM1 |  | 遇见数字，字符，或者小数点儿，还是NUM1，至于letter是否是合法的由编译器判断 |
| NUM1 | ‘E’ | ’e’ | NUM3 |  | 科学记数法 |
| NUM2 | all | END | NUMBER | 由编译器判断合法值与否 |
| NUM2 | digit | NUM1 |  |  |
| NUM2 | . | DOT1 |  | 遇见两个小数点，有可能是 …运算符（函数定义成可变参数的运算符） |
| NUM3 | all | END | NUMBER |  |
| NUM3 | + | - | NUM1 |  | 正负号 |
| NUM3 | digit | letter | NUM1 |  | e后面可以直接接数字 |
| … | DOT1 | all | END |  | 错误符号 |
| DOT1 | . | END\_FORWARD | ELLIPS(省略号) | …运算符(函数可变参数) |
| string | START | “L” | STR1 |  | 宽字符，一个字符占两个字节 |
| START | “ | STR2 |  | 双引号 |
| STR1 | all | END | NAME | 标识符 |
| STR1 | letter | digit | \_ | ID1 |  |  |
| STR1 | “ | STR2 |  | 双引号 |
| STR1 | ‘ | CC1 |  | 单引号 |
| STR2 | all | STR2 |  | 引号内所有字符都属于字符串 |
| STR2 | “ | END | STRING | 遇见结尾双引号 |
| STR2 | \ | STR3 |  | 字符串换行符 |
| STR2 | \n | STR\_NL |  | 字符串不能跨行定义 |
| STR2 | EOF | STR\_EOF |  | 字符串定义错误 |
| STR3 | all | STR2 |  |  |
| STR3 | \n | STR\_NL |  |  |
| STR3 | EOF | STR\_EOF |  |  |
| character const | CC1 | all | CC1 |  | 单引号内的字符 |
| CC1 | ‘ | END | CHAR\_CONST | 语句结束单引号 |
| CC1 | \ | CC2 |  | 跨行定义 |
| CC1 | \n | STR\_NL |  |  |
| CC1 | EOF | STR\_EOF |  |  |
| CC2 | all | CC1 |  |  |
| CC2 | \n | STR\_NL |  |  |
| CC2 | EOF | STR\_EOF |  |  |
| 空白符 | START | 空格 | WS1 |  |  |
| START | 水平制表符 | WS1 |  |  |
| START | 垂直制表符 | WS1 |  |  |
| WS1 | all | END | WS |  |
| WS1 | \t | \v | ‘ ‘ | WS1 |  |  |