

Trabalho de Implementação de Analisador Léxico

Objetivos: Projetar e Implementar Analisador Léxico para a linguagem que escolheram para o projecto da disciplina. Para tanto, utilizem a linguagem Java e tomem como base as aulas teóricas e práticas do curso de Compiladores até o momento.

Os estudantes deveram baixar e estudar a ferramenta jFlex e aplicar ao projecto

(<http://www.jflex.de/>).

NB: Este trabalho pesará em 30% na nota do projecto

I. O que deve ser feito?

1. Construir a classe AnaLex e métodos para análise léxica. Pelo menos um método (AnaLex) deve existir.
2. O método deve “ler” o conteúdo do programa passado como parâmetro e retornar os tokens encontrados.
3. Controlar a numeração das linhas assim como deve-se ignorar comentários. Para cada token reconhecido, deve ser possível exibir: <Número da linha do token, Token, Atributo (quando possuir atributo)>
4. Observe que o método AnaLex retorna um token sempre que for chamado. Esse método não deve escrever o token na tela, mas sim produzir um token para um outro método (método main, por exemplo) que será o responsável pela saída
5. As palavras-chave são consideradas palavras reservadas, i.e. não podem ser utilizadas como identitificadores. Usem como base a tabela de palavras reservadas da linguagem Java, e construam a vossa base de paravras reservadas e construam uma tabela que mostra um mapeamento entre a vossa palavra reservada e seu equivalente em Java.
6. Criar e gerir uma tabela de palavras reservadas (tabela hash ou vetor de strings) para as palavras reservadas da linguagem. Toda vez que um identificador for encontrado, deverá ser feita uma busca na tabela de palavras reservadas. Uma busca com sucesso nessa tabela indica que uma palavra-chave foi reconhecida e então seu token será retornado. Caso contrário, um identificador foi reconhecido e o mesmo deverá ser inserido na tabela de símbolos, caso ainda não tenha sido inserido.

7. Implementar tabela de símbolos como uma tabela hash ou tabela de dispersão. Quando um identificador é reconhecido, deve-se inicialmente fazer uma busca na tabela de palavras reservadas para verificar se é uma palavra reservada. Caso a busca não tenha sucesso, deve-se fazer uma nova busca na tabela de símbolos, e retornar o token associado, caso o identificador seja encontrado. Em caso negativo, o identificador deve ser inserido na tabela de símbolos e então retornar o token associado.
8. A vossa linguagem não faz distinção entre identificadores com letras maiúsculas ou minúsculas. ARRAY é equivalente à array. Sugere-se que todas as letras sejam convertidas para uma das duas formas antes de qualquer comparação.

Sugestão de metodologia para desenvolvimento do trabalho

- i. Trabalho cooperativo e participativo em grupo. Na medida do possível, utilizem a metodologia XP para realização desse trabalho.
- ii. Façam o projeto do analisador léxico. Organizem as classes, atributos e métodos. Lembrem-se que, mais tarde, acrescentaremos um analisador sintático nesse projeto!
- iii. Implemente o reconhecimento de um token, faça testes, passe para o próximo. Fazendo isso, erros simples detectados nessa implementação não se repetirão nas próximas.
- iv. Não apaguem os programas que estão a criar para testar os vossos analisador léxico. O conjunto de programas testes deve incrementar de tamanho a medida que novas características são implementadas.