

Processamento de Linguagens I

LESI (3º ano)

Trabalho Prático nº 1
(Lex e Yacc)

Ano lectivo 06/07

1 Objectivos e Organização

Este trabalho prático tem como principais **objectivos**:

- aumentar a experiência de uso do ambiente linux, da linguagem imperativa C (para codificação das estruturas de dados e respectivos algoritmos de manipulação), e de algumas ferramentas de apoio à programação;
- rever e aumentar a capacidade de escrever *gramáticas independentes de contexto* que satisfaçam a condição LR());
- desenvolver processadores de linguagens segundo o método da *tradução dirigida pela sintaxe*, suportado numa *gramática tradutora*;
- utilizar *geradores de compiladores* como o par `lex/yacc`

Para o efeito, esta folha contém 5 enunciados, dos quais deverá resolver pelo menos um¹.

O programa desenvolvido será apresentado a um dos membros da equipa docente, totalmente pronto e a funcionar (acompanhado do respectivo relatório de desenvolvimento) e será defendido por todos os elementos do grupo (3 alunos), em data a marcar.

O **relatório** a elaborar, deve ser claro e, além do respectivo enunciado, da descrição do problema, das decisões que lideraram o desenho da linguagem e a concepção da gramática, do esquema de tradução e respectivas acções semânticas (incluir as especificações `lex` e `yacc`), deverá conter exemplos de utilização (textos fontes diversos e respectivo resultado produzido). Como é de tradição, o relatório será escrito em \LaTeX .

O pacote de software desenvolvido (um ficheiro compactado, ".tgz", contendo os ficheiros ".l", ".y", algum ".c" ou ".h" que precise, os ficheiros de teste ".txt", o relatório ".tex" e a respectiva "makefile") deve ser entregue através do sistema de submissão de TPs, até ao dia **16 de Abril**.

¹Se resolver mais, cada um será avaliado separadamente e a nota final será a média das notas individuais obtidas.

2 Enunciados

Para sistematizar o trabalho que se lhe pede em cada uma das propostas seguintes, considere que deve, em qualquer um dos casos, realizar a seguinte lista de tarefas:

1. Especificar a gramática concreta da linguagem de entrada.
2. Desenvolver um reconhecedor léxico e sintáctico para essa linguagem com recurso ao par de ferramentas geradoras *flex/yacc*.
3. Construir o gerador de código que produza a resposta solicitada. Esse gerador de código é construído associando acções semânticas de tradução às produções da gramática, recorrendo uma vez mais ao gerador *yacc*.

2.1 Rede Semântica do Museu da Emigração

Neste projecto pretende-se descrever parte da *rede semântica* que suporta o Museu da Emigração, das Comunidades e da Luso-descendência (www.museu-emigrantes.org) e gerar um grafo (usando *GraphViz / WebDot*) que permita uma navegação conceptual sobre esse repositório de conhecimento; além de poder percorrer os caminhos do grafo, pode seleccionar um nodo e saltar para uma página com info sobre esse elemento.

Para o efeito, deve definir e processar uma linguagem que lhe permita descrever três tipos de nodos, ou vértices, da rede—*emigrante* (com dados pessoais e dados relativos ao processo de emigração), *obra* (palácio, fábrica, hospital, escola, capela, etc.), *evento* (baile, sarau literário, sarau musical, etc.)—e os respectivos arcos—*fez* (liga emigrante a obra), *participou* (liga emigrante a evento).

2.2 Dicionário Cinematográfico

Considere que, no âmbito da realização de filmes, se pretende criar um **Dicionário** de palavras utilizadas nesta área. Associado a cada palavra pretende-se ter: o significado; a designação comum em inglês; e uma lista de sinónimos.

O seu sistema de tratamento de textos cinematográficos, **STTC**, deve aceitar uma linguagem que especifique o ficheiro onde está o dicionário (escrito na linguagem que terá de definir) e uma lista de ficheiros onde se encontram textos a anotar.

O objectivo do **STTC** é ler o dicionário e carregá-lo para um estrutura interna de tal modo que depois possa processar os ditos textos. Para isso, deve procurar em cada um as palavras que estejam presentes no dicionário assinalando-as (sublinhado) e associando-lhes uma **footnote** com o termo em inglês. No fim do texto será gerado um apêndice com a lista de palavras encontradas e respectivos significados.

2.3 Escolha espiritual de música

O objectivo deste projecto é construir uma musiteca espiritual, a **MEspiritual** – a escolha de músicas consoante o estado de espírito.

Para o efeito será necessário definir uma linguagem para descrever cada música disponível na musiteca (nome, autor(es), ano, editor, letra da música, estilo (rock, pop, clássico, ...), tipo, ...). Quanto ao tipo, podem associar-se um ou mais descritores (tais como: calmo, agitada, relaxante, alegre, triste, espevitante, dançante, ...) afectados por uma percentagem.

A partir da descrição desse repositório musical, pretende-se gerar um site que aceite um pedido (escrito numa linguagem específica) baseado no estado de espírito do visitante (enérgico, calmo, ...) e mostre um grafo com todas as músicas disponíveis que se adequem melhor ou pior a esse pedido espiritual.

A noção de adequar é deixada ao critério do grupo (quer o conceito, quer a métrica).

2.4 A famosa ODume

O objectivo deste projecto é construir uma *site* para a famosa ourivesaria de Dume, ODume.

A pequena ourivesaria ODume vende: **jóias** de vários tipos—anéis, brincos, colares e alfinetes, em ouro ou prata, lisos ou com pedras preciosas (rubis, ametistas topázios, esmeraldas); **objectos de prata**—jarras, taças, molduras, cigarreiras, candelabros, faqueiros; e **relógios**—pulso, mesa (despertadores) e parede (com ou sem pé).

Cada jóia será descrita, no mínimo, por uma *referência*, o *tipo*, a *data de entrada*, o *fabricante*, o *preço base*, o *peso* (em gramas) da peça relativo ao *material de base* (ouro ou prata), o *número e tipo de pedras preciosas*.

O site pretendido deve reflectir os 3 expositores da loja (para cada um dos 3 tipos de artigo) e cada expositor corresponderá ao tipo de mercadoria acima explicada.

Pretende-se ainda que no site haja uma linguagem de query para pedir um determinado tipo de objecto a consultar. O sistema produzirá um link directo para a respectiva página.

2.5 Informação Geográfica

O objectivo deste projecto é construir um *site* para as localidades de Portugal face à sua hierarquia geográfica definida à custa da linguagem OrgGeo abaixo descrita.

OrgGeo é uma Linguagem de Domínio Específico (DSL) que se destina a descrever a organização geo-política de um país, ou sua região. Para o efeito, pretende-se identificar as Localidades de cada Concelho de cada Distrito da região em causa.

A gramática independente de contexto G , abaixo apresentada, define o tipo de linguagem pretendida—cada grupo pode criar a sua variante.

O Símbolo Inicial é OrgGeo, os Símbolos Terminais são escritos em minúsculas (pseudo-terminais), ou em maiúscula (palavras-reservadas), ou entre apostrofes (sinais de pontuação) e a string nula é denotada por &; os restantes serão os Símbolos Não-Terminais.

```
p1: OrgGeo --> Distritos
p2: Distritos--> IdD DescD '.'
```

```

p3:          | Distritos IdD DescD ','
p4: DescD    --> Concelhos
p5: Concelhos--> Concelho
p6:          | Concelhos ';' Concelho
p7: Concelho --> Locais ':' IdC
p8: Locais   --> Local LstIds
p9: LstIds   --> &
p10:         | ',' Locais
p11: IdD     --> id
p12: IdC     --> id
p13: Local   --> IdL .....
p14: IdL     --> id

```

A informação a incluir em cada `Local` fica ao cuidado do grupo, sendo que se deve contemplar a possibilidade de fornecer um URL para um site do local; caso tal suceda, no site gerado existirá um link para esse site do local.

Uma possível forma de desenvolver o projecto é definida a seguir: Transforme G numa **gramática tradutora**, GT , reconhecível pelo `yacc`, para gerar um sítio HTML com uma página por cada Distrito, contendo uma lista de items com os Concelhos e para cada um uma outra lista com suas localidades. Complete a GT anterior de modo a gerar também uma página inicial (a *homepage* do seu sítio) com um título (fixo) do projecto e links para as páginas dos Distritos.

Adicionalmente, estenda o seu processador para incluir a construção duma **Tabela de Localidades** que associe a cada identificador de Localidade os respectivos Concelho e Distrito. Note que podem existir Localidades com nomes repetidos, desde que tal não aconteça no mesmo Concelho. Nomes de Concelhos e de Distritos é que não podem ser duplicados.