

## Ficha de Trabalho nº 4

### Extração de dados da Web: Wrappers

#### 1. Introdução

Neste ficha de trabalho pretende-se que implementem *wrappers* para extrair dados das sites web usando a função `HttpRequest`. No Moodle encontra um ficheiro com a implementação desta função:

- **`HttpRequestFunctions.java`**: para aceder aos dados/source de um determinado site

#### 2. Wrappers: extrair dados da web, wikipedia

No seu Browser, aceda a <https://pt.wikipedia.org/>

Vamos extrair deste site alguns dados sobre escritores. Por exemplo, faça a pesquisa de Fernando Pessoa e veja o resultado. Analise o *source* do ficheiro.



**Fernando Pessoa**

Esta página cita fontes, mas que não cobrem todo o conteúdo. Ajude a inserir referências. Conteúdo não verificável pode ser removido.—Encontre fontes: [NSW](#) · [CAPES](#) · [Google](#) (N · L · A) (Abrir de 2017)

**Fernando António Nogueira Pessoa** (Lisboa, Mártires, 13 de junho de 1888 – Lisboa, Santa Catarina, 30 de novembro de 1935)<sup>[a]</sup> foi um poeta, filósofo, dramaturgo, ensaísta, tradutor, publicitário, astrólogo, inventor, empresário, correspondente comercial, crítico literário e comentarista político português.

Fernando Pessoa já foi considerado por especialistas de sua obra como o mais universal poeta português.<sup>[b][c]</sup> Por ter sido educado na *África do Sul*, numa escola católica irlandesa de *Durban*, chegou a ter maior familiaridade com o idioma inglês do que com o português ao escrever os seus primeiros poemas nesse idioma. O crítico literário *Harold Bloom* considerou Pessoa como "*Whitman renascido*"<sup>[d]</sup> e incluiu-o no seu cânone entre os 26 melhores escritores da civilização ocidental,<sup>[e]</sup> não apenas da literatura portuguesa mas também da inglesa.<sup>[f]</sup>

Das quatro obras que publicou, três são na *língua inglesa* e apenas uma em *língua portuguesa*, intitulada *Mensagem*.<sup>[g][h]</sup> Fernando Pessoa traduziu várias obras em inglês (e.g., de *Shakespeare* e *Edgar Allan Poe*) para o português, e obras portuguesas (nomeadamente de *António Botto*<sup>[i]</sup> e *Almada Negreiros*) para o inglês e o francês.

Enquanto poeta escreveu sob diversas personalidades – a que ele próprio chamou *heterónimos*, como *Ricardo Reis*, *Álvaro de Campos* e *Alberto Caeiro* –, sendo estes últimos objeto da maior parte dos estudos sobre a sua vida e obra. *Robert Hass*, poeta americano, diz: "outros modernistas como *Yeats*, *Pound*, *Eliot* inventaram máscaras pelas quais falavam ocasionalmente... Pessoa inventava poetas inteiros".<sup>[j][k]</sup> Buscou também inspirações nas obras

**Fernando Pessoa**

Fernando Pessoa em 1914

<b>Nome completo</b>	Fernando António Nogueira Pessoa
<b>Outros nomes</b>	Alberto Caeiro, Álvaro de Campos, Ricardo Reis, Bernardo Soares etc.
<b>Nascimento</b>	13 de junho de 1888 Lisboa

Crie um novo projeto **Ficha4**.

Na *package* **ficha4** crie uma Classe **Escritor** com os seguintes campos:

**nome** – **String**  
**nome\_completo** – **String**  
**nacionalidade** – **String**  
**nascimento** – **String**  
**fotografia** – **String**  
**ocupações** – **ArrayList**

Acrescente à classe o construtor, getters e setters.

- a) Na *package* `ficha4` crie um novo ficheiro de nome **Wrappers.java** e implemente 5 wrappers que, dado o nome de um escritor, obtenham do site o seu nome completo, nacionalidade, data de nascimento, o link da fotografia e a lista de ocupações.
- **public static String obter\_nome\_completo(String escritor)**
  - **public static String obter\_nacionalidade(String escritor)**
  - **public static String obter\_dnascimento(String escritor)**
  - **public static String obter\_fotografia(String escritor)**
  - **public static ArrayList<String> obter\_ocupacoes(String escritor)**

Em cada Wrapper deve chamar a função `HttpRequest` e obter o source de cada escritor:

```
String link = "https://pt.wikipedia.org/wiki/"
String escritor = "Fernando Pessoa";
HttpRequestFunctions.httpRequest1(link, escritor, "escritor.html");
```

As ERs devem ser aplicadas sobre o ficheiro **escritor.html**

- b) Implemente a função **criaEscritor** que chama as funções anteriores e agrega os dados recolhidos num objeto da classe **Escritor**

- **public static Escritor criaEscritor (String escritor)**

- c) No *main* teste as funções anteriores.  
Experimente os wrappers para outros escritores:

Paulo Coelho  
José Saramago  
Mário Cesariny, ...

corrija eventuais erros.