

## Integração de Dados

Licenciatura em Engenharia Informática: 2º ano - 2º semestre

2018/2019

### Trabalho Prático

### **Integração de Dados com XML**

**Nota prévia:** O enunciado é propositadamente vago, genérico e incompleto em alguns pontos. O que se pretende é que os alunos avaliem as várias opções existentes e escolham a que considerarem mais apropriada para cada uma das situações com que se depararem. Todas as escolhas devem ser referidas e devidamente justificadas no relatório a entregar.

## 1. OBJECTIVOS

Com este trabalho pretende-se criar um programa em Java composto por vários Wrappers que obtenham dados de fontes heterogéneas, distribuídas e autónomas e possibilitem ao utilizador a visualização dos dados de forma integrada.

O utilizador terá ainda a possibilidade de fazer pesquisas, acrescentar dados que respeitem os esquemas adoptados e gerar ficheiros com informação seleccionada.

Para a realização deste trabalho deve usar a Linguagem Java, Expressões regulares e o API JDOM2 para manipulação de XML estudado nas aulas práticas.

## 2. RESULTADOS DA APRENDIZAGEM

Com este trabalho prático pretende-se que se adquiram as seguintes competências:

- Saber analisar uma situação típica de Integração de Dados e apresentar propostas válidas para um modelo de integração funcional, eficaz e correcto;
- Capacidade de criação e manipulação de XML
- Utilização de expressões regulares
- Capacidade de realização de pesquisa de informação em ficheiros XML usando XPath ou XQuery
- Capacidade de efetuar transformações de ficheiros XML
- Capacidade de efetuar validação de ficheiros XML usando DTD e/ou XSD

### 3. DESCRIÇÃO DO TRABALHO

O objectivo do trabalho consiste na criação de uma aplicação integradora que apresente uma vista unificada de informação relativa a países, proveniente de várias fontes de fontes.

As fontes de dados a usar são:

- S1 - Detalhes sobre um país: [http://pt.wikipedia.org/wiki/nome\\_pais](http://pt.wikipedia.org/wiki/nome_pais)  
(por exemplo: <http://pt.wikipedia.org/wiki/Portugal>)
- S2 - [https://www.sport-histoire.fr/pt/Geografia/Paises\\_por\\_ordem\\_alfabetica.php](https://www.sport-histoire.fr/pt/Geografia/Paises_por_ordem_alfabetica.php)

Nota: **nome\_do\_país** é dado pelo utilizador

Outros links da wikipedia que pode explorar:

- Países do continente Europeu: <http://pt.wikipedia.org/wiki/Europa>
- Países do continente Africano: <http://pt.wikipedia.org/wiki/África>
- Países do continente Asiático: <http://pt.wikipedia.org/wiki/Ásia>
- Países do continente Oceânia: <http://pt.wikipedia.org/wiki/Oceania>
- Países do continente Americano: <http://pt.wikipedia.org/wiki/América>
- Lista de códigos ISO de países:  
[http://pt.wikipedia.org/wiki/Comparação\\_entre\\_códigos\\_de\\_países\\_COI,\\_FIFA,\\_e\\_ISO\\_3166](http://pt.wikipedia.org/wiki/Comparação_entre_códigos_de_países_COI,_FIFA,_e_ISO_3166)
- Lista códigos ISO de Moeda: [http://pt.wikipedia.org/wiki/ISO\\_4217](http://pt.wikipedia.org/wiki/ISO_4217)
- Lista de países por continente:  
[http://pt.wikipedia.org/wiki/Lista\\_de\\_Estados\\_soberanos\\_e\\_territórios\\_dependentes\\_por\\_continente](http://pt.wikipedia.org/wiki/Lista_de_Estados_soberanos_e_territórios_dependentes_por_continente)
- Lista de países válidos: [http://pt.wikipedia.org/wiki/Lista\\_de\\_Estados\\_soberanos](http://pt.wikipedia.org/wiki/Lista_de_Estados_soberanos)

Podem ser usadas outras fontes de dados, mas a sua utilização deve ser devidamente justificada. Se a estrutura das fontes de dados usadas não for equivalente às sugeridas acima, poderão existir penalizações na nota final.

- As duas fontes de dados **S1** e **S2** são heterogéneas, autónomas e distribuídas e contêm informação relevante sobre demografia, política, e outros atributos geográficos de um país.
- O objectivo do trabalho prático consiste em efectuar integração de dados provenientes de várias fontes de dados e integrar a informação num modelo global G contendo pelo menos dois ficheiros XML (**países.xml** e **factos.xml**). Pode optar por modelos que usem mais ficheiros, caso ache relevante.
- O ficheiro **países.xml** deve conter (pelo menos) a seguinte informação: código ISO do país, nome do país, continente a que pertence, nome do Presidente da República, link para a imagem da bandeira do país.
- O ficheiro **factos.xml** deve conter (pelo menos) a seguinte informação: código ISO do país, código telefónico, código da internet, capital, cidade mais populosa, hino, moeda, população, área e lista de idiomas oficiais.

Os ficheiros devem usar o código ISO-3 de um país para manterem a informação relacionada.

O esquema a adoptar na vista unificada deve ser decidido pelos alunos e validado usando o XSD e o DTD apropriado.

Depois de realizado o processo de integração dos dados, o utilizador poderá fazer pesquisas sobre a vista unificada.

## 4. TAREFAS A REALIZAR

Encontram-se em seguida as tarefas principais a desenvolver neste trabalho prático. As descrições são genéricas e os exemplos apresentados servem apenas para uma melhor compreensão do que é pretendido. Os alunos devem ser criativos e apresentar uma solução integradora completa e funcional que permita efectuar uma grande diversidade de pesquisas.

### 4.1. IMPLEMENTAR WRAPPERS

Implementar os *Wrappers* que permitam obter a informação relevante de cada fonte de dados. Estes *Wrappers* devem ser implementados usando expressões regulares. No relatório deve ser descrito detalhadamente cada um dos wrappers, indicando que informação é retirada por cada um deles da fonte de dados em que cada um opera.

Para cada atributo a encontrar, deve(m) ser seleccionada(s) a(s) fonte(s) de dado(s) relevante(s). No caso de encontrar inconsistências ou conflitos os alunos terão de propor uma solução.

Para saber como implementar os Wrappers deve analisar a fonte das páginas HTML onde vai procurar a informação.

Use a função *HttpRequest* dada nas aulas práticas para aceder às páginas e gravá-las em disco.

O número e a estrutura dos wrappers depende da forma e da quantidade de informação que se quer encontrar e deve ser analisada pelos estudantes.

A palavra de pesquisa introduzida pelo utilizador é sempre o nome do país.

### 4.2. FICHEIROS XML: ACRESCENTAR, EDITAR E ELIMINAR INFORMAÇÃO DOS FICHEIROS

Usando menus, ou outras formas de interacção, o interface da aplicação deve possibilitar ao utilizador:

- Adicionar um novo país, desde que este não exista no(s) ficheiro(s) XML. Se os ficheiros não existirem ainda, devem ser criados com a inserção do 1º país. A informação obtida pelos Wrappers deve ser usada para a criação dos ficheiros XML de acordo com os modelos descrito na secção 3.
- Eliminar um país.
- Editar/alterar alguns atributos dos ficheiros XML (alterar capital, alterar população, alterar um idioma, etc)

**NOTA:** A informação dos dois ficheiros deve estar coerente e relacionada usando um campo à escolha do aluno. A remoção de um país do ficheiro **países.xml** deve ter como consequência a remoção dos seus factos do ficheiro **factos.xml**. Após as alterações, o modelo deve ser validado, para garantir a sua integridade.

### 4.3. VALIDAR O MODELO G

Os ficheiros do modelo **G** devem ser validados usando os XSD/DTD escolhidos.

Esta tarefa deve ser feita usando o API JDOM2 dado nas aulas práticas.

#### 4.4. FAZER PESQUISAS XPATH

Permitir ao utilizador efectuar diferentes pesquisas:

- 1) Sobre os ficheiros XML criados
- 2) Sobre as fontes de dados originais
  - **Exemplo: obter detalhes de um país introduzido pelo utilizador**
    - País já está no ficheiro países.xml? → mostrar dados do país.
    - País não está no ficheiro? → procurar nas fontes de dados usando os wrappers, mostrar ao utilizador os dados obtidos e adicionar a nova informação aos ficheiros XML, respeitando o modelo G. Se o país não for encontrado, informar o utilizador do insucesso da pesquisa.
  - Os passos 1) e 2) devem ser aplicados a diferentes tipos de pesquisas:
  - Exemplos de pesquisas:
    - Procurar países de um dado continente
    - Procurar países que falem um determinado idioma
    - Introduzir um código ISO de um país e obter os dados do país respectivo
    - Qual o país do ficheiro com maior área?
    - Efectuar pesquisas que combinem dois ou mais atributos
      - ...
      - ...

As pesquisas devem ser definidas pelos alunos, devendo ser variadas e versáteis com combinação de um diferente número de atributos.

A escolha dos atributos de pesquisa deve ser escolhido pelo utilizador usando Interface Java.

Os resultados devem ser apresentados de forma atrativa e organizada usando Interfaces Java.

As pesquisas devem ser implementadas usando XPath.

#### 4.5 GERAR FICHEIROS DE OUTPUT

O programa deve possibilitar ao utilizador gerar ficheiros de resultados. Estes ficheiros devem ser transformações dos ficheiros XML da vista global.

Três transformações **obrigatórias**:

- Gerar um ficheiro HTML contendo as bandeiras de todos os países listadas no ficheiro XML;
- Gerar um novo ficheiro XML que junte atributos dos dois ficheiros.
- Gerar um novo ficheiro TXT que ofereça a lista dos 5 países mais populosos.

Os alunos devem propor no mínimo **mais três** transformações adicionais. Devem implementar as transformações usando XSLT / XQuery

### 5. NORMAS PARA REALIZAÇÃO DO TRABALHO

O trabalho deverá ser realizado **individualmente ou em grupos de dois alunos**.

O trabalho vale 6 valores e é necessário um mínimo de 35% para aprovação na Unidade Curricular.

O trabalho final deve ser entregue até **26 de Maio de 2019** às 23h55 GMT. Os trabalhos serão sujeitos a **defesa obrigatória** na aula prática da semana 27-31 de Maio. Neste dia, deverá ser entregue o relatório em papel.

## 6. CRITÉRIOS DE AVALIAÇÃO

O trabalho vale **6 valores** na nota final da Unidade Curricular.

Será avaliado segundo os seguintes critérios:

- Qualidade e correcção na implementação das tarefas solicitadas
- Funcionalidade do programa
- Originalidade e diversificação dos conteúdos abordados, nomeadamente as funcionalidades extras
- Justificação das opções tomadas
- Qualidade do relatório entregue
- Qualidade da defesa

**Bom trabalho!**  
**©2019 Anabela Simões**