

Processamento de Linguagens – LEI (3ºano)

Processamento de Linguagens e Compiladores – LCC (2ºano)

1º Teste

Data: 21 de Junho de 2012

Hora: 11:00

Dispõe de 2 horas para realizar este exame

1 Filtros de Texto e Analisadores Léxicos (6v)

Supondo que vai usar o gerador Flex para construir automaticamente os programas abaixo pedidos, escreva a respectiva especificação:

- a) Um documento de texto gerado pelo Bb (Blackboard, o sistema de eLearning adotado pela UM) descreve todos os grupos inscritos a uma dada UC conforme se exemplifica abaixo. Cada descrição começa pela palavra "GRUPO" seguido por dois pontos (":") seguido pelo número do grupo (espaços são permitidos entre qualquer um destes elementos). Depois vem blocos onde se descreve cada aluno do grupo (número, nome, estatuto, etc.). O nome do aluno aparece em um linha começada pela palavra "NOME" seguida de zero ou mais espaços e do carater "=" e inclui todos os caracteres a partir daí até terminar com o carater "."

Exemplo

```
GRUPO: 1
NOME=Antonio Carlos Mendes.
.....
NOME =Ana Gabriela Dias Moura.
GRUPO : 2
NOME=Rui Ramos Costa.
.....
```

Desenvolva um Filtro que receba o texto acima descrito e:

- a1) conte o número total de grupos;
- a2) verifique que todos os grupos tem entre 2 e 3 alunos (escreva uma mensagem caso tal não ocorra);
- a3) produza uma listagem ordenada com o nome de todos os alunos seguido do número de grupo a que está inscrito
- b) Desenvolva um Filtro que receba o texto em L^AT_EX e o devolva como o recebeu apenas substituindo os comandos de formatação

`\textbf`, `\texttt`, `\emph`

pelos comandos correspondentes em HTML (Nota: Lembre-se que os comandos podem aparecer aninhados):

``, `<TT>`, `<I>`

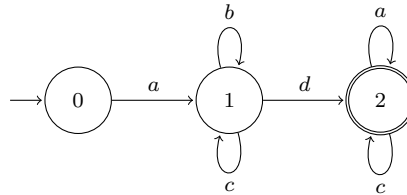
- c) Desenvolva um Analisador Léxico para a linguagem definida pela GIC do exercício 4, a integrar no tradutor gerado pelo Yacc. O Analisador Léxico deve devolver o **código** dos Símbolos Terminais, conforme valor estipulado pelo Yacc, bem como o respectivo **valor semântico** (sequência de caracteres lida) para o caso dos *terminais-variáveis* que se descrevem abaixo; deve ainda eliminar todos os *comentários* (qualquer texto entre “{” e “}”). As *palavras-reservadas* poderão ser formadas apenas por letras maiúsculas; Um **id** é uma palavra formada por letras, podendo eventualmente terminar por um algarismo; um **texto** é uma sequência eventualmente nula de caracteres entre apóstrofes ('); um **num** é um numero inteiro; uma **data** é formada por dígitos segundo o formato `aaaa-mm-dd`.

2 Expressões Regulares e Autômatos (5v)

Considere as seguintes ERs:

$$\begin{aligned} e1 &= a^+ (c d + a c) (d b)^+ \\ e2 &= a (b + c)^* d (a c)^* \\ e3 &= (a b^* + a c^*) (d a + d c)^* \end{aligned}$$

e autômato determinista AD1



Responda, então, às seguintes questões:

- depois de explicar por palavras suas a linguagem definida pela ER $e2$, dê 3 exemplos de frases válidas dessa linguagem incluindo a frase mais curta.
- usando *cadeias de derivação*, verifique se a sequência de símbolos “**abdac**” é uma frase válida (pertence) à linguagem gerada por cada uma das três ERs acima.
- construa, aplicando as regras sistemáticas, um autômato não-determinista (AND) equivalente a $e1$.
- construa, informalmente, um autômato determinista (AD) equivalente a $e2$.
- indique a expressão regular equivalente ao autômato AD1 e depois mostre que a ER $e3$ acima não é equivalente a esse autômato AD1, apesar da frase “**abda**” pertencer a ambas as linguagem.

3 Desenho/especificação de uma Linguagem (3v)

Um colecionador de lápis de grafite tradicionais pretende ter uma linguagem de domínio específico simples que lhe permita ir descrevendo cada nova peça que adiciona ao espólio—forma (cilíndrico, facetado, achatado), estado (novo, usado), estilo (liso, colorido, decorado), fabricante (se conhecido), país origem (se conhecido), remate do topo (sem nada, com boneco). Para questão de catalogação, o colecionador pretende dividir a coleção em dois grandes grupos: lápis normais e publicitários. No caso dos lápis publicitários em vez de registrar o fabricante e o país origem, pretende-se guardar o nome da empresa/instituição que publicita.

Escreva então uma Gramática Independente de Contexto, *GIC*, que especifique a Linguagem pretendida—note que o estilo da linguagem (mais ou menos verbosa) e o seu desenho são da sua responsabilidade.

4 Gramáticas, Linguagens, Parsing e Tradução (6v)

A gramática independente de contexto, *GIC*, abaixo escrita em *BNF*, define uma linguagem de domínio específico para descrição de uma base de dados de perguntas de resposta múltipla para exames tipo americano de uma dada UC (unidade curricular).

O Símbolo Inicial é **Questoes**, os Símbolos Terminais são escritos só em minúsculas (terminais-variáveis) ou só em maiúsculas (palavras-reservadas) ou entre apostrofes (sinais-de-pontuação), e a string nula é denotada por **&**; os restantes (sempre começados por maiúsculas) serão os Símbolos Não-Terminais.

```
p0: Questoes    --> UC Nome Qs '.'
```

```
p1: Qs          --> Questao
```

```
p2:            | Qs ';' Questao
```

```
p3: Questão     --> id Assunto Grau Enunciado Resps Solucao MetaInfo
```

```
p4: Assunto     --> TEMA texto
```

```
p5: Nome       --> texto
```

```
p6: Grau       --> FACIL
```

```
p7:           | MEDIO
```

```
p8:           | DIFICIL
```

```
p9: Enunciado  --> texto
```

```
p10: Solucao   --> num
```

```
p11: Resps     --> Rsp
```

```
p18:          | Resps ',' Rsp
```

```
p12: Rsp       --> num Afirmacao
```

```
p13: Afirmacao --> texto
```

```
p14: MetaInfo  --> &
```

```
p15:          | AUTOR Nome DATA data
```

Neste contexto e após analisar a *GIC* dada, responda às alienas seguintes.

- Usando uma Árvore de Derivação escreva uma frase válida da linguagem gerada por essa *GIC*.
- Escreva a função de um parser RD (recursivo-descendente puro) para reconhecer o símbolo não-terminal **MetaInfo** e acrescente apenas a função genérica para reconhecer qualquer símbolo terminal dado.
- Recorde o que sabe sobre as Tabelas de Parsing LL(1) e construa as 2 linhas correspondentes aos símbolos não-terminais **Questao** e **Grau**.
- Construa o estado 0 (estado inicial) do Autômato LR(0) e os estados dele imediatamente derivados.
- transforme a *GIC* dada numa **gramática tradutora**, *GT*, reconhecível pelo *Yacc*, que:
 - calcule e imprima o número total de questões contidas na base e dados e indique também os subtotais por cada um dos 3 graus de dificuldade.
 - verifique que a cada questão estão associadas sempre exatamente 4 respostas.