

Processamento Estruturado de Documentos 2001

LESI + LMCC (5º ano)

Ficha Teórico-Prática nº2

Data: 31 de Outubro de 2001

Exercício 1 (Arqueossítios) *Vá à página de Material Didático e descarregue o documento XML que descreve os arqueossítios do Noroeste Português.*

Realize então, cada uma das seguintes tarefas:

- 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 (faça isto e ciclicamente vá alterando o DTD, há subtilezas que só serão encontradas no processo de validação):*

- *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, um índice numérico para todos os arqueossítios sendo que cada arqueossítio deveria estar numerado, também pode usar a identificação do arqueossítio para o índice).*

Exercício 2 (Arqueossítios e XPath) *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 de uma 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
```

```
sabcmd xpath.xsl q1.xml > q1.xsl
```

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

```
sabcmd q1.xsl documento.xml > q1-res.xml
```

Tente resolver os seguintes exercícios respectivos ainda ao documento dos *arqueossítios*:

- a) Crie um documento baseado no DTD acima com expressões XPath que respondam às seguintes questões:
 - os *arqueossítios* pertencentes ao concelho de Amares
 - os *arqueossítios* que tenham ligações a termos do dicionário
 - a identificação dos *arqueossítios* que se encontram a uma latitude superior a 180
 - a identificação dos *arqueossítios* que tenham o nome "Martins" na bibliografia
- b) Faça o mesmo com o documento dos sumários de PED2000 (também estão na página) 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) *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. Analise previamente os documentos resultado deste sistema e pense no mapeamento mais correcto dos elementos deste tipo de documento para elementos HTML.*

Exercício 3 (Bibliografia) *Descarregue o ficheiro relativo à Bibliografia do JCR da página da disciplina e tente resolver as alíneas seguintes:*

- a) *Desenvolva uma script XML::DT para gerar uma versão HTML do documento atendendo aos seguintes pormenores:*
- *Cada entrada deverá estar individualizada das restantes (linhas horizontais, espaçamento, ...).*
 - *No topo da página, deverá aparecer um índice (disposto horizontalmente ou verticalmente) que permita um acesso directo a cada entrada. Utilize o identificador da entrada.*
 - *Tente criar um índice de autores em que a cada autor está associada a lista de identificadores das respectivas entradas.*
- b) *Se ainda tiver tempo especifique um DTD para este tipo de documentos.*