Instituto Superior de Engenharia de Coimbra

Departamento de Informática

No âmbito da disciplina de Integração de Dados

Componente prática

Leccionada por: Filipe Sá

Relatório Trabalho Pratico



Realizado por:

Bruno Sousa nº21260212 TPSI

Manuel Brás nº21260202 TPSI

C:\Users\tiago\Desktop\white.pngTiago Correia nº21260205 TPSI

Indice

[1. Expressões Regulares 3](#_Toc504172432)

[1.1 – dre.pt 3](#_Toc504172433)

[1. Procura Titulo 3](#_Toc504172434)

[2. Procura Titulo Noticia 3](#_Toc504172436)

[3- Procura data da noticia 3](#_Toc504172438)

[4- Procura Autor 4](#_Toc504172440)

[5- Procura Descrição 4](#_Toc504172442)

[6- Procura Link da Noticia 4](#_Toc504172444)

[1.2 – anmp.pt 5](#_Toc504172446)

[1- Procura código do município 5](#_Toc504172447)

[2- Procura o nome do presidente da câmara 5](#_Toc504172449)

[3- Procura email do município 5](#_Toc504172451)

[4- Procura site do município 5](#_Toc504172453)

[5- Procura número de telefone 6](#_Toc504172455)

[6- Procura número de freguesias por município 6](#_Toc504172457)

[7- Procura número de freguesias por município 6](#_Toc504172459)

[8- Procura número de habitantes do município 6](#_Toc504172461)

[9- Procura número a imagem do brasão do município 6](#_Toc504172463)

[2. Classes 7](#_Toc504172465)

[2.1– Noticia 7](#_Toc504172466)

[1- Variáveis 7](#_Toc504172467)

[2.2– Município 7](#_Toc504172469)

[1- Variáveis 7](#_Toc504172470)

[2.2– Município 8](#_Toc504172472)

[1- Variáveis - continuação 8](#_Toc504172473)

[2.3 – Webpage 1 8](#_Toc504172475)

[1- Sumario 8](#_Toc504172476)

[2.4– Webpage2 10](#_Toc504172477)

[1- Função ProcuraSite2 10](#_Toc504172478)

[2- Função procuraPageMunicip 10](#_Toc504172479)

[3- Função Exemplo para procura de dados 11](#_Toc504172480)

[4- Função retirar municípios duplicados 12](#_Toc504172481)

[5-Função para formatar os nomes dos presidentes 13](#_Toc504172482)

[2.4 – XML Maker 14](#_Toc504172483)

[1- Função Exemplo para criar XML 14](#_Toc504172484)

[2- Função para guardar a chave de pesquisa 15](#_Toc504172485)

[2.5– XML to Array 16](#_Toc504172486)

[1- Função Noticias para array 16](#_Toc504172487)

[2- Função Municípios para array 17](#_Toc504172488)

[3- Função top 5 pesquisas 18](#_Toc504172489)

[1- XQuery Autor 19](#_Toc504172490)

[2- Xquery Tipo 19](#_Toc504172491)

[3- Xquery Câmara 20](#_Toc504172492)

[4- Xquery Presidente 20](#_Toc504172493)

[5- Municipios 21](#_Toc504172494)

[5.1- Xquery Autores das noticias 21](#_Toc504172495)

[5.2- Xquery Autores das noticias count 21](#_Toc504172496)

[5.3- Xquery Autores das noticias count Desc 22](#_Toc504172497)

[5.4- Xquery top 5 municípios com mais publicações 22](#_Toc504172498)

[2- Painel inicial (Data) 23](#_Toc504172499)

[3- Menu de pesquisas 24](#_Toc504172500)

[Este menu serve para alterar parâmetros da pesquisa. 24](#_Toc504172501)

[4- Painel para pesquisa de noticias 24](#_Toc504172502)

[6- Painel mostrar notícias 26](#_Toc504172503)

[7-Painel mostra Municípios 26](#_Toc504172504)

[8- Painel mostra municípios com mais noticias na data pesquisada 27](#_Toc504172505)

# 1. Expressões Regulares

## – dre.pt

## Procura Titulo



## Esta ER é composta pela iniciação “title=\”” que foi apenas usado para que possamos encontrar com mais facilidade os títulos pretendidos. Isto é seguido de ([a-z A-Z 0-9 À-ú\\s\\(\\)]+) que irá procurar um grupo de caracteres que podem ser letras maiúsculas e minúsculas, números de 0 a 9, espaços e parênteses. Todos estes poderão aparecer mais do que uma vez devido ao “+”, sendo que terá que aparecer pelo menos um deles. Isto é seguido de “n.º” para se saber onde terminar a pesquisa apesar de não aparecer na pesquisa em si.

## Procura Titulo Noticia



## Esta ER é composta pela iniciação “title=\”” que foi apenas usado para que possamos encontrar com mais facilidade os títulos pretendidos. Isto é seguido de ([a-z A-Z À-ú\\s\\(\\)]+) que irá procurar um grupo de caracteres que podem ser letras maiúsculas e minúsculas, espaços e parênteses. Todos estes poderão aparecer mais do que uma vez devido ao “\*”, sendo que terá que aparecer pelo menos um deles. Isto é seguido de “\\sn.º\\s” que foi apenas usado para que possamos encontrar com mais facilidade os títulos pretendidos. Isto é seguido de (\\d{1,5}(\\-[A-Z])?/\\d{4}) que é o grupo que realmente pretendemos retirar com esta ER. Neste pedaço da ER procuramos qualquer numero de 0 a 9 entre uma a 5 vezes, seguido da possibilidade de existir uma “\” com um carácter, que é seguido por 4 números de 0 a 9.

## Procura data da noticia



## Esta ER é composta pela iniciação (19[0-9]{2}|20[0-9]{2}) que foi usado para que possamos encontrar o ano da publicação. Isto é seguido de [\\-|\\/] que irá procurar um hífen “-” ou uma barra “/”. Isto é seguido de (0[1-9]|1[0-2]) que foi apenas usado para que possamos encontrar com mais facilidade os títulos pretendidos. Isto é seguido de [\\-|\\/]([1,2][0-9]|3[0,1]|0[1-9]))</a> que é constituído novamente um hífen “-” ou uma barra “/” seguido de ([1,2][0-9]|3[0,1]|0[1-9]), que irá procurar o dia. Finalmente, a tag </a> serve para sinalizar o final da pesquisa uma vez que existiam varias datas que a ER iria procurar se não a incluíssemos.

## Procura Autor



## Esta ER é composta pela iniciação "class=\"author\">\\s\*” que foi usado para que possamos encontrar o autor da publicação sendo que, irá procurar pela class indicada seguido de um numero de espaços. Isto é seguido de ([A-Z a-z À-ú\\s]+) que irá procurar uma qualquer disposição de letras e espaços, incluindo caracteres especiais. Esta segunda parte é a que realmente contem o nome do autor e é a única que pretendemos retirar.

## Procura Descrição



## Esta ER é composta, no inicio e no fim, por uma tag de paragrafo”<p>” e </p>, respectivamente, uma vez que a descrição é, em todo o texto, o único pedaço de texto dentro destas tags. Em seguida procuramos quaisquer caracteres que iram constituir a descrição.

## Procura Link da Noticia



## Esta ER é composta, no inicio, por <a href=\"(<https://dre.pt/web/guest/pesquisa-avancada/-/asearch/1>, que nos ajuda a identificar o local de onde iremos retirar uma vez que a descrição é, em todo o texto, o único pedaço de texto dentro destas tags. Em seguida procuramos quaisquer caracteres que iram continuar o link. No fim, adicionamos “\" title" de modo a ficarmos a saber onde termina a pesquisa.

## – anmp.pt

## Procura código do município



## Esta ER é composta, no inicio, por <OPTION value\\=\"mun101w3\\.php\\?cod\\=, seguido de (M[0-9]{4}) que vai procurar um “M” seguido de 4 números entre 0 e 9. Isto representa o código do município. De seguida concatenamos com o “Municip”, que é igual ao nome do município, e que é concatenado com [\\b](file:///\\b) que indica que termina naquele local devido a encontrar um espaço.

## Procura o nome do presidente da câmara



## Esta ER é composta, no inicio, por ([a-z A-Z À-ú \\s]+), que vai procurar um conjunto de letras e espaços, seguido “, Presidente da Câmara" que serve apenas para auxiliar a pesquisa.

## Procura email do município



## Esta ER é composta, no inicio, por "mailto\\:” que vai indicar-nos onde começar a pesquisa, e é seguido por ([a-z A-Z Á-ú] que vai iniciar a procura pela expressão que pretendemos encontrar. Este pedaço vai procurar por um conjunto de letras, seguido por um @. Em seguida, voltamos a procurar por um conjunto de letras ou hífenes. Logo em seguida temos um ponto “.” e um conjunto de letras novamente.

## Procura site do município

## C:\Users\Utilizador\Desktop\Prints\2-4.PNG

## Esta ER é composta, no inicio, por “(<http://www>.” que vai iniciar a pesquisa do website, e é seguido por ([a-z A-Z \\-] que vai iniciar a procurar o inicio do website composto por um conjunto de letras ou hífenes. Isto será seguido de um conjunto de letras que ira proporcionar o final do website.

## Procura número de telefone



## Esta ER é composta, no inicio, por "Telefone\\:” que vai indicar onde começar a pesquisa do numero que realmente queremos retirar. Em seguida procuramos por um conjunto de números entre 0 e 9, em conjunto com espaços, neste caso, 2 espaços entre conjuntos de números (xxx xxx xxx).

## Procura número de freguesias por município



## Esta ER é composta, no inicio, por "([0-9]+)” que vai procurar por um conjunto de números. Em seguida procuramos por “freguesia[s]?"”, que vai indicar onde devemos parar a procura.

## Procura número de freguesias por município



## Esta ER é composta, no inicio, por " [0-9\\,]+”, que vai procurar por um conjunto de números que pode conter uma virgula. Em seguida procuramos por “km2”, que vai indicar onde devemos parar a procura.

## Procura número de habitantes do município



## Esta ER é composta, no inicio, por " [0-9]\*”, que vai procurar por um conjunto de números vai corresponder aos milhares. Em seguida procuramos por um espaço, “\\s”, e novamente por 1 ou mais números. Para finalizar, procuramos por “habitante[s]” que vai finalizar a pesquisa.

## Procura número a imagem do brasão do município



## Esta ER é composta, no inicio, por “(SRC=\"../../../image/munap/”, que vai iniciar a pesquisa. Em seguida procuramos por um conjunto de letras ou numeros que podem aparecer uma vez ou mais, “([a-z A-Z Á-ú 0-9]+”. Para finalizar, procuramos por “+b.jpg)]” que vai finalizar a pesquisa.

# 2. Classes

## 2.1– Noticia

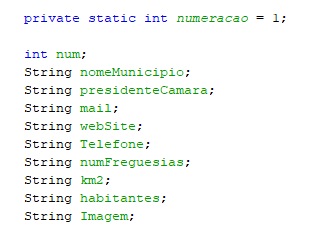
## Variáveis

## C:\Users\Utilizador\AppData\Local\Microsoft\Windows\INetCache\Content.Word\variaveis.png

## A “numeracao” é uma variável que nos permite fazer um controlo de quantas noticias guardamos. Esta numeração será incrementada em 1 cada vez que guardamos uma notícia, na variável num. A variável “tipo” irá guardar o tipo da noticia. A variável “autor”, irá guardar a o autor da notícia. A variável “discricao” irá guardar a discrição da notícia. A variável “link” irá guardar a totalidade do lik da noticia e a variável “numPub” irá guardar o numero da publicação.

## 2.2– Município

## Variáveis



## A “numeracao” é uma variável que nos permite fazer um controlo de quantos municípios guardamos. Esta numeração será incrementada em 1 cada vez que guardamos uma notícia, na variável num. A variável “tipo” irá guardar o tipo do município.

## 2.2– Município

## Variáveis - continuação

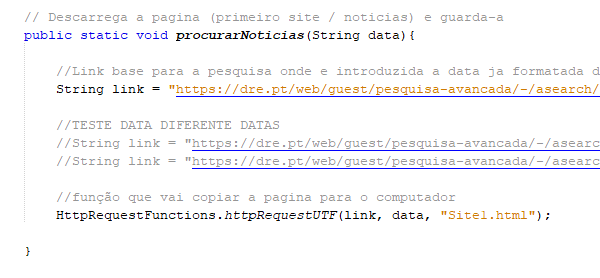
## A variável “nomeMunicipio”, irá guardar o nome do municipio. A variável “presidenteCamara” irá guardar o nome do Presidente da câmara em questão. A variável “mail” irá guardar o email do município. A variável “website” irá guardar o website do município. A variável “Telefone” irá guardar o telefone da câmara do município. A variável “numFreguesias” irá guardar o número de freguesias pertencentes ao município. A variável “km2” irá guardar o número de km2 do município. A variável “habitantes” irá guardar o número de habitantesdo município. A variável “imagem” irá guardar o link directo para a imagem, online.

## 2.3 – Webpage 1

## Sumario

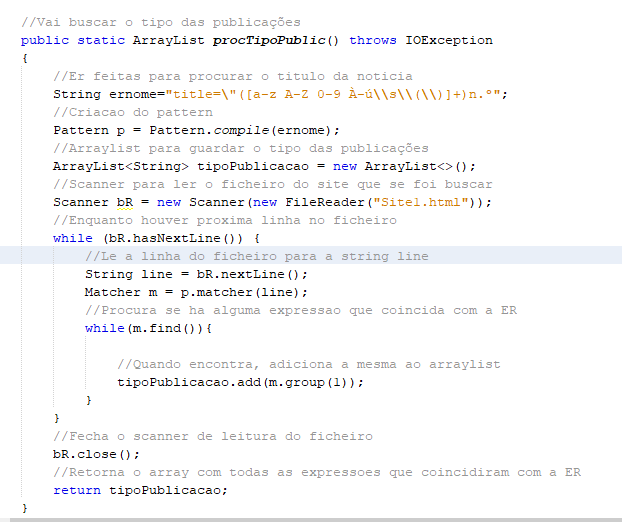
Esta classe irá aceder ao primeiro website e retira toda a informação que necessitamos, além de o guardar em disco.

1. Função procurarNoticias

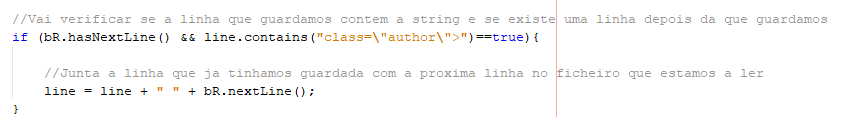


Esta função irá receber um string com a data já formatada, concatenando-a com o resto do link que depois irá usar para procurar e guardar a pagina.

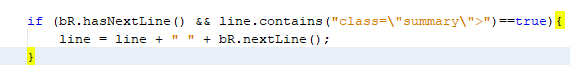
1. Função Exemplo para procura de dados



Esta função irá ser a base para todas as outras que se seguem na classe. Em termos de estrutura, a única coisa que muda para além das ER’s, para a procura dos dados que queremos são:



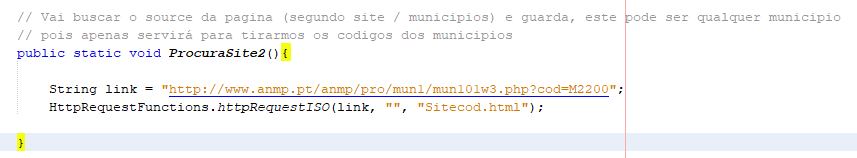
Na função da procura de autor, incluímos este IF que vai verificar se a linha que guardamos contem a string e se existe uma linha após esta. Caso isto se aplique, irá concatenar a string com essa próxima linha.



E na função da descrição em que incluímos este IF que vai verificar se a linha que guardamos contem a string e se existe uma linha depois da que guardamos. Caso isto se aplique, irá concatenar a string com essa próxima linha.

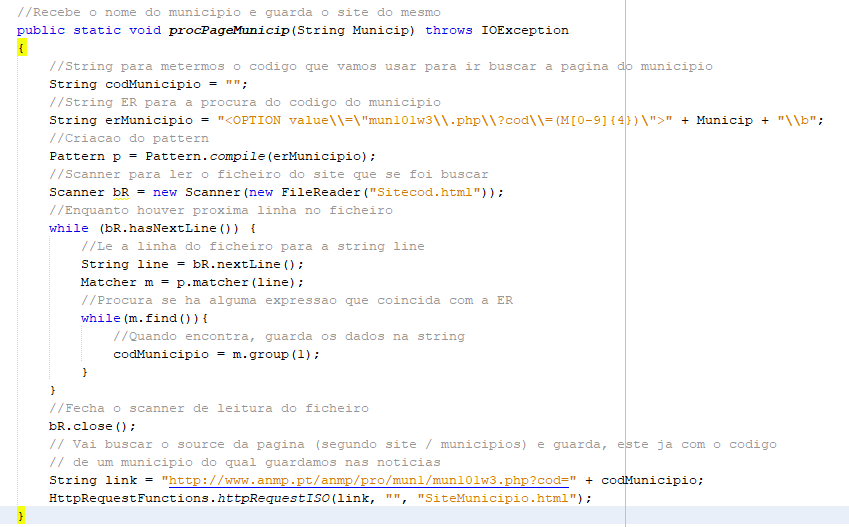
## 2.4– Webpage2

## Função ProcuraSite2



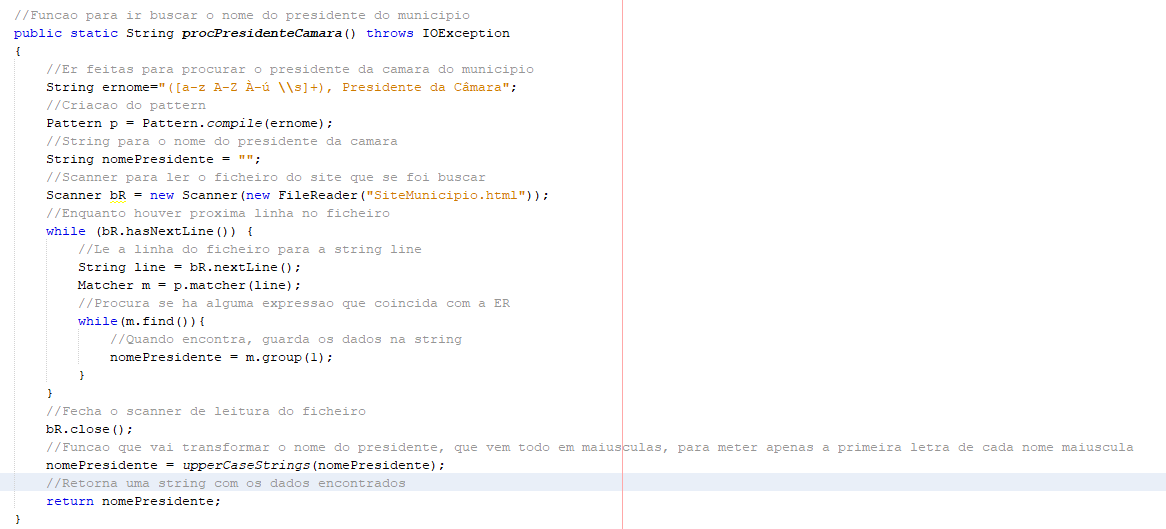
Esta função permite-nos obter o segundo site, não interessando o município, pois, apenas servirá para obter os códigos dos municípios presentes nas notícias.

## Função procuraPageMunicip

****

Esta função irá receber uma string com o nome do município, já formatado, de modo a ser possível procurar o seu código do site que guardamos anteriormente. Depois de obter o código, este vai ser concatenado com o link incompleto de modo a podermos guardar a pagina do município com as informações que pretendemos.

## Função Exemplo para procura de dados

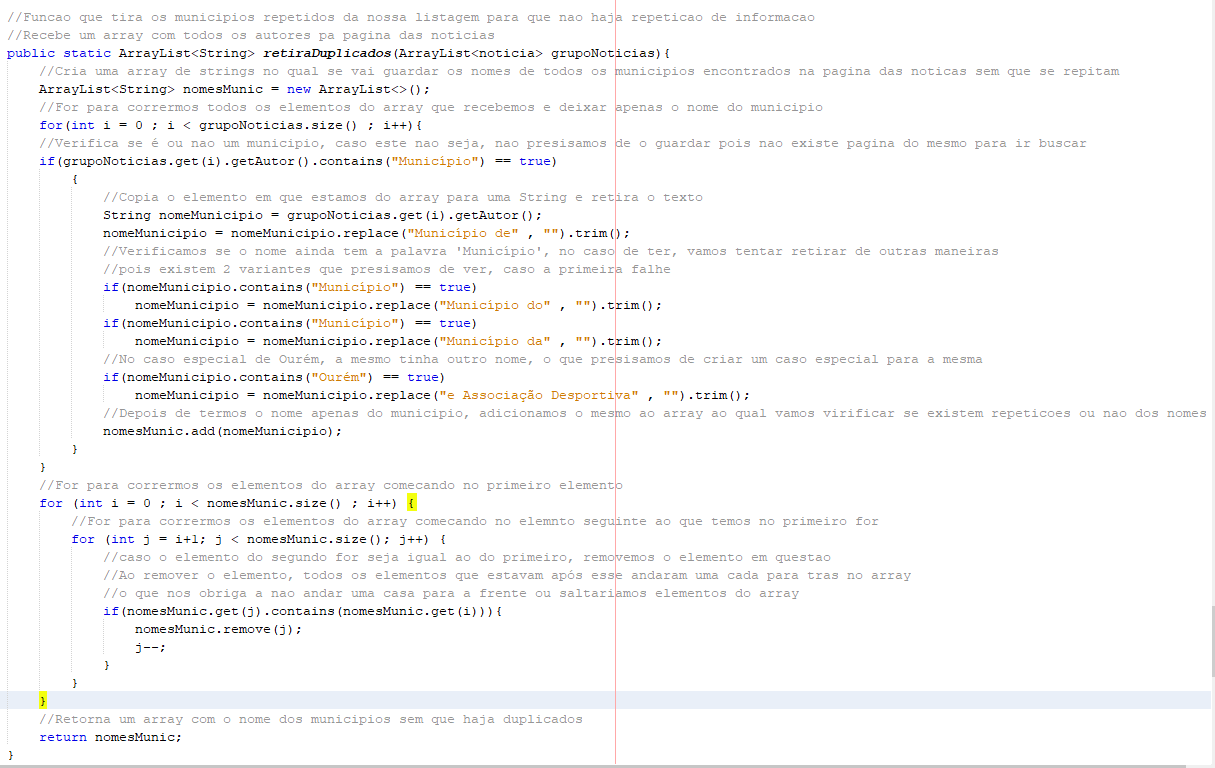


Esta função irá ser a base para todas as outras que se seguem na classe. Em termos de estrutura, a única coisa que muda para além das ER’s, para a procura dos dados que queremos é:



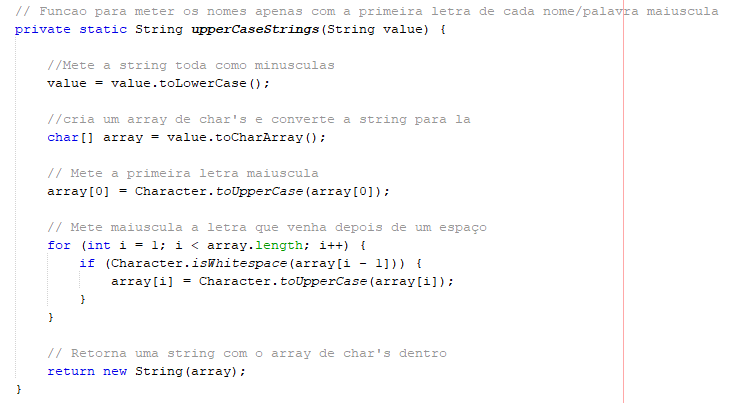
Este pedaço de código, vai substituir, no link da imagem, quando guardamos o source que vem nos dados encontrados pelo link online, para que tenhamos acesso à imagem.

## Função retirar municípios duplicados



Esta função, recebe um array com todos os autores que escreveram noticiam na data pesquisada. Utilizando esta informação fornecida pelo array, irá reformular cada posição do array de modo a esta ficar apenas com o nome do município e guardará essa informação num novo array. Usando este novo array com os dados formatados, faz-se uma verificação para apagar quaisquer duplicados no array. No fim, retornamos o mesmo para este ser usado para ir buscar as páginas dos municípios.

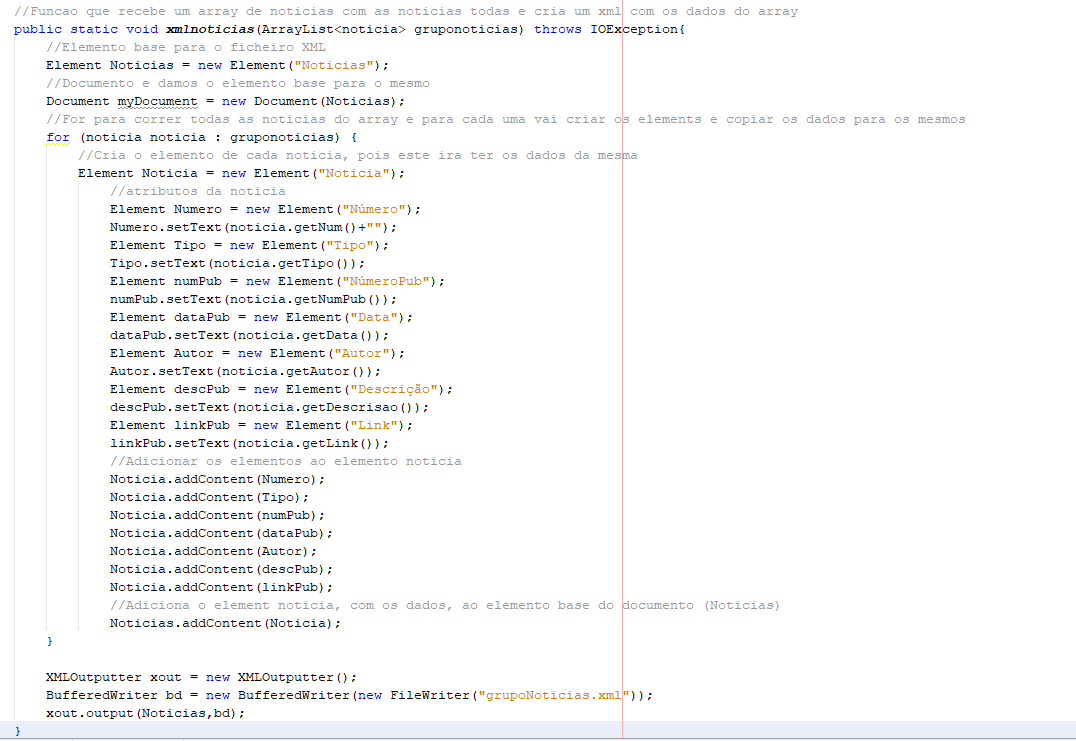
## 5-Função para formatar os nomes dos presidentes



Esta função recebe uma string com o nome do presidente e vai fazer com que todas as letras do nome fiquem minúsculas. Em seguida convertemos a string num array, no qual, sempre que houver um espaço antes de uma letra, a mesma vai ser convertida em maiúscula, utilizando para isto um ciclo “for”. No final, retornamos o array em forma de string.

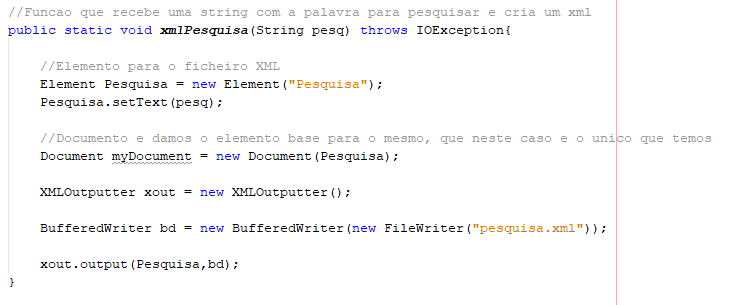
## 2.4 – XML Maker

## Função Exemplo para criar XML



Esta função irá ser a base para todas as outras que se seguem. Em termos de estrutura, serão muito similares.

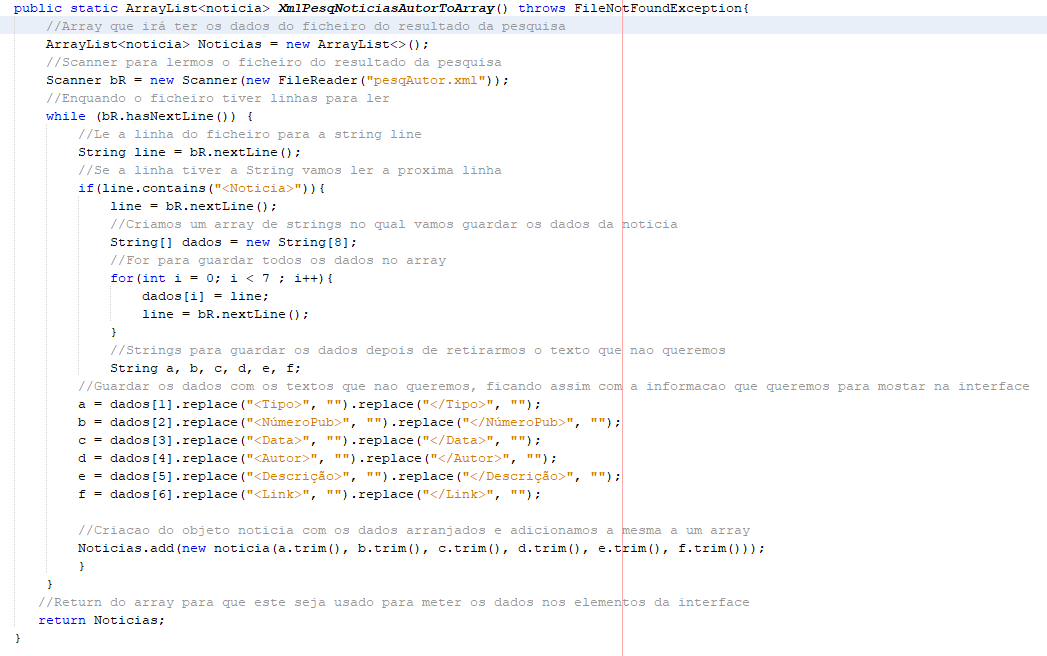
## Função para guardar a chave de pesquisa



Esta função irá fazer uma pesquisa num documento XML. Recebemos uma string, a qual vai ser usada como palavra-chave para a pesquisa. Abrimos o documento, e escrevemos o que encontramos com a palavra-chave, no documento “pesquisa.xml”.

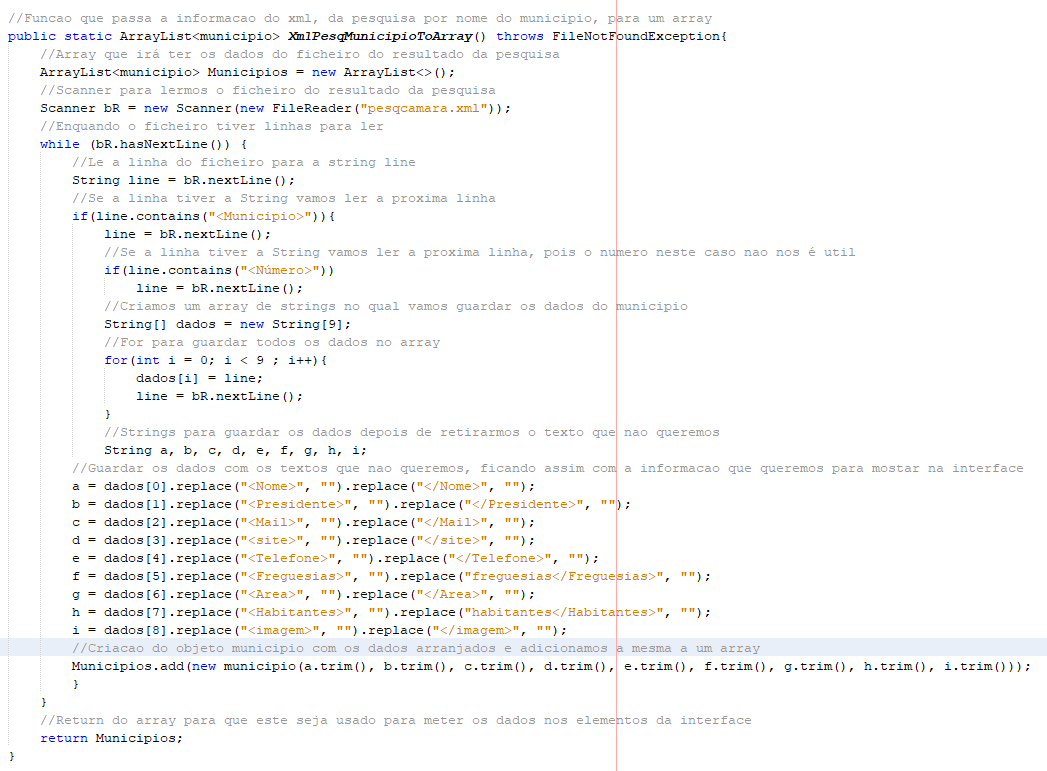
## 2.5– XML to Array

## Função Noticias para array



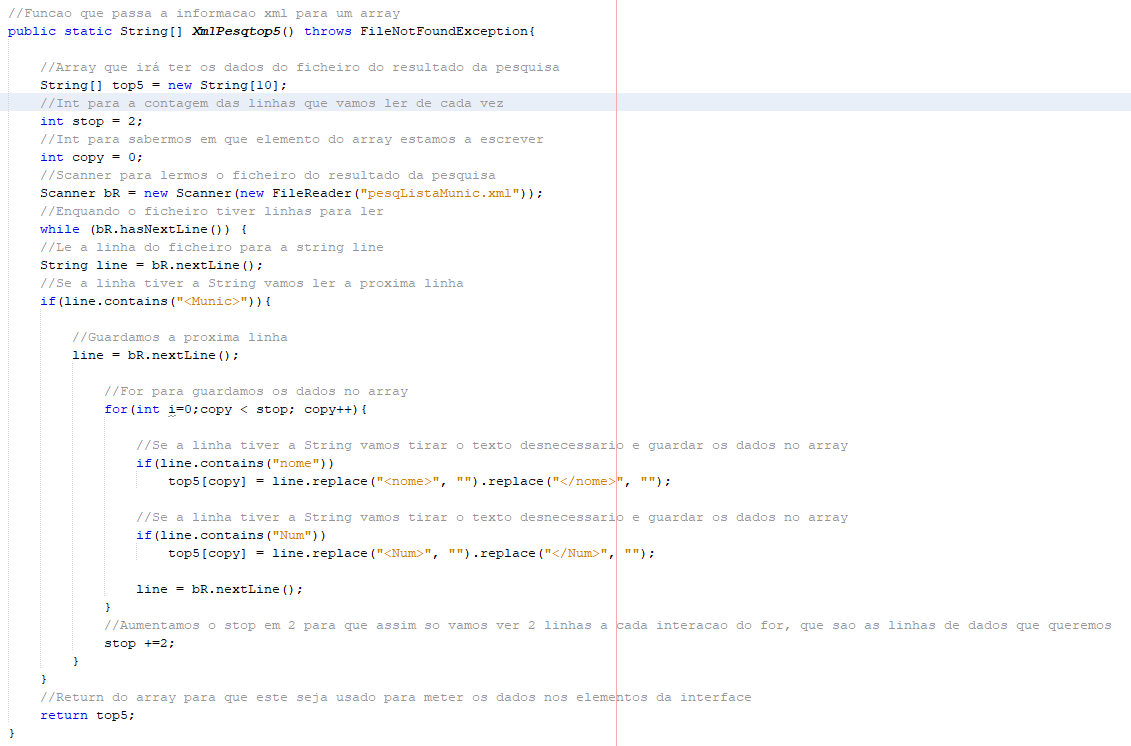
Esta função irá passar toda a informação do ficheiro XML criado pela pesquisa por notícias para um array, de modo a que este seja mostrado, formatado, na interface gráfica. Esta função é usada tanto na pesquisa por autor, como por tipo de notícia.

## Função Municípios para array



Esta função irá passar toda a informação do ficheiro XML criado pela pesquisa por município(s) para um array, de modo a que este seja mostrado, formatado, na interface gráfica. Esta função é usada tanto na pesquisa por nome do município, como por presidente da câmara.

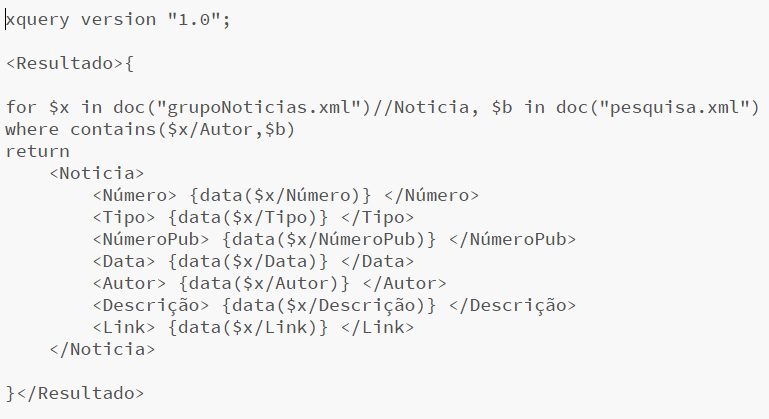
## Função top 5 pesquisas



Esta função irá passar toda a informação do ficheiro XML criado pela pesquisa dos 5 municípios que tenham mais noticias publicadas na data pesquisada.

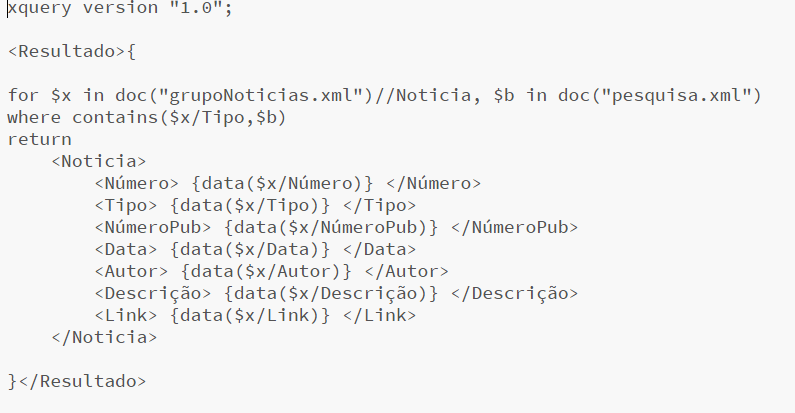
**2.5 – Xquery**

## XQuery Autor



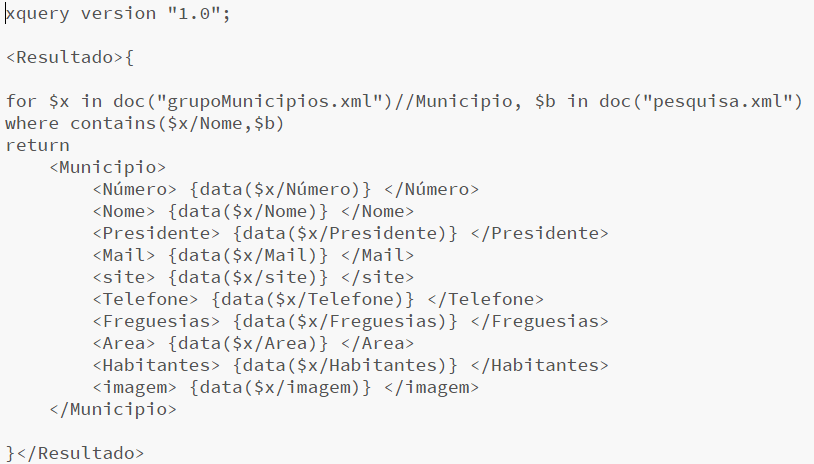
Esta query, procura no ficheiro XML das notícias se alguma delas tem a chave de pesquisa do XML em que esta guardada. Devolve as notícias que contenham a chave no seu conteúdo. Esta pesquisa é feita por nome de autor.

## Xquery Tipo



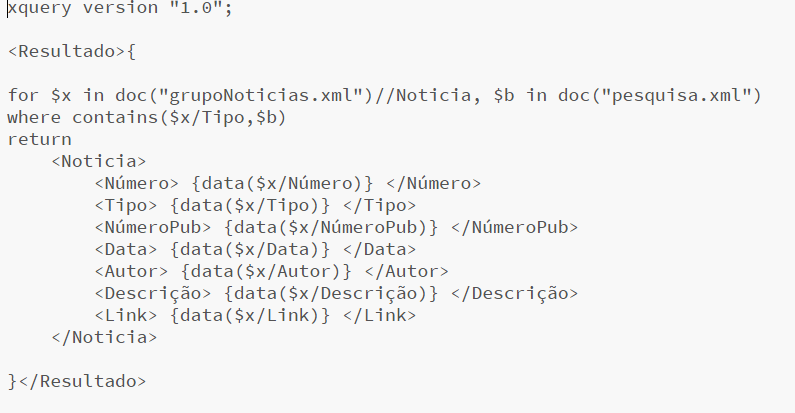
Esta query, procura no ficheiro XML das notícias se alguma delas tem a chave de pesquisa do XML em que esta guardada. Devolve as notícias que contenham a chave no seu conteúdo. Esta pesquisa é feita por tipo de noticia.

## Xquery Câmara



Esta query, procura no ficheiro XML dos municípios se algum deles tem a chave de pesquisa do XML em que esta guardada. Devolve os municípios que contenham a chave no seu conteúdo. Esta pesquisa é feita por nome de municipio.

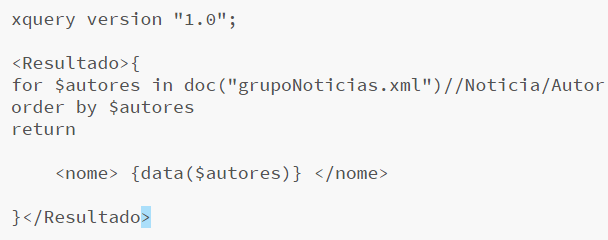
## Xquery Presidente



Esta query, procura no ficheiro XML dos municípios se algum deles tem a chave de pesquisa do XML em que esta guardada. Devolve os municípios que contenham a chave no seu conteúdo. Esta pesquisa é feita por nome do presidente da câmara.

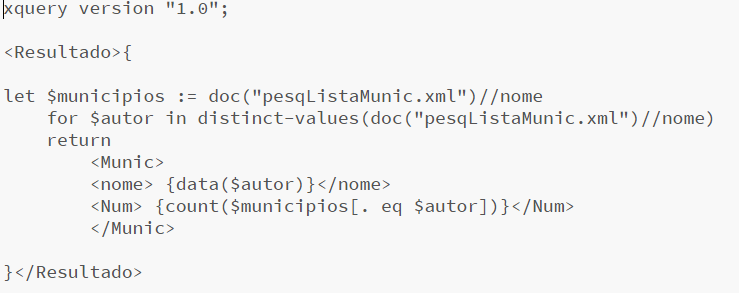
## Municipios

## Xquery Autores das noticias



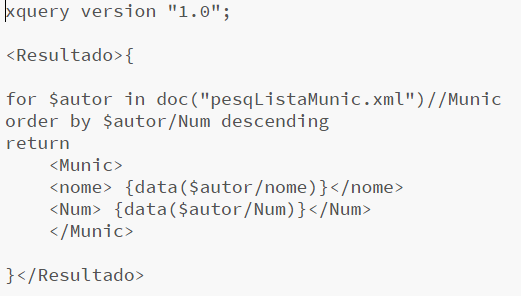
Esta query, procura no ficheiro XML das notícias, por todos os autores das mesmas, mesmo que sejam repetidos. Devolve a lista dos autores das notícias.

## Xquery Autores das noticias count



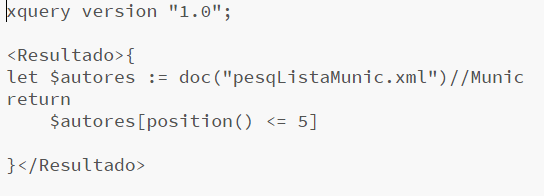
Esta query, vai ao XML dos nomes dos autores e vai fazer o count de cada um para que se retire os duplicados. Devolve a lista dos autores das notícias, sem duplicados e com a contagem.

## Xquery Autores das noticias count Desc



Esta query, vai ao XML dos nomes dos autores com a contagem. Devolve a lista dos autores das notícias, sem duplicados e por ordem decrescente de número de notícias publicadas.

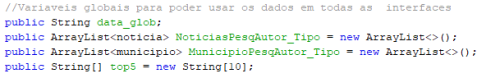
## Xquery top 5 municípios com mais publicações



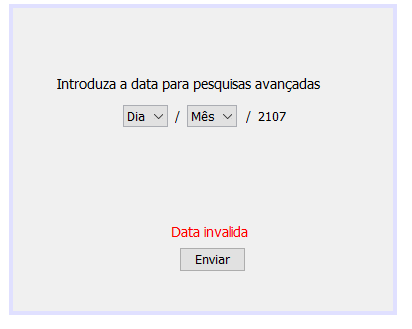
Esta query, vai ao XML dos nomes dos autores com a contagem e por ordem decrescente de número de publicações. Devolve apenas os 5 primeiros elementos da lista.

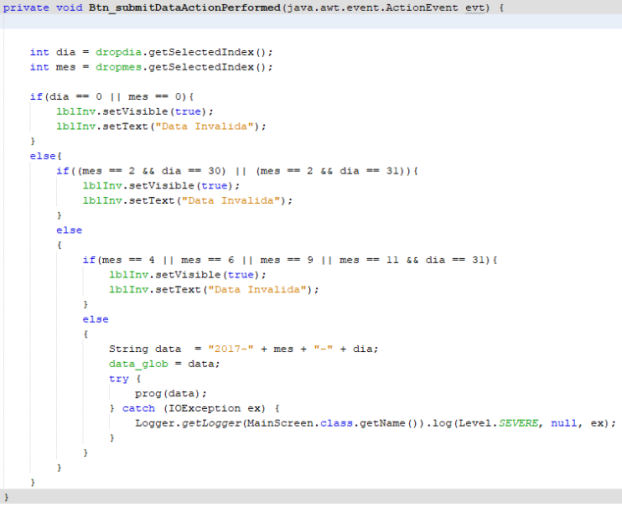
**2.6 – Interface Gráfica**

1. Variáveis Globais



## Painel inicial (Data)



Neste painel temos 2 dropboxes, uma para os dias e outra para os meses. 

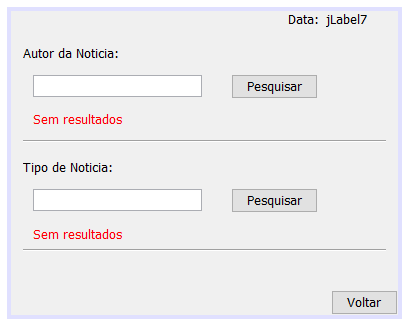
Ao Clicar no botão iniciar, verificamos se a data é valida. Caso esta não o seja, mostramos uma mensagem a dizer que a mensagem é inválida. Caso seja valida chamamos a função prog que recebe a data como parâmetro e irá criar o XML das notícias e dos municípios.

## Menu de pesquisas

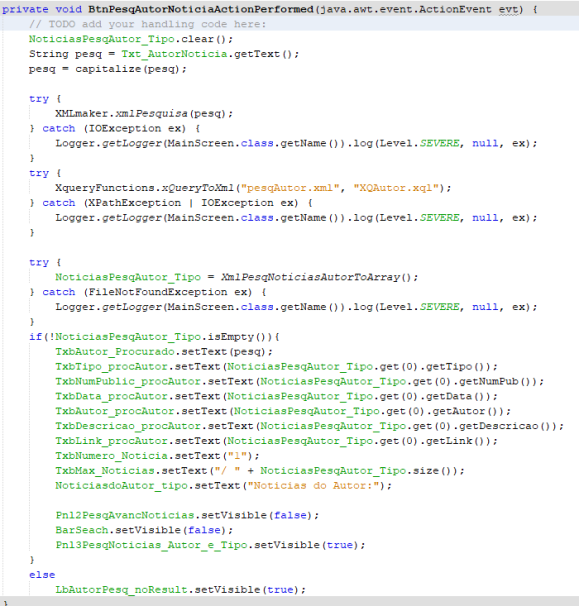
## C:\Users\Utilizador\AppData\Local\Microsoft\Windows\INetCache\Content.Word\menu.png

## Este menu serve para alterar parâmetros da pesquisa.

## Painel para pesquisa de noticias



Este painel permite-nos fazer a pesquisa de noticias por autor ou tipo. No canto superior mostra-nos a data da pesquisa em questão. Dá-nos também a possibilidade de voltar atrás para mudarmos a data de pesquisa.

Código do botão “pesquisar” por autor.

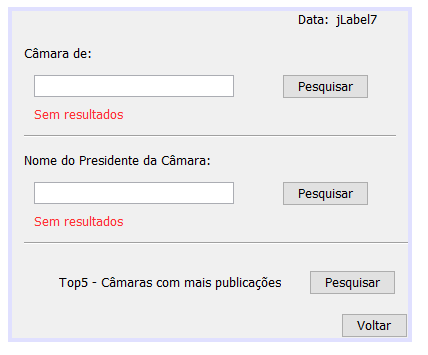
Guarda numa string o texto que for introduzido na caixa para a pesquisa e cria um XML com o mesmo.

Chama a função XquerytoXML que criará um ficheiro com o resultado final da pesquisa.

Passa o resultado da pesquisa do XML para um array e no caso do array não conter dados, este mostra uma mensagem “sem resultados”. Caso contrario, irá mostrar a informação no ecrã.

Na pesquisa por tipo, o procedimento irá ser o mesmo.

1. Painel para pesquisa de municípios



Este painel permite-nos fazer a pesquisa de municípios por nome ou presidente. No canto superior mostra-nos a data da pesquisa em questão. Dá-nos também a possibilidade de voltar atrás para mudarmos a data de pesquisa.

Código do botão “pesquisar” por nome do municipio.



Guarda numa string o texto que for introduzido na caixa para a pesquisa e cria um XML com o mesmo.

Chama a função XquerytoXML que criará um ficheiro com o resultado final da pesquisa.

Passa o resultado da pesquisa do XML para um array e no caso do array não conter dados, este mostra uma mensagem “sem resultados”. Caso contrário, irá mostrar a informação no ecrã.

Para que se consiga mostrar a imagem, é necessário criar um objeto url, ao qual damos o link da imagem, depois, criamos um objeto imagem ao qual adicionamos este url e para que a imagem seja mostrada, apenas temos que associar à label, um objeto *imageicon* com a imagem que criamos.

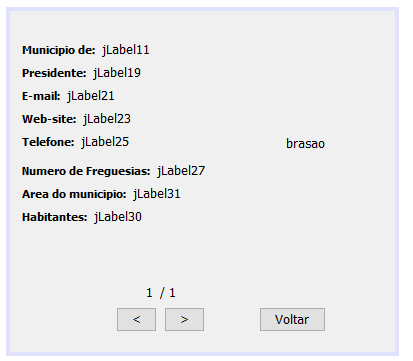
Na pesquisa por presidente da câmara, o procedimento irá ser o mesmo.

## Painel mostrar notícias



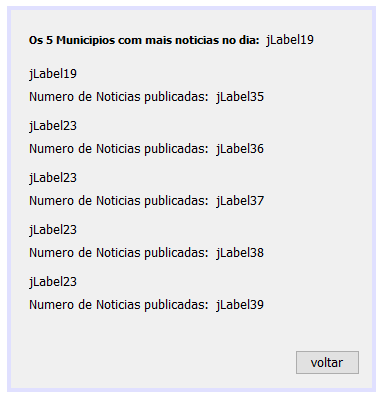
Este painel mostra as noticias resultantes das pesquisas realizadas, dando-nos a hipótese de navegar entre elas com os botões “anterior” e “seguinte”.

## 7-Painel mostra Municípios



Este painel mostra os municípios resultantes das pesquisas realizadas, dando-nos a hipótese de navegar entre elas com os botões “<” e “>”.

## 8- Painel mostra municípios com mais noticias na data pesquisada

****

Este painel mostra-nos os 5 municípios com mais publicações e o número de das mesmas por cada município, dando apenas a opção de voltar ao painel das pesquisas.

Código do botão “pesquisar” por top 5 mais publicações.



Cria um XML com os 5 municípios com mais publicações e o número de publicações correspondentes.

Passa o resultado da pesquisa do XML para um array e mostra os dados no ecrã.