Test Document
XML - FEEDS

october 2016

TIMES OF MALTA *Credibility matters*

ÍNDICE

1 VERSION CONTROL 4

2 DOCUMENT PURPOSE 4

3 SETUP 4

4 FEATURES 6

4.1 FEEDS 6

4.2 LIST OF VACANCIES OF A CHANNEL 6

4.2.1 Cache 9

4.3 CATEGORIES OF ALL FEEDS 10

4.3.1 Vacancies by category 10

5 FUTURE IMPROVEMENTS 12

6 QUESTIONS OF THE TEST 13

6.1 JOB ITEMS ACROSS FEEDS ARE SLIGHTLY DIFFERENT. HOW WOULD YOU NORMALIZE ITEMS? 13

6.2 HOW WOULD YOU CACHE FEED? 13

6.3 HOW WOULD YOU CACHE RESULTS? PSR-6 13

6.4 WHAT TO DO WITH FUNCTION LOADFEED()? 13

6.5 WHY ELEMENT CHANNEL WAS ADDED IN GET.PHP? (LINE 50) 13

7 ISSUES 14

1 Version control

Autor	Versión	Fecha	Descripción
Rodrigo Isidro	01r00	22/10/2016	First Version

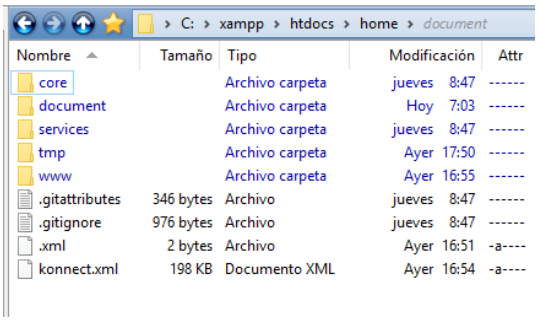
2 Document purpose

This document has the purpose of documenting the changes of the Xml platform of Times Malta. It will be explained the setup process of the application and the changes developed in the test.

Furthermore it will be commented the future improvements that must be done in the platform.

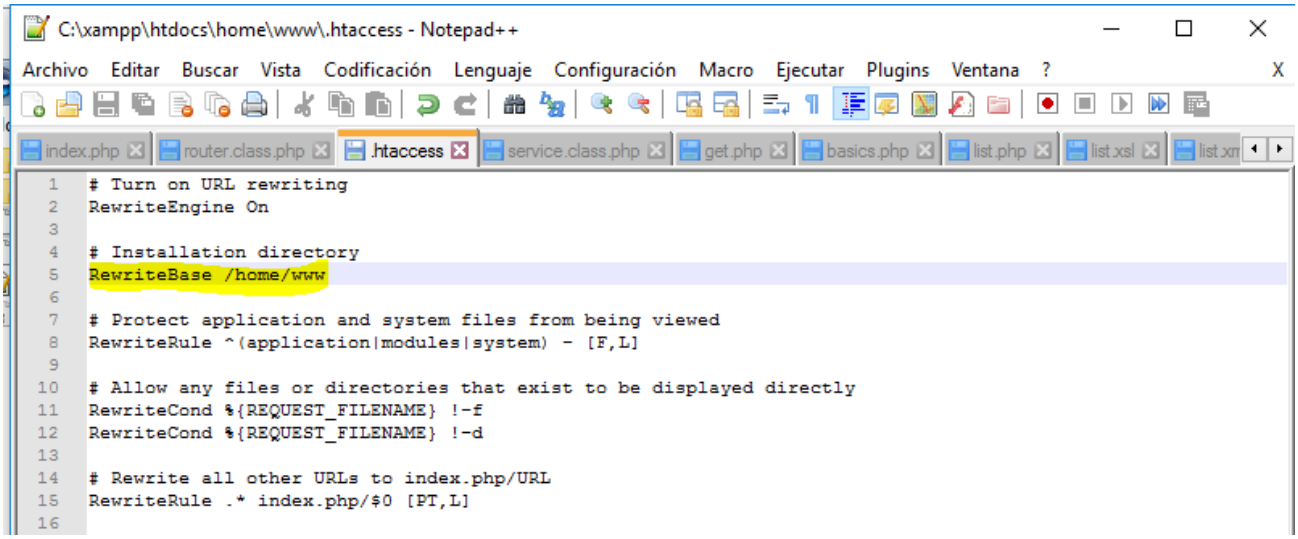
3 Setup

In this section it will be explained the setup process of the platform. The scructure of the original platform has the following list of folders:

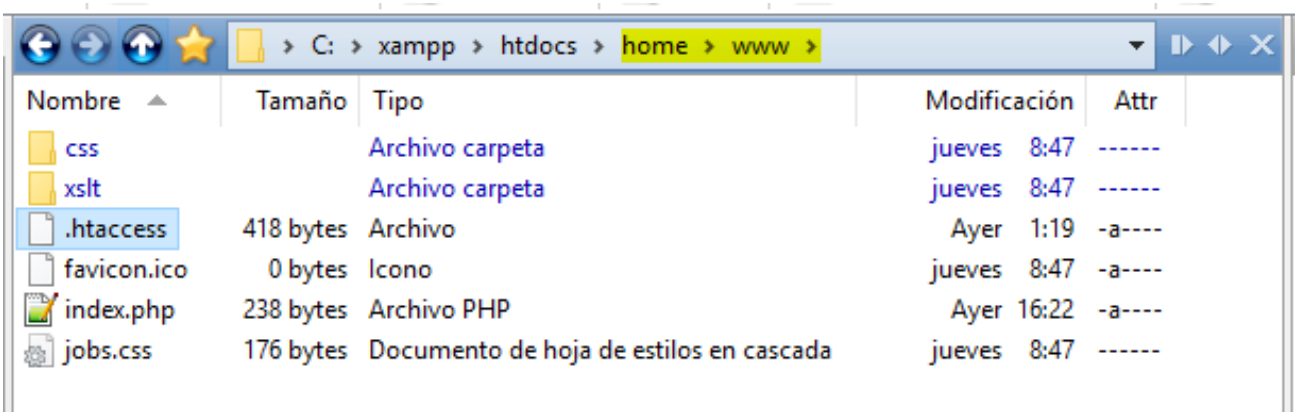


The web interface is contained in the "www" folder. This level must be copied below the root in a Production environment. You have to change de .htaccess for the URL variables will be availables and working on.

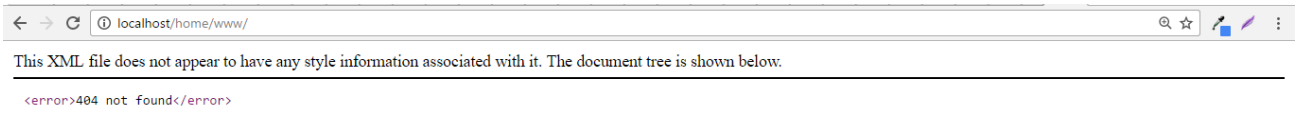
If the root directory of the server is not the setup folder you have to specify the path in the .htaccess file:



In our Windows Xampp environment:



In this moment, we have to test the path URL entering it to our web browser and test the URL localhost/home/www. (Replace localhost with the production server url). This application returns a Xml view, because is a XML feed of another Work Applications websites.



4 Features

4.1 Feeds

This app has configured 3 XML feeds:

- Times of Malta
- Castille Resources
- Konnect

These feeds are configured in an array in the following files: get.php and list.php

```
8   //Array of feeds
9   $feeds = [
10          "jobsinmalta" => "https://jobsinmalta.com/jobs.rss?exclude_recruitment_agencies=1&limit=5",
11          "konnect" => "https://www.konnekt.com/opportunities/feed",
12          "castille" => "https://www.castilleresources.com/en/rss"
13  ];
```

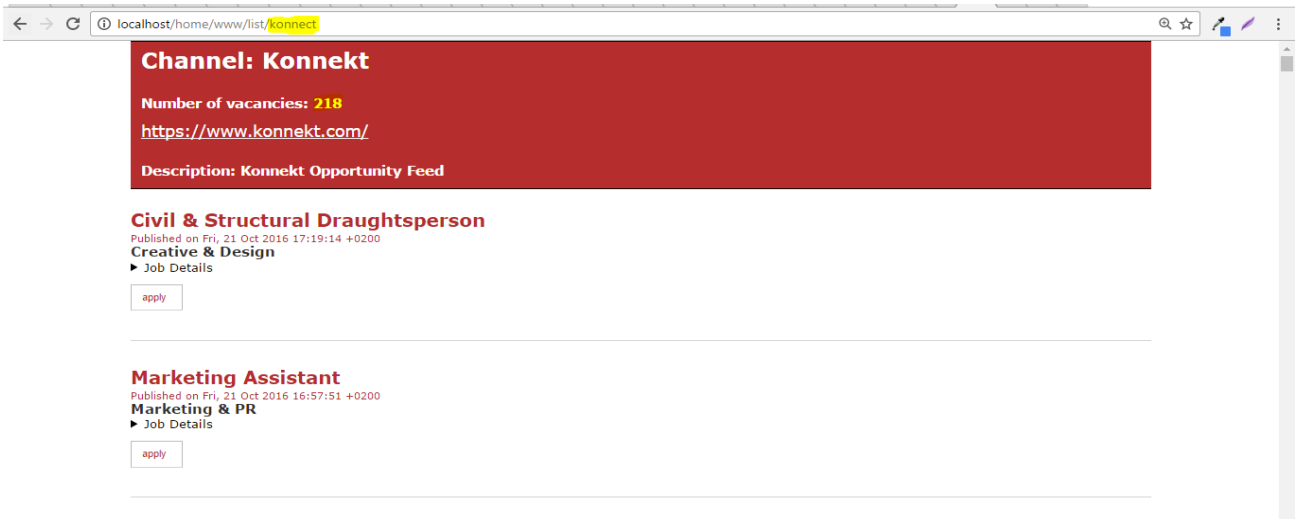
If we want to remove or add feeds we have to modify this array called $feeds.

Notef: These feeds are external of our aplication and we don't have control over them.

4.2 List of vacancies of a channel

This feature permits us to view an xml of the vacancies of the choosen channel passed by URL. You have to write the following URL in the website:

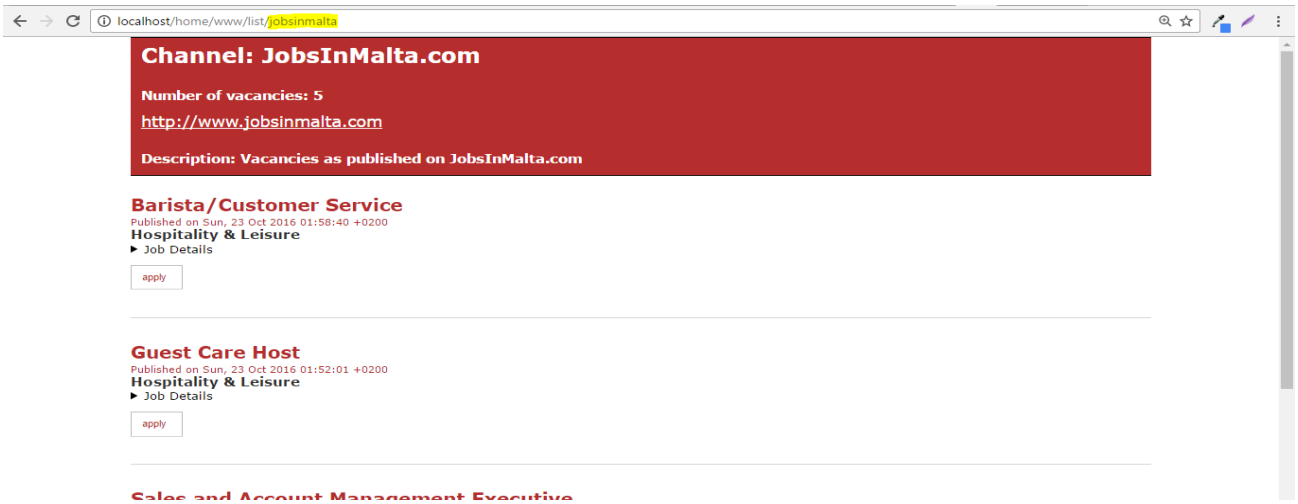
$_SERVER/www/list/$channel



In the top of the page we will view:

- The name of the Channel
- Number of vacancies of the channel
- Link of the channel
- Description of the channel

And then, they are a list of vacancies of the choosen channel. If we change the URL variable, the feed will change:



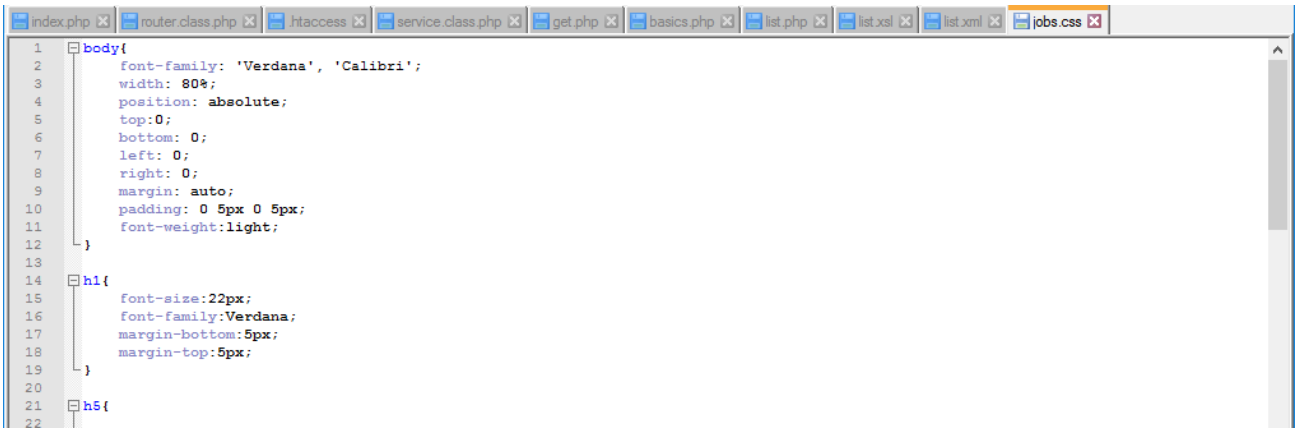
The list of vacancies is ordered by an XSL template placed in the www/xsl folder. The file is list.xsl

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0"
                xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <xsl:output method="html"  encoding="UTF-8"/>
    <xsl:template match="/">
        <html>
            <head>
                <title><xsl:value-of select="root/channel/title" /></title>
                <link href="./../css/jobs.css" rel="stylesheet" type="text/css" />
            </head>
            <body>
                <section>
                    <section>
                        <h1>Channel:  <xsl:apply-templates select="root/channel/title"/></h1>
                        <h5>Number of vacancies: <xsl:apply-templates select="root/nodes"/></h5>
                            <a>
                                <xsl:attribute name="href"><xsl:value-of select="root/channel/link"/></xsl:attribute>
                                <xsl:value-of select="root/channel/link" />
                            </a>
                        <h5>Description: <xsl:apply-templates select="root/channel/description"/></h5>
                    </section>
                    <xsl:apply-templates select="root/channel/item"/>
                </section>
            </body>
        </html>
    </xsl:template>

    <xsl:template match="pubDate">
        <time><xsl:value-of select="." /></time>
    </xsl:template>

    <xsl:template match="item">
```

The CSS design is got by the file /www/css/Jobs.css If we want to change the design of the XML feeds we have to change this file:

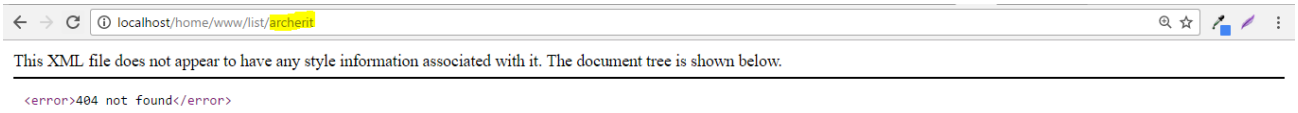


```
body{
    font-family: 'Verdana', 'Calibri';
    width: 80%;
    position: absolute;
    top:0;
    bottom: 0;
    left: 0;
    right: 0;
    margin: auto;
    padding: 0 5px 0 5px;
    font-weight:light;
}

h1{
    font-size:22px;
    font-family:Verdana;
    margin-bottom:5px;
    margin-top:5px;
}

h5{
```

If we enter in a url without associated channel, the website will return a 404 error:



4.2.1 Cache

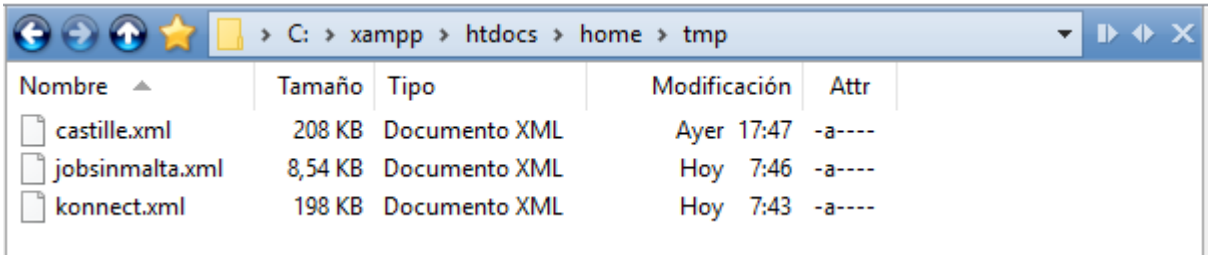
This funcionality has the feature of cache. Cache permits to obtain a local versión of the feeds placed in our server. With this feature we get a better performance of the website. Cache must be turned on or turned off. In the top of the file list.php we can find the following lines:

```php
<?php
//We decide if we want to use cache or not.
$cacheMode = true;
$timeCache = 120;
$getChannel = htmlspecialchars(Router::$params[2]);
```

$cacheMode: True / False to turn on or to turn off the cache system.

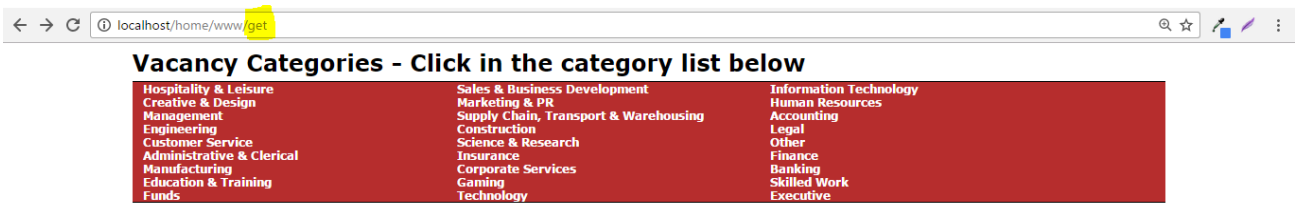
$timecache: Time in seconds that the feed is saved in our local system. We can specify the time changing this value.

The system will check if there is a local versión of the XML feed in the time we have decided to cache. If not, or neighter the cache time is over, the system will go to the original website to get it. Cache files are stored in the /tmp/ folder:



4.3 Categories of all feeds

This feature permits us to view all the categories of vacancies of every channel. If you go to /get/ url you will receive a list of all categories of all feed's channels:



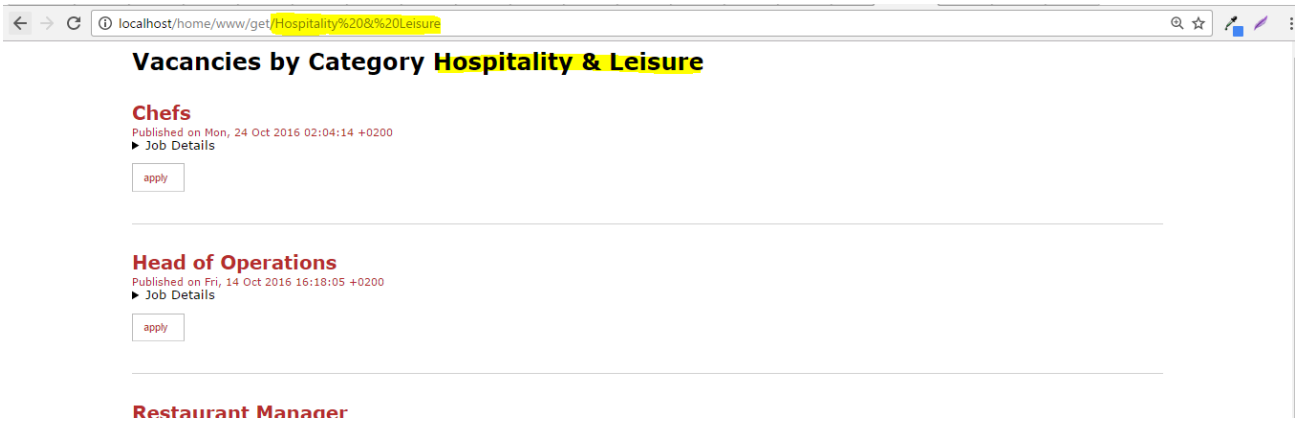
This page uses another XSLT template called list2.xsl with its own styles.

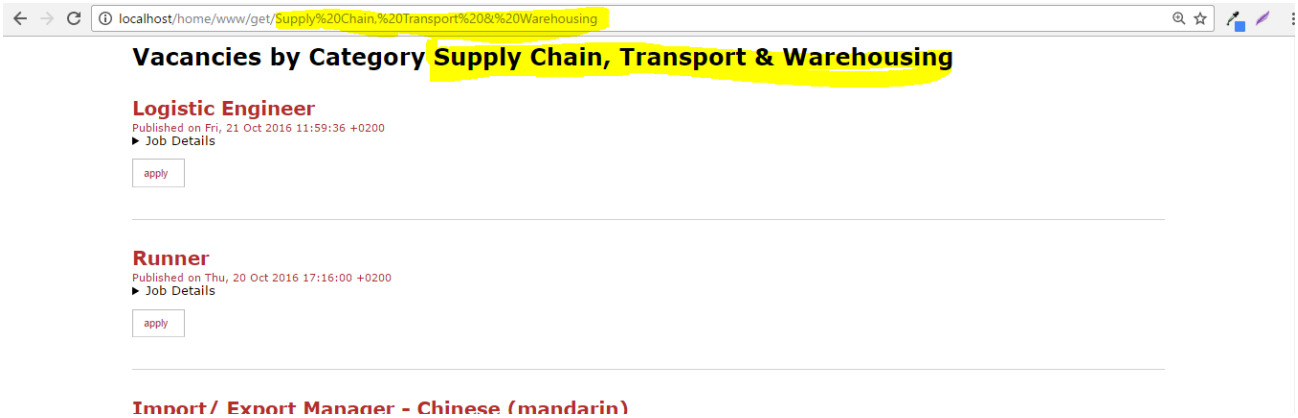
If you click in the name of any category, the page will link you to a list of vacancies of that category.

NOTE: This page is not cached. It get every categories each it's called by URL.

4.3.1 Vacancies by category

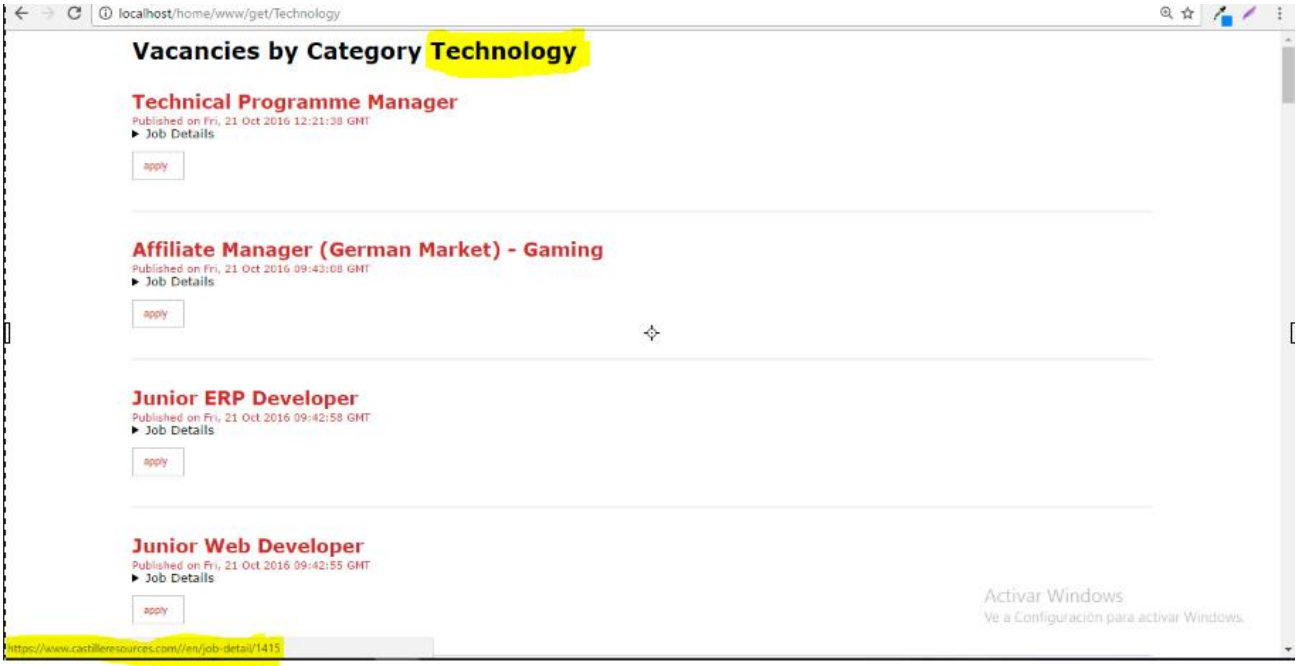
If we click over any element of the list /get/ we will go to a page with the results of only vacancies of that category.


localhost/home/www/get/Hospitality%20&%20Leisure
Vacancies by Category Hospitality & Leisure
Chefs
Published on Mon, 24 Oct 2016 02:04:14 +0200
Job Details
apply
Head of Operations
Published on Fri, 14 Oct 2016 16:18:05 +0200
Job Details
apply
Restaurant Manager


localhost/home/www/get/Supply%20Chain,%20Transport%20&%20Warehousing
Vacancies by Category Supply Chain, Transport & Warehousing
Logistic Engineer
Published on Fri, 21 Oct 2016 11:59:36 +0200
Job Details
apply
Runner
Published on Thu, 20 Oct 2016 17:16:00 +0200
Job Details
apply
Import/ Export Manager - Chinese (mandarin)

Vacancies by Category Sales & Business Development
Telemarketer
Published on Mon, 24 Oct 2016 02:55:44 +0200
Job Details
apply
Sales Specialist
Published on Mon, 24 Oct 2016 01:42:24 +0200
Job Details
apply
Sales Manager
Published on Thu, 20 Oct 2016 07:54:40 +0200
Job Details
apply

These lists are obtained of the three feeds but how the organization of categories are different, there is no point in common in some categories. For Example, IT (Konnect) Technology (Castille-Resources).



5 Future Improvements

In this section, we explain future improvements that will be desired to implement in the application.

- Centralice the feeds. We must put on feeds in one place with the idea to a better organization. We could do it in a database or in a file.
- Cache category feed in get.php
- Category order alphabeticaly in get.php
- Why not Rest Apis with json communication? I think the communication is quicker than with XML and XSLT.

6 Questions of the test

6.1 Job items across feeds are slightly different. How would you normalize items?

You have to see the difference in the nodes of the 3 feeds and there are two options to do:

- Quit nodes for normalize the feeds with a function removeChild('node');
- Add Nodes with null value or known data. For example images of the feed. We can use the method of SimpleXml ->addNode($node) in a foreach.

6.2 How would you cache feed?

Cache is used in the list.php (Please refer to 4.2.1)

6.3 How would you cache results? PSR-6

To be honest, I don't know PSR-6 and I don't have experience about this to answer this question.

6.4 What to do with function loadFeed()?

I would develope a class with this function beacuse is used along the code.

6.5 Why element channel was added in get.php? (line 50)

It's used for searching with the feature xpath of SimpleXml.

7 Issues and other consderations

In this section I expose the problems of the platform if there were:

- Sometimes /get/ appears in blank. Perhaps for timeout or there is a problem with the original Router::.
- In Unix system is probably that the app failes because of path directory separator.
- I add the sentence CURL_OPT_VERIFIER = FALSE to communicate with the feeds from localhost. If you consider delete it, do it.