

Processamento Estruturado de Documentos 2001

LESI + LMCC (5º ano)

Ficha Teórico-Prática nº1

Data: 24 de Outubro de 2001

Exercício 1 Pretende-se criar uma agenda para gerir contactos de pessoas e empresas.

Uma agenda deste tipo é essencialmente constituída por uma lista de entradas.

Uma entrada pode ser simples, correspondendo a uma pessoa ou empresa, ou pode ser composta, neste caso corresponde a uma lista de pessoas e/ou empresas.

A agenda poderá conter um número livre de grupos aninhados.

Cada entrada é constituída pelos seguintes itens de informação:

id uma string que corresponderá ao identificador único da entrada.

tipo campo com dois valores possíveis: pessoa ou empresa.

nome nome da pessoa ou da empresa.

email email da pessoa ou empresa.

telefone telefone da empresa ou pessoa.

gid no caso da entrada se tratar dum grupo este campo deverá conter o identificador do grupo.

Nas alíneas que se seguem considere que uma frase concreta desta linguagem tem a seguinte forma:

```
<AGENDA>
  <ENTRADA id="e1" tipo="pessoa">
    <NOME>José Carlos Ramalho</NOME>
    <EMAIL>jcr@di.uminho.pt</EMAIL>
    <TELEFONE>253 604479</TELEFONE>
  </ENTRADA>
  <GRUPO gid="ep1">
    <ENTRADA id="e2" tipo="pessoa">
      <NOME>Pedro Henriques</NOME>
      <EMAIL>prh@di.uminho.pt</EMAIL>
      <TELEFONE>253 604469</TELEFONE>
    </ENTRADA>
    <ENTRADA id="e3" tipo="pessoa">
      <NOME>João Saraiva</NOME>
      <EMAIL>jas@di.uminho.pt</EMAIL>
      <TELEFONE>253 604479</TELEFONE>
    </ENTRADA>
  </GRUPO>
  <ENTRADA id="e4" tipo="empresa">
    <NOME>Lavandaria Tudo Branco</NOME>
    <TELEFONE>253 604433</TELEFONE>
  </ENTRADA>
</AGENDA>
```

Desenvolva então, cada uma das seguintes alíneas:

- a) Especifique um DTD para este tipo de documento.
- b) Valide a instância acima com o DTD que especificou usando o Omnimark e a script parser.xom:

- o Omnimark está instalado no orunner e corre com a seguinte linha de comando:

```
omnimark -s nome-script documento-xml
```

- a script parser.xom está disponível na página da disciplina, em Material Didático:

```
http://www.di.uminho.pt/~jcr/AULAS/ped2001
```

- c) Crie uma script XML::DT para transformar estes documentos em HTML:

- lembre-se da geração automática de scripts de transformação:

```
perl -MXML::DT -e 'mkdtskel "f.xml"' > f.pl
```

- seria interessante que a página HTML gerada apresentasse, no topo, o número de entradas na agenda (um grupo deverá ser contabilizado com um número igual ao das suas subentradas e assim sucessivamente).

Exercício 2 Para, nesta aula, poderes testar os teus conhecimentos de XPath abstraindo-te da restante complexidade duma stylesheet XSL, foi desenvolvida uma metodologia: foi criado um pequeno DTD para documentos que irão conter expressões XPath e que depois de processados em conjunto com uma determinada instância documental dão como resultado um novo documento XML composto pelas subárvores resultantes da aplicação das expressões XPath à instância documental.

O DTD para documentos XPath é simples:

```
<!ELEMENT query-set (query+)>  
<!ELEMENT query (#PCDATA)>
```

O conteúdo do elemento query deverá ser a expressão XPath a ser calculada.

Para se obter o resultado do cálculo duma expressão XPath sobre uma instância documental (por exemplo um poema) é necessário seguir os dois passos seguintes:

1. usar um gerador de stylesheets XSL que aplicado ao documento com as expressões XPath vai gerar uma stylesheet XSL que calcula essas expressões; foram desenvolvidos dois geradores, um em XML::DT — o xpath.pl, e outro que não é mais nem menos do que uma stylesheet XSL — o xpath.xsl; ambos os geradores estão disponíveis na página da disciplina; para se obter a stylesheet em causa é necessário executar uma das duas linhas seguintes:

```
perl xpath.pl q1.xml > q1.xsl
```

```
testXSLT -XSL xpath.xsl -IN q1.xml -OUT q1.xsl
```

2. a seguir utiliza-se a stylesheet gerada para processar a instância e obter o resultado do cálculo das expressões:

```
testXSLT -XSL q1.xsl -IN documento.xml -OUT q1-res.xml
```

Tente resolver os seguintes exercícios:

- a) Se não tem a instância do poema da última aula faça o seu download. Crie um documento baseado no DTD acima com expressões XPath que respondam às seguintes questões:

- o terceiro verso da última quadra
- todos os nomes
- os versos do segundo terceto
- os nomes ou lugares começados pela letra L

- b) *Faça o mesmo com o documento dos sumários de PED2000 e obtenha uma resposta para as seguintes questões:*
- *todas as aulas teóricas*
 - *todas as aulas práticas*
 - *a lista das datas das aulas*
 - *o primeiro parágrafo do sumário de cada aula*
- c) *Faça o mesmo com a agenda de contactos depois de lhe ter adicionado mais umas entradas. As interrogações são as seguintes:*
- *todas as entradas de Braga (indicativo 253)*
 - *todas as entradas de pessoas com nome contendo "José" ou "Maria"*
 - *os identificadores das entradas pertencentes a grupos*
 - *os identificadores das entradas que não são do tipo "pessoa"*
- d) *Crie uma script XML::DT que transforme os documentos criados em páginas HTML que têm no início um índice de queries e que permitem navegar nos resultados.*