Tiger Lexical Analysis Documentation

Xinyi Yu
515030910029
aureole@sjtu.edu.cn

Handle comments

- Use states. On encountering "/*", state transits from <INITIAL> to <COMMENT>.
- Use @variable `commentLayer` to handle nested comments, determine when to transit back to <INITIAL>. The pseudocode:

```
commentLayer = 0;
if encounters '/*' in initial state:
    commentLayer ++;
    transit to comment state;
if encounters '/*' in comment state:
    commentLayer ++;
if encounters '*/' in comment state:
    commentLayer --;
    if commentLayer == 0:
        transit to initial state;
```

Handle strings

- Use states, similar with above.
- Use @variable `stringBuf` to temporarily store the current String object content and @variable `stringPos` to record the header of `stringBuf`.
- Handle special char, for example '\n' should not be interpreted as char '\' and char 'n'. (Future work) This part should be improved.

Handle errors

Not did much. Only trigger error when encountering illegal char in <INITIAL> state.

Handle EOF

When writing code, I didn't think about that at all. When I searched online, I found ([link](#)):

> "Flex has <<EOF>> pattern which matches only end-of-file. Probably you can return some token indicating end-of-file to the parser as the following.
> `<<EOF>>      return END_OF_FILE;`"

That means the EOF pattern is `<<EOF>>`. If we want to return tokens or do things on EOF, we can use this pattern.