

# TRABALHO PRÁTICO

**Integração de Dados com XML 2023/2024**

**Licenciatura em Engenharia Informática**

**Dinis Meireles de Sousa Falcão | [a2020130403@isec.pt](mailto:a2020130403@isec.pt)**

**Telmo Eduardo Fonseca Silva | [a2019127444@isec.pt](mailto:a2019127444@isec.pt)**

# ÍNDICE

<b>DESCRIÇÃO DO TRABALHO</b>	<b>2</b>
<b>TAREFAS REALIZADAS</b>	<b>3</b>
<b>DEFINIR O ESQUEMA GLOBAL (G)</b>	
<b>IMPLEMENTAR OS WRAPPERS (MAPEAMENTOS M)</b>	
<b>GERAR/MANIPULAR FICHEIROS XML</b>	
<b>VALIDAÇÃO DO MODELO G</b>	
<b>PESQUISAS XPATH</b>	
<b>XSLT/XQUERY</b>	
<b>INTERFACE GRÁFICO</b>	<b>9</b>
<b>OVERVIEW</b>	<b>11</b>

# DESCRIÇÃO DO TRABALHO

No âmbito da Unidade Curricular de **Integração de Dados** no ano letivo **2024/2025**, foi proposta a realização de um trabalho prático com o objetivo de criar uma aplicação de integração de dados que apresentasse uma visão unificada de informações relativas a países.

A informação foi extraída de três sites a seguir apresentados, tratada e integrada em ficheiros XML.

As fontes de dados usadas foram:

- **S1** - [https://en.wikipedia.org/wiki/nome\\_país](https://en.wikipedia.org/wiki/nome_país)
- **S2** - [https://www.countryreports.org/country/nome\\_país.htm](https://www.countryreports.org/country/nome_país.htm)
- **S3** - [https://www.countryreports.org/country/nome\\_país/geography.htm](https://www.countryreports.org/country/nome_país/geography.htm)

Sendo que “nome\_país” é dado pelo utilizador, em inglês.

Os dados foram provenientes destas fontes de dados e o modelo global **G** é composto por dois ficheiros XML que agregam a seguinte informação de forma organizada e coerente:

- Ficheiro **países.xml** contém: código ISO país, nome do país, nome oficial do país, continente a que pertence, nome do Presidente da República / Monarca, link para a imagem da bandeira do país.
- Ficheiro **factos.xml** contém: código ISO país, nome do país, nome oficial do país, capital, moeda, população, área em km<sup>2</sup>, crescimento da população, cidades mais populosas, domínio internet, lista de idiomas oficiais e lista de países vizinhos.

Os esquemas adotados nas vistas unificadas foram validados usando **XSD** e o **DTD** apropriados.

# TAREFAS REALIZADAS

## Definir o Esquema Global (G):

### países.xml

```
<países>
  <país códigoISO="PT" nome="Portugal">
    <nome_oficial>Portuguese Republic</nome_oficial>
    <continente>Europe</continente>
    <presidente>Marcelo Rebelo de Sousa</presidente>
    <bandeira>/upload.wikimedia.org/wikipedia/commons/thumb/5/5c/Flag_of_Portugal.svg/125px-Flag_of_Portugal.svg.png</bandeira>
  </país>
  <país códigoISO="DE" nome="Germany">
    <nome_oficial>Federal Republic of Germany</nome_oficial>
    <continente>Europe</continente>
    <presidente>Frank-Walter Steinmeier</presidente>
    <bandeira>/upload.wikimedia.org/wikipedia/en/thumb/b/ba/Flag_of_Germany.svg/125px-Flag_of_Germany.svg.png</bandeira>
  </país>
  <país códigoISO="RW" nome="Rwanda">
    <nome_oficial>Republic of Rwanda</nome_oficial>
    <continente>Africa</continente>
    <presidente>Paul Kagame</presidente>
    <bandeira>/upload.wikimedia.org/wikipedia/commons/thumb/1/17/Flag_of_Rwanda.svg/125px-Flag_of_Rwanda.svg.png</bandeira>
  </país>
</países>
```

### factos.xml

```
<factos>
  <país códigoISO="PT" nome="Portugal">
    <nome_oficial>Portuguese Republic</nome_oficial>
    <capital>Lisbon</capital>
    <moeda>Euro (EUR)</moeda>
    <populacao>10,302,674</populacao>
    <area>92,090</area>
    <crecimento>0.15%</crecimento>
    <idades_mais_populosas>
      <cidade>LISBON (capital) 2.843 million</cidade>
      <cidade>Porto 1.367 million</cidade>
    </idades_mais_populosas>
    <dominioInternet>.pt</dominioInternet>
    <idiomas_oficiais>
      <idioma>Portuguese (official)</idioma>
      <idioma>Mirandese (official - but locally used)</idioma>
    </idiomas_oficiais>
    <países_vizinhos>
      <país_vizinho>Portugal</país_vizinho>
    </países_vizinhos>
  </país>
  <país códigoISO="DE" nome="Germany">
    <nome_oficial>Federal Republic of Germany</nome_oficial>
    <capital>Berlin</capital>
    <moeda>Euro (EUR)</moeda>
    <populacao>80,159,662</populacao>
    <area>357,022</area>
    <crecimento>0.19%</crecimento>
    <idades_mais_populosas>
      <cidade>Hamburg 1.796 million</cidade>
      <cidade>Munich 1.364 million</cidade>
      <cidade>Cologne 1.006 million</cidade>
    </idades_mais_populosas>
    <dominioInternet>.de</dominioInternet>
    <idiomas_oficiais>
      <idioma>German</idioma>
    </idiomas_oficiais>
    <países_vizinhos>
      <país_vizinho>Austria</país_vizinho>
    </países_vizinhos>
  </país>
</factos>
```

## Implementar Wrappers (Mapeamentos M)

Como referido anteriormente, os ficheiros países.xml e factos.xml deviam conter algumas informações que deveriam ser encontradas com o recurso a **expressões regulares**, através da implementação de **Wrappers**.

Foram implementados **15 Wrappers** na totalidade, dos quais **1 extra**. Destes 15 Wrappers, 12 retornam **String** e os restantes 3 **ArrayList<String>**.

- String obtemCodigoISO(String nome\_pais);
  - Obtém o código ISO do país com o nome = nome\_pais.
- String obtemNome(String nome\_pais);
  - Obtém o nome do país com o nome = nome\_pais.
- String obtemNomeOficialPais(String nome\_pais);
  - Obtém o nome oficial do país com o nome = nome\_pais.
- String obtemContinente(String nome\_pais);
  - Obtém o continente do país com o nome = nome\_pais.
- String obtemPresidente(String nome\_pais);
  - Obtém o presidente ou monarca do país com o nome = nome\_pais.
- String obtemBandeira(String nome\_pais);
  - Obtém a bandeira do país com o nome = nome\_pais.
- String obtemCapital(String nome\_pais);
  - Obtém a capital do país com o nome = nome\_pais.
- String obtemMoeda(String nome\_pais);
  - Obtém a moeda do país com o nome = nome\_pais.
- String obtemPopulacao(String nome\_pais);
  - Obtém a população do país com o nome = nome\_pais.
- String obtemArea(String nome\_pais);
  - Obtém a área do país com o nome = nome\_pais.
- String obtemCrescimentoPopulacao(String nome\_pais);
  - Obtém o crescimento da população do país com o nome = nome\_pais.
- ArrayList<String> obtemCidadesMaisPopulosas(String nome\_pais);
  - Obtém as cidades mais populosas do país com o nome = nome\_pais.
- String obtemDominioInternet(String nome\_pais);
  - Obtém o domínio da internet do país com o nome = nome\_pais.
- ArrayList<String> obtemIdiomaOficial(String nome\_pais);
  - Obtém os idiomas do país com o nome = nome\_pais.
- ArrayList<String> obtemPaisesVizinhos(String nome\_pais);
  - Obtém os países vizinhos do país com o nome = nome\_pais.

Durante a realização destes Wrappers, notaram-se algumas situações de exceção, das quais conseguimos identificar: alguns países não apresentam presidente, sendo que

optámos por colocar o monarca nesses países; o país Denmark não apresentava o código ISO 3166 na página da Wikipédia; o Wrapper de idiomas oficiais não está a funcionar a 100% em alguns países; o Wrapper de países vizinhos apenas apanha um país vizinho.

Como também já foi referido anteriormente, a informação foi extraída de três sites, sendo especificada origem de cada informação aqui:

- [https://en.wikipedia.org/wiki/nome\\_país](https://en.wikipedia.org/wiki/nome_país)
  - Código ISO;
  - Nome;
  - Nome Oficial;
  - Presidente;
  - Bandeira;
  - Domínio Internet;
- [https://www.countryreports.org/country/nome\\_país.htm](https://www.countryreports.org/country/nome_país.htm)
  - Continente;
  - Capital;
  - Moeda;
  - População;
  - Área;
  - Crescimento da População;
  - Cidades mais Populosas;
  - Idiomas Oficiais.
- [https://www.countryreports.org/country/nome\\_país/geography.htm](https://www.countryreports.org/country/nome_país/geography.htm)
  - Países Vizinhos.

## Gerar/Manipular Ficheiros XML

Após a implementação dos Wrappers, foram realizadas criadas algumas opções de manipulação dos ficheiros XML:

- **Adicionar um novo país** e respetivos factos nos ficheiros XML, sendo que este não permite a repetição de países e, caso os ficheiros não estejam criados, são criados no ato de adicionar o primeiro país;
- **Eliminar um país** e os respetivos factos através do nome do mesmo;
- **Alterar um qualquer atributo de um país** através do nome do país e do nome do atributo;
- **Adicionar um qualquer elemento a uma lista de elementos de um país** através do nome do país e do nome da lista de elementos;
- **Alterar um qualquer elemento a uma lista de elementos de um país** através do nome do país e do nome da lista de elementos;

# Validação Do Modelo G

Como foi dito anteriormente, os esquemas adotados nas vistas unificadas foram validados usando **XSD** e o **DTD** apropriados, sendo que esta tarefa foi feita usando o **API JDOM2** dado nas aulas práticas da unidade curricular de Integração de Dados.

Os ficheiros utilizados nesta validação foram:

- **países.xsd**
- **países.dtd**
- **factos.xsd**
- **factos.dtd**

## factos.dtd

```
<!ELEMENT factos (pais*)>
<!ELEMENT pais (nome_oficial, capital, moeda, populacao, area, crescimento, cidades_mais_populosas?, dominioInternet, idiomas_oficiais, países_vizinhos?)>
<!ATTLIST pais codigoISO CDATA #REQUIRED>
<!ATTLIST pais nome CDATA #REQUIRED>
<!ELEMENT nome_oficial (#PCDATA)>
<!ELEMENT capital (#PCDATA)>
<!ELEMENT moeda (#PCDATA)>
<!ELEMENT populacao (#PCDATA)>
<!ELEMENT area (#PCDATA)>
<!ELEMENT crescimento (#PCDATA)>
<!ELEMENT cidades_mais_populosas (cidade*)>
<!ELEMENT cidade (#PCDATA)>
<!ELEMENT dominioInternet (#PCDATA)>
<!ELEMENT idiomas_oficiais (idioma*)>
<!ELEMENT idioma (#PCDATA)>
<!ELEMENT países_vizinhos (pais_vizinho*)>
<!ELEMENT pais_vizinho (#PCDATA)>
<!ATTLIST factos xmlns:xsi CDATA #IMPLIED>
<!ATTLIST factos xsi:noNamespaceSchemaLocation CDATA #IMPLIED>
```

## países.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="países">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="pais" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="pais">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="nome_oficial"/>
        <xs:element ref="continente"/>
        <xs:element ref="presidente"/>
        <xs:element ref="bandeira"/>
      </xs:sequence>
      <xs:attribute name="codigoISO" type="xs:string" use="required"/>
      <xs:attribute name="nome" type="xs:string" use="required"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="bandeira" type="xs:string"/>
  <xs:element name="presidente" type="xs:string"/>
  <xs:element name="continente" type="xs:string"/>
  <xs:element name="nome_oficial" type="xs:string"/>
</xs:schema>
```

## Pesquisas XPath

Além disto, também foi permitido ao utilizador a realização de diferentes pesquisas sobre os ficheiros XML, das quais:

- **Pesquisar pelo nome do país informação relevante** (nome, continente e presidente);
  - “//pais[@nome=" + nome\_pais + "]/concat(@nome, ', ', continente, ', ', presidente)”
- **Pesquisar nomes dos países de um determinado continente;**
  - “//pais[continente=" + continente + "]/@nome”
- **Pesquisar os factos de um determinado país;**
  - “//pais[@nome=" + nome\_pais + "]"
- **Qual o país mais populoso?;**
  - “//pais[number(translate(populacao, ',', '')) = max(//pais[number(translate(populacao, ',', '')))]/@nome”
- **Quais os países de um determinado continente, com população acima de um dado valor;**
  -

```
String xp1 = "//pais[continente = ' " + continente + "']/@nome";
XdmValue res1 = XPathFunctions.executaXpath(xp1, "países.xml");
String s1 = XPathFunctions.listaResultado(res1);

String[] array = s1.split("\n");

String xp = "//pais[number(translate(populacao, ',', '')) > " + valor + " and ";
for(String a : array)
    xp = xp + "@nome = ' " + a + " ' or ";

xp = xp.substring(0, xp.length() - 3);
xp = xp + "]/@nome";

XdmValue res = XPathFunctions.executaXpath(xp, "factos.xml");
String s = XPathFunctions.listaResultado(res);
```

Como podemos ver nesta anterior pesquisa, sendo que a informação do continente e da população se encontrava em **.xml** diferentes, foram realizadas duas pesquisas para obter o resultado. A primeira pesquisa guardou os seus resultados numa **string**, que foi posteriormente dividida num **array de strings**, de maneira a poder percorrer todos os países encontrados pertencentes ao continente escolhido.

- **EXTRA: Pesquisar países com população abaixo de um dado valor.**
  - “//pais[number(translate(populacao, ',', '')) < " + valor + "]/@nome”

Todos os dados necessários a estas pesquisas são introduzidos pelo utilizador.



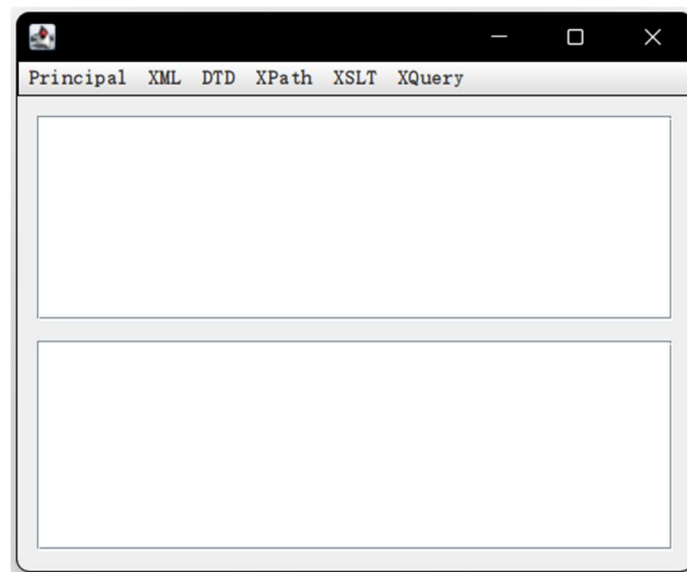
## XSLT/XQuery

Também foi possibilitado ao utilizador a geração de ficheiros de resultados, com recurso a ficheiros **XSLT** e **XQuery**. Estas transformações são as permitidas:

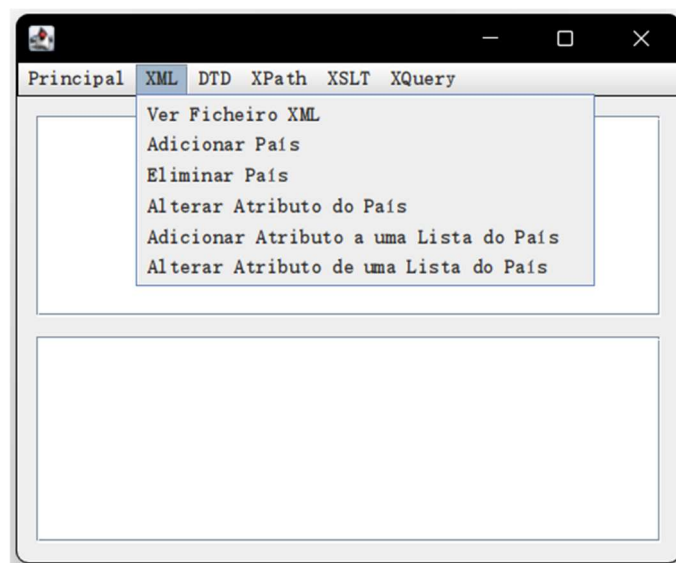
- **XQuery:**
  - **Gerar HTML com uma tabela** (nomes e bandeiras dos países);
    - **query1.xql**
  - **Gerar TXT que mostre a listagem de todos os países;**
    - **query2.xql**
  - **Indicar o nome de um continente e gerar ficheiro HTML de fotos das bandeiras dos países.**
    - **query3.xql**
- **XSLT:**
  - **Gerar ficheiro XML que faça a junção da informação de um país e dos respetivos factos;**
    - **trf2.xslt;**
  - **Gerar ficheiro XML com informação dos 5 países mais populosos.**
    - **trf1.xslt;**
- **EXTRA:**
  - **XSLT:**
    - **Gerar ficheiro XML com países ordenados alfabeticamente;**
      - **trf3.xslt;**
    - **Gerar ficheiro HTML com listagem de todas as cidades.**
      - **trf4.xslt**
  - **XQuery:**
    - **Gerar TXT com países com taxa de crescimento acima de um determinado valor.**
      - **query4.xql**

# INTERFACE GRÁFICO

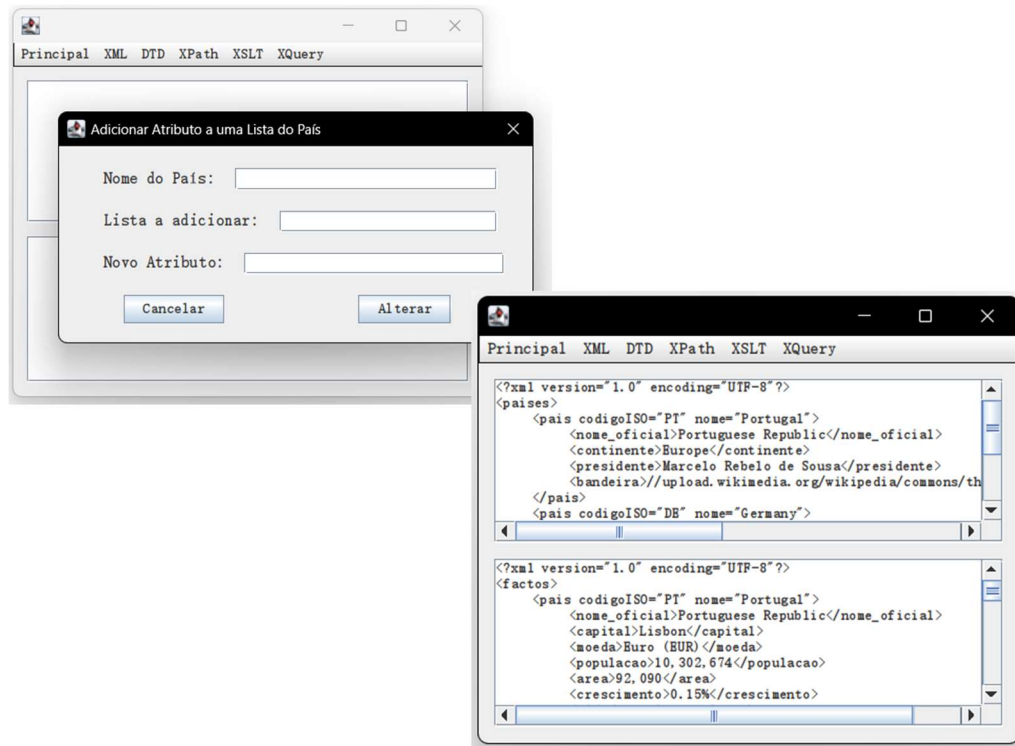
A aplicação desenvolvida apresenta uma interface gráfica que disponibiliza ao utilizador um conjunto de opções e ações que englobam tudo o que foi dito anteriormente. Esta aplicação foi desenvolvida com o objetivo de aplicar todos os conteúdos abordados durante as aulas, como podemos ver em seguida.



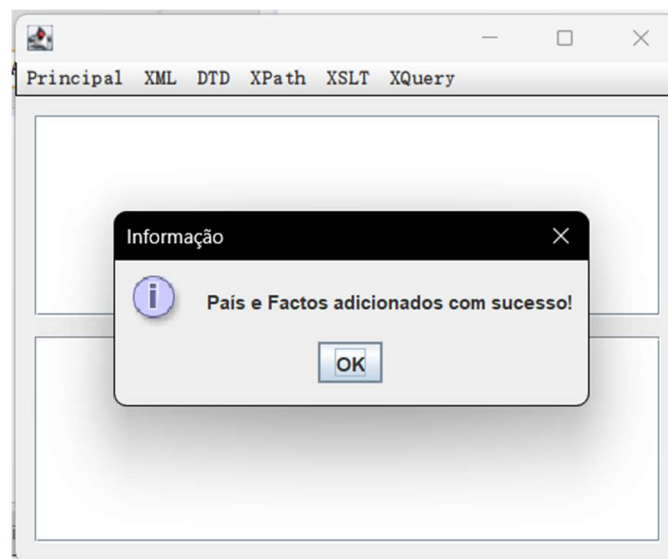
Como podemos ver através da imagem anterior, a aplicação possui 6 menus (**Principal**, **XML**, **DTD**, **XPath**, **XSLT** e **XQuery**) e 2 áreas de texto, sendo que cada menu possui itens como podemos ver na seguinte imagem.



Após a seleção de um destes itens, podem ser originadas algumas ações, como: aparecimento de uma janela Dialog, aparecimento de texto nas áreas de texto e geração de ficheiros (.txt, .html, .xml). Podemos ver algumas destas ações nas imagens seguintes.



Além disso, como referido anteriormente, com o objetivo de abordar os conteúdos lecionados, em algumas janelas Dialog, é originada uma janela a confirmar o sucesso da operação realizada, como podemos ver na imagem seguinte:



# OVERVIEW

Fazendo uma pequena análise do trabalho realizado, podemos concluir que foram realizadas todas as tarefas presentes no enunciado do trabalho prático com precisão, excetuando algumas irregularidades nos Wrappers para obtenção da lista de idiomas e da lista de países vizinhos.

Além disto, apenas foram utilizados conteúdos e funções lecionados nas aulas, complementados por aqueles realizados por nós. Considerando os pontos mencionados, acreditamos que desempenhámos este trabalho de maneira satisfatória e com um bom nível de qualidade.