

Ficha 9 – XML

Bases de Dados, FCT-NOVA

Ano letivo 2020/21

Grupo 1. Suponha que pretende definir a estrutura para documentos XML que irão conter uma lista de alunos do Departamento de Informática. Considere os seguintes requisitos:

- uma lista de alunos contém um ou mais alunos.
 - um aluno tem um número de matrícula, pelo qual pode ser identificado. Uma entrada de um aluno contém o número do aluno, o nome do aluno, o curso em que o aluno está inscrito, as cadeiras já feitas e as cadeiras em que está inscrito. Para cada aluno pretendemos igualmente apresentar o ano da primeira inscrição.
 - o nome do aluno consiste num primeiro nome, nos nomes “intermédios”, que são opcionais, e no último nome.
 - os cursos são “Engenharia Informática”, “Mestrado Europeu em Lógica Computacional” e “Doutoramento em Informática”.
 - De cada cadeira pretendemos manter o nome da cadeira. As cadeiras são de licenciatura, de mestrado, de doutoramento, ou de vários níveis.
 - das cadeiras já feitas, queremos manter informação sobre a nota obtida. Das cadeiras em que o aluno está inscrito, queremos manter informação sobre a nota eventualmente já obtida na componente prática.
1. Escreva um DTD que defina a estrutura para documentos do tipo `listadealunos`.
 2. Escreva uma instância de `listadealunos` que seja válida de acordo com o DTD.
 3. Escreva uma instância de `listadealunos` que contenha apenas elementos definidos no DTD, mas que seja inválida.

Pode testar os erros, colocando os seus ficheiros online e usando um dos muitos validadores de ficheiros XML que existem para uso livre.

Grupo 2. Considere o documento XML `countries.xml` que acompanha a ficha.

Escreva expressões XPATH que devolvam os resultados das perguntas abaixo¹.

1. A área da Mongólia;
2. Os nomes dos países com população superior a 100 milhões;
3. Os nomes dos países onde mais de 50% da população fala alemão. (Dica: dependendo da solução, poderá usar “.”, que se refere ao nó do elemento corrente dentro de uma expressão XPATH.)

¹Para testar as queries em XPath pode usar um dos muitos sistemas livres que existem para isso. O próprio Google Chrome permite colocar queries XPath em Tools–Developer Tools. Uma forma mais simpática de correr essas perguntas é usando o addon XPath Checker do Firefox. Também existem vários plugins para o Chrome que permitem fazer queries em XPath (e.g. o XPath Helper). Também pode usar directamente um tool online, por exemplo <https://www.online-toolz.com/tools/xpath-tester-online.php>, <https://codebeautify.org/Xpath-Tester> ou <http://videlibri.sourceforge.net/cgi-bin/xidelcgi>.

4. Os nomes dos países onde uma das suas cidades contém mais de um terço da população desse país.
5. Os nomes dos países cuja população é inferior a um milésimo da de alguma cidade (num qualquer país).
6. Os nomes das cidades que têm o mesmo nome que os países onde estão localizadas.
7. Os nomes dos países que têm pelo menos três cidades com população superior a 3 milhões.
8. Os nomes dos países para os quais não há dados sobre língua nem cidades, mas que têm uma população superior a 10 milhões.
9. Nomes dos países cujo nome contenha, textualmente, uma língua falada nesse país. Por exemplo, Uzbek é falado no Uzbekistan pelo que Uzbekistan deve ser devolvido. (Dica: o predicado `contains/2` poderá ser útil.)
10. Nomes dos países nos quais seja falado uma língua cujo nome contenha, textualmente, o nome do país. Por exemplo, Japanese é língua de Japan, pelo que Japan deve ser devolvido.
11. Nomes das línguas faladas num país cujo nome contenha, textualmente, o nome da língua. Por exemplo, Uzbek é falado no Uzbekistan pelo que Uzbek deve ser devolvido.
12. Nomes das línguas que contenham, textualmente, o nome de um país onde sejam faladas. Por exemplo, Japanese é língua de Japan, pelo que Japanese deve ser devolvido.

Grupo 3. Considere o documento XML `courses.xml` que acompanha a ficha.
Escreva expressões XPATH que devolvam:

1. As cadeiras com menos inscrições. (Note que há cadeiras sem inscrições)
2. As cadeiras com pelo menos dois docentes.
3. Os nomes das cadeiras dadas pelo Prof. Aiken.
4. Os docentes das cadeiras dado por Jerry Cain.
5. As cadeiras dadas por Jerry Cain, mas não por Mehran Sahami.
6. Os nomes das cadeiras dadas por docentes do departamento de linguística.
7. As cadeiras que têm como precedências cadeiras dadas noutra departamento.
8. Cadeiras dadas por docentes que também dão uma das cadeiras que é precedência.
9. As cadeiras que têm menos inscrições do que alguma das suas precedências.

Grupo 4. Defina um DTD que capture da forma mais restritiva possível a informação descrita no ficheiro XML `recipes.xml` que acompanha a ficha.

Note que:

- o documento XML deve ser válido de acordo com o DTD;
- se se impuser alguma restrição adicional no DTD, o documento deixa de satisfazer o DTD. Impor restrições adicionais significa, por exemplo, mudar uma expressão `element*` para `element+`, `element?` ou `element\verb`. Outro exemplo de impor restrições adicionais inclui eliminar uma alternativa numa expressão.