



Licenciatura em Engenharia Informática
2022/2023

Relatório do Trabalho Prático de
Integração de Dados

Integração de Dados sobre Escritores

Coimbra, 11 de junho de 2023

Elaborado por:

Rodrigo Bernarda nº 2021136740 email a2021136740@isec.pt

Arlindo Martins nº 2021150822 email a2021150822@isec.pt

Índice

1. Introdução	3
2. Estrutura do Trabalho	4
3. Implementação dos Wrappers	5
4. Criação dos Ficheiros XML	6
5. Validar XSD e DTD	6
6. Interface	6
7. Conclusão	6

1. Introdução

O objetivo deste trabalho consiste na criação de uma aplicação integradora usando linguagem Java que seja composto por Wrappers e que apresenta uma vista unificada de informação relativa a diversos escritores e as suas respetivas obras.

Esta informação encontra-se nas seguintes fontes de dados: <https://pt.wikipedia.org/> e bertrand.pt, para que o utilizador possa ver, pesquisar, acrescentar dados que respeitem os esquemas adotados e gerar ficheiros com informação selecionada.

Usando estas fontes de dados, é suposto construir um modelo global G composto por um ficheiro XML que agregue a informação de forma organizada e coerente.

Criamos dois ficheiros, um chamado obras.xml que recolhe informações de obras escritas por certo autor no site da bertrand.pt e outro chamado escritores.xml que recolhe informações pessoais dos autores através do site <https://pt.wikipedia.org/>, e o terceiro ficheiro chamado global.xml que é o modelo global que junta as informações mais relevantes do escritores.xml e obras.xml.

Depois do processo de integração de dados estar realizado, e ser validado usando DTD e XSD, será possível ao utilizador fazer pesquisas e edições sobre a vista unificada.

2. Estrutura do Trabalho

O trabalho encontra-se dividido em diferentes ficheiros, são eles:

- EscritoresXML.java
- GlobalXML.java
- HttpRequestFunctions.java
- JDOMFUnctions_Validar.java
- JdomFunctions_XSLT.java
- Livro.java
- MainFrame.java
- SaxonFunctions_Xquery.java
- TPID2023.java
- Utils.java
- ValidarXML.java
- WrapperBertrand.java
- WrapperWikipedia.java
- XMLJDomFunctions.java
- XpathFunctions.java
- Efemeride.java
- Escritor.java
- Premio.java
- EscritorModeloXML.java

Usamos também as libraries:

- Jdom2-2.0.6.1.jar
- Saxon-9.jar
- Saxon9-s9api.jar

3. Implementação dos Wrappers

Wrapper da Wikipedia:

- Para cada componente que precisamos de encontrar usamos uma função específica, por exemplo, a função *escritor_dataNascimento(String nome)* devolve a data de nascimento do autor em causa, para isso chega a usar mais que 2 expressões regulares. O ficheiro que reúne todas as funções ligadas à extração de dados do site da wikipedia é o *WrappersWikipedia.java*. No geral, consegue encontrar todas as informações precisas para os escritores, só o autor Raymond Carver não conseguiu encontrar as informações necessárias por não ter sequer as informações que pretendemos.

Wrapper da Bertrand:

- Para o wrapper da bertrand criamos uma função *obtem_links(String autor)* que guarda todos os links das obras encontradas, depois as funções destinadas a extrair os dados pretendidos, como a *obtem_titulos(List<String> links)* e *obtem_autores(List<String> links)*, acedam os links das obras e extraem todos os dados precisos. Com os dados recolhidos procedemos depois a uma verificação se o autor dos livros recolhidos são do autor que procuramos. Estes wrappers funcionam completamente, porém não foi possível recolher mais que uma obra do autor Raymond Carver.

4. Criação do ficheiro XML

No ficheiro `EscritoresXML.java` estão as funções relacionadas com a criação do `obras.xml`, no ficheiro `EscritorModeloXML.java` estão as funções que criam o `escritores.xml` e no `GlobalXML.java` as funções que junta as informações do `obras.xml` e `escritores.xml` e cria o `global.xml`. Todas estas funções criam corretamente os ficheiros xml, com a informação recolhida pelos wrappers que criamos.

5. Validar DTD e XSD

No ficheiro `ValidarXML.java` temos as funções que ajudam a validar os documentos xml, quer por DTD, quer por XSD. Os ficheiros `global.dtd`, `obras.dtd`, `escritores.dtd`, `global.xsd`, `escritores.xsd` e `obras.xsd` são os ficheiros responsáveis pela validação dos respetivos ficheiros xml.

6. Interface

A Interface deste trabalho permite ver o conteúdo dos ficheiros XML, alterar dados como acrescentar e remover escritores, pesquisar informações dentro do `global.xml` como obras de um certo autor e livros de uma certa editora com um preço superior a um dado pelo utilizador. Também é possível fazer exportações XSLT como devolver um ficheiro html com as fotografias e nomes dos autores e exportações XQuery, como mostrar a lista de autores.

7. Conclusão

Com este trabalho, conseguimos aperfeiçoar os nossos conhecimentos de Integração de Dados, nomeadamente no que diz respeito à utilização de Wrappers, gerar ficheiros XML, validações DTD e XSD e exportações XSLT e Xquery.