

Mestrado em Informática

UCE30- Engenharia de Linguagens

Módulo: Engenharia Gramatical

Ficha Prática 1: exercício sobre GAs e geração de PLs
Ano Lectivo 2008/09

2008.out.01

Exercício 1

Nos tempos que correm, até a produção científica de uma instituição académica tem de ser contabilizada. Para isso é preciso colectar determinados resultados obtidos pelos membros do centro de investigação em causa e fazer contagens por anos em avaliação, ou considerando outros parâmetros. Para organizar toda essa informação, descrevendo os resultados atingidos e associando-os a quem os produziu, pretende-se criar uma linguagem específica¹ (que se denominará GCI – *Gestão de Centros de Investigação*) que facilite a recolha e permita automatizar os cálculos, além de servir para consolidar certos dados.

A gramática independente de contexto G , abaixo apresentada, define essa linguagem específica para apoio à contabilização dos processos de orientação de pós-graduações (PGs)—isto é, das supervisões de Doutoramentos (phd) e Mestrados (msc)—de um dado grupo de investigação, permitindo descrever cada projecto concluído ou em andamento dentro do grupo.

```
T = { num, id, str, PHD, MSC, CO-ORIENT, INI, FIM, ".", ";", "(", ")" }
N = { Gci, PGs, Pg, IdOrient, Tipo, CoOrient, Aluno, Titulo, Nome, Inic, Fim, Ano }
S = Gci
P = {
  p0: Gci      --> PGs '.'
  p1: PGs      --> Pg
  p2:         | PGs ';' Pg
  p3: Pg       --> IdOrient Tipo CoOrient Aluno '(' Titulo ')' Inic Fim
  p4: IdOrient --> id
  p5: Tipo     --> PHD
  p6:         | MSC
  p7: CoOrient --> &
  p8:         | CO-ORIENT Nome
  p9: Aluno    --> Nome
  p10: Titulo  --> str
  p11: Nome    --> str
  p12: Inic    --> INI Ano
  p13: Fim     --> &
  p14:         | FIM Ano
  p15: Ano     --> num
}
```

sabendo-se ainda que os símbolos terminais variáveis e os comentários válidos são definidos pelas seguintes expressões regulares:

¹Uma textsfDSL.

```
num : [0-9]+
id  : [a-zA-Z][a-zA-Z0-9/_]*
str : \"[^\"]*\"
comentario1 : \"//\".*
comentario2 : \"\{^\}*\}
```

Desenvolva, então, com recurso ao gerador AnTLR um reconhecedor para aceitar as frases válidas da linguagem GCI. Do ponto de vista semântico, uma frase só será considerada válida se se observarem as seguintes restrições:

- um orientador não pode ter mais de duas teses de cada tipo;
- a data de fim de uma tese concluída nunca pode ser menor ou igual à sua data de início.

Quanto aos resultados a produzir ao processar automaticamente as frases da linguagem GCI (semântica dinâmica da linguagem) pretende-se:

- gerar uma listagem com as siglas de todos Orientadores (sem repetições);
- imprimir, por ordem crescente de ano de início, os alunos de Doutoramento (nome, título da tese e orientador) com teses em andamento (iniciadas mas não acabadas);
- contar e imprimir o número total de teses concluídas e não-concluídas por ano de conclusão e por tipo (phd, ou msc).

Para resolver este problema escreva uma gramática de atributos (GA) que especifique formalmente as restrições definindo as frases com significado (semanticamente válidas) e o resultado a obter do processamento das frases (semântica dinâmica) e use de novo o AnTLR para produzir o pretendido Processador para a linguagem GCI.