

采用 Java 语言实现词法分析器分析 C 语言

关键字:

```
keywords =  
"auto,double,int,struct,break,else,long,switch,case,enum,register,typedef,  
char,extern,return,union,const,float,short,unsigned,continue,for,signed,  
void,default,goto,sizeof,volatile,do,if,while,static";
```

数字:

```
digit = "0123456789";
```

字符:

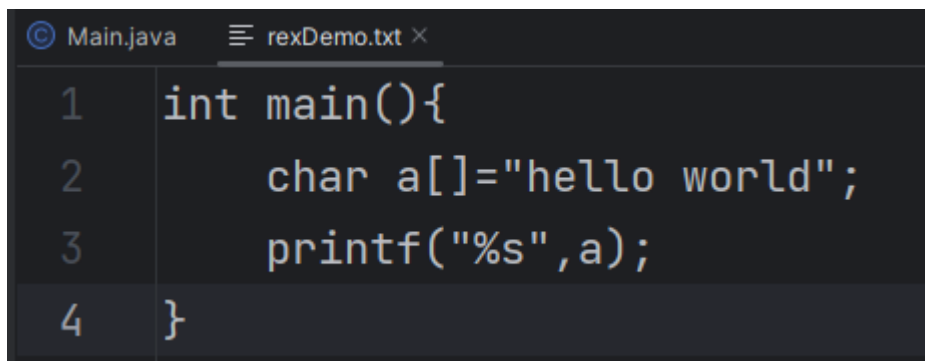
```
symbol = "{,},(,),\",>,<,>=,=,==,<=,&,&&,%/,/,;,[,]";
```

字母:

```
letter = "abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ";
```

****皆为自定义，字符可能不全****

被解析文件: rexDemo.txt



```
1 int main(){  
2     char a[]="hello world";  
3     printf("%s",a);  
4 }
```

程序代码:

```
import java.util.*;  
  
/**  
 * @author 俞家宝  
 * @version 1.0  
 * @className 实现词法解析器  
 * @dateTime 2023/10/7 1:47  
 */  
import java.io.*;  
  
public class Main {  
    public static String keywords =  
"auto,double,int,struct,break,else,long,switch,case,enum,register,typedef,char,ex  
tern,return,union,const,float,short,unsigned,continue,for,signed,void,default,got  
o,sizeof,volatile,do,if,while,static";  
    public static String symbol = "{,},(,),\",>,<,>=,=,==,<=,&,&&,%/,/,;,[,]";  
    public static String digit = "0123456789";  
    public static String letter =  
"abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ";  
    public static List<String> Token = new ArrayList<>();  
    public static int status = 0;  
    public static int index = 0;
```

```

public static void main(String[] args) throws FileNotFoundException {
    String filePath = "E:\\JavaProject\\Homework\\src\\rexDemo.txt";
    String textLine = getText(filePath);
    checkTerms(textLine);
    System.out.println(textLine);
    System.out.println(Token);
}

public static String getText(String filePath) {
    StringBuilder demo = new StringBuilder();
    try (Scanner sc = new Scanner(new FileReader(filePath))) {
        while (sc.hasNextLine()) { //按行读取字符串
            String line = sc.nextLine();
            demo.append(line);
        }
    } catch (FileNotFoundException e) {
        throw new RuntimeException(e);
    }
    return demo.toString();
}

public static void checkTerms(String textLine) {
    while (index != textLine.length() - 1) {
        String temp = getIndex(textLine);
        if (!temp.equals(" ") && !temp.isEmpty())
            getTokenType(temp);
    }
}

public static String getIndex(String text) {
    int start = index;
    while (true) {
        String temp = String.valueOf(text.charAt(index));
        if (temp.equals(" ")) {
            index++;
            return text.substring(start, index - 1);
        }
        if (!changeStatus(temp)) {
            if (status == 0) status = 1;
            else status = 0;
            return text.substring(start, index);
        }
        if (index + 1 < text.length())
            index++;
        else
            return text.substring(start);
    }
}

public static boolean changeStatus(String checked) {
    if (symbol.contains(checked) && status == 1 || !symbol.contains(checked)
    && status == 0) {

```

```

        return true;
    }
    return false;
}

public static void getTokenType(String tokenValue) {
    if (keywords.contains(tokenValue) || symbol.contains(tokenValue)) {
        Token.add(tokenValue);
    } else {
        for (int i = 0; i < tokenValue.length(); i++) {
            if (symbol.contains(String.valueOf(tokenValue.charAt(i))))
                Token.add(String.valueOf(tokenValue.charAt(i)));
            else
                Token.add("id" + "->" +
String.valueOf(letter.indexOf(String.valueOf(tokenValue.charAt(i)))));
        }
    }
}
}

```

运行结果：

输出读入数据
int main(){ char a[]="hello world"; printf("%s",a);}
输出 Token 流
[int, id->12, id->0, id->8, id->13, (,), {, char, a, [,], =, ", id->7, id->4, id->11, id->11, id->14, id->22, id->14, id->17, id->11, id->3, ", :, id->15, id->17, id->8, id->13, id->19, id->5, (, ", %, s, ", , a,), ;, . }