OSS AUDITOR

**License Aggregator Project**

Rogue wave software | 1315 W Century Dr #150, Louisville, CO 80027

**this porject allows OSS auditors to research license in different sources in one internal webpage**

Chavirat Burapadecha

2016

Contents

[Overview 2](#_Toc457139871)

[Prepare data 2](#_Toc457139872)

[Web scraping technique 2](#_Toc457139873)

[How to extract information from the website? 2](#_Toc457139874)

[Match license keys 3](#_Toc457139875)

[Install LAMP stack in server 3](#_Toc457139876)

[Create the directory for stored web pages 3](#_Toc457139877)

[Test PHP service on the server 3](#_Toc457139878)

[Create a database and tables in MySQL 4](#_Toc457139879)

[Import data from .CSV files into tables 5](#_Toc457139880)

[Check the number of row for each imported data 6](#_Toc457139881)

[Implement the license aggregator website. 6](#_Toc457139882)

[Sitemap. 6](#_Toc457139883)

[Implement back-end pages 9](#_Toc457139884)

[Connect PHP to MySQL 9](#_Toc457139885)

[Query code 9](#_Toc457139886)

[SELECT 9](#_Toc457139887)

[INSERT 10](#_Toc457139888)

[UPDATE 10](#_Toc457139889)

[DELETE 10](#_Toc457139890)

[Develop front-end pages 10](#_Toc457139891)

[Install Bootstrap CDN and customize it 10](#_Toc457139892)

# Overview

The license aggregator project aggregates open source software(OSS) license information from multiple sources and mashes it up with Olex license information into a single database that help OSS auditors view through a web page. In the webpage, the auditor can query, insert, update and delete information on standard licenses and obligation. Furthermore, the auditor can compare license information from different sources. This project uses many methods to collect the information from these sources and demonstrate into the web-based application. These methods includes web scraping, mashing information, installation LAMP stack, creating a database, and implementation of a web page.

# Prepare data

## Web scraping technique

Web scraping (web harvesting or web data extraction) is a computer software technique of extracting information from websites. I used the web scraping with Python (BeautifulSoup version 4) and PHP (HTML DOM) to extract information from such websites as Nexb, SPDX, and Tldrlegal websites.

* Nexb website <https://enterprise.dejacode.com/license_library/>
* SPDX website <http://spdx.org/spdx-license-list>
* Tldrlegal website <https://tldrlegal.com/>

### How to extract information from the website?

After inspecting the source code in the website, I use BeautifulSoup package which is a Python library to pull data out of HTML page to .CSV file. This technique can be used for Nexb and SPDX websites, but not for Tldrlegal website. Therefore, I use PHP (HTML DOM) to pull only the URL link and need to manually fetch data from this website.

The scripts are located in 10.75.1.12 server in   
/home/oluser/Documents/import\_data/web scraping scripts

The results of data are located in 10.75.1.12 server in   
/home/oluser/Documents/import\_data

The olex license information is located in [Audit Report Resources](https://rwsoftware-my.sharepoint.com/personal/dave_mcloughlin_roguewave_com/_layouts/15/onedrive.aspx#FolderCTID=0x0120004F8AAA037273A9489FBE611278BF4F0B&AjaxDelta=1&isStartPlt1=1469120535956&id=%2Fpersonal%2Fdave%5Fmcloughlin%5Froguewave%5Fcom%2FDocuments%2FNotebooks%2FOpenLogic%20Audit%20Folder%2FAudit%20Report%20Resources).

The table below shows the relationship between the scripts that use to fetch data into .csv files. For instead, the Spdx-license.py script uses to pull the license list from SPDX website and stores the list into SDPX-license-info.csv file. And then the SPDX-license-info.py script will select each license of the list to get license information and repeat into SPDX-license.info.csv file.

*The process of fetching SPDX license information from SPDX.org into the .csv file.*

|  |  |
| --- | --- |
| Script | Data |
| Spdx-license.py (for license list)  SPDX-license-info.py (for license information) | SPDX-license-info.csv |
| Match license keys in excel with fussy lookup  spdx-license-information.py (for getting spdx key in nexb web site) | match\_license.csv |
| Nexb-license.py (for license list)  Nexb-license-information.py (for license information) | nexb-license-information.csv |
| nexb-license-requirements.py (for license requirements for each license) | nexb-license-requirements.csv |
| nexb\_requirements.py (for license requirements) | nexb\_requirements.csv |
| From bash script and shared excel documents on One Drive | olex\_license.csv |
| olex\_license\_obligation.csv |
| olex\_obligation.csv |
| PHP Web Scraping Script | tldr\_license.csv |
| Manual matching | tldr\_requirements1.csv |

## 

## Match license keys

After I created the license files for each website, I matched the key of each license in one table in order to compare the information from different resources by using fussy lookup in Microsoft excel.

# Install LAMP stack in server

LAMP stack is a popular open source web platform commonly used to run dynamic web sites and servers. It includes Linux, Apache, MySQL, and PHP/Python/Perl and is considered by many the platform of choice for development and deployment of high performance web applications which require a solid and reliable foundation.

Before installation of the LAMP stack, you must update the application which is running on the server.

This command is to update the applications.

$ sudo apt-get update

This command is to install lamp stack in server.

$ sudo apt-get install lamp-server^

## Create the directory for stored web pages

This command is to make directory in this path /var/www/html/license

$ sudo mkdir /var/www/html/license/

This command is to change permission in order to allow the developer to insert, update, or delete web pages under this folder.

$ sudo chmod 777 /var/www/html/license/

## Test PHP service on the server

Use the command to create and edit the test.php.

$ sudo touch /var/www/html/license/test.php

$ sudo nano /var/www/html/license/test.php

Then, open the test.php and insert the code below.

<? Php Print\_r (phpinfo());?>

After that, open the web browser to type 10.75.1.12/license/test.php. The result will be shown if the PHP service already run in the server. Otherwise, the web browser will show an error.

## Create a database and tables in MySQL

MySQL is an open-source relational database management system (RDBMS).

Use the command to access to MySQL in Linux.

Mysql –u root –p

The password is oluser12.

Use the command to show the database.

SHOW DATABASES;

Use the command to create the database name as ‘license’.

CREATE DATABASE license;

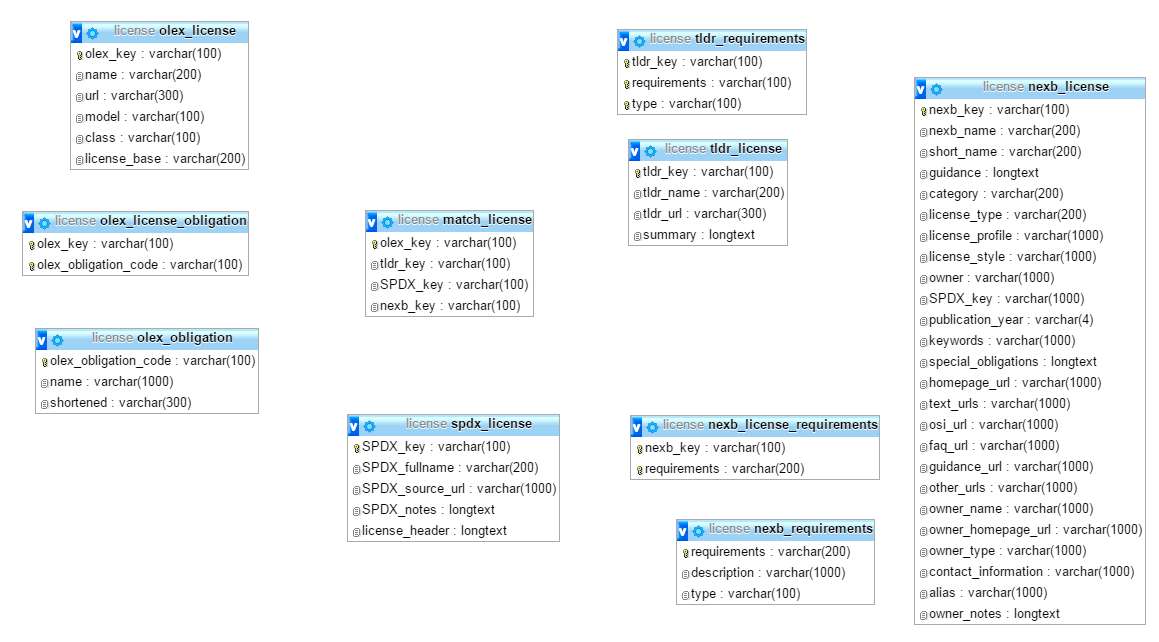
Use the command to select the license database.

USE license;

Create the 10 tables in license database as

1. spdx\_license
2. match\_license
3. nexb\_license
4. nexb\_license\_requirements
5. nexb\_requirements
6. olex\_license
7. olex\_license\_obligation
8. olex\_obligation
9. tldr\_license
10. tldr\_requirements





Here is the relationship between tables in the license database.

## Import data from .CSV files into tables

The table below shows the relationship between an imported data file and its table in the database. For example, the SPDX-license-info.csv was imported into spdx\_license table. The SQL code for importing data is provided in the import data sql document.

|  |  |
| --- | --- |
| Table | Data |
| spdx\_license | SPDX-license-info.csv |
| match\_license | match\_license.csv |
| nexb\_license | nexb-license-information.csv |
| nexb\_license\_requirements | nexb-license-requirements.csv |
| nexb\_requirements | nexb\_requirements.csv |
| olex\_license | olex\_license.csv |
| olex\_license\_obligation | olex\_license\_obligation.csv |
| olex\_obligation | olex\_obligation.csv |
| tldr\_license | tldr\_license.csv |
| tldr\_requirements | tldr\_requirements1.csv |



## Check the number of row for each imported data

This SQL queries shows the number of row for each table to ensure that all data are imported correctly.

Implement the license aggregator website.

## Recommended tools for implement

I strongly recommend to use Atom which is a text editor that's modern, approachable, yet hackable to the core—a tool you can customize to do anything but also use productively without ever touching a configure file to implement the code. (<https://atom.io/>)

I also installed WinSCP to transfer the pages into the internal web server (10.75.1.12). WinSCP is an open source free SFTP client, FTP client, WebDAV client and SCP client for Windows. Its main function is file transfer between a local and a remote computer. Beyond this, WinSCP offers scripting and basic file manager functionality. <https://winscp.net/eng/download.php>

I developed this site in my local server before I transfer it into the internal server (10.75.1.12). Therefore, VPN connection may be required if the transferring is not from the Rogue Wave environment.

## Sitemap.

The web pages of this project is stored in license folder (/var/www/html/license/). Here is the sitemap of the web-based application and the table below explains the role of each page.

|  |  |
| --- | --- |
| Webpage | Description |
| Index.php | Is a homepage or a landing page |
| License\_list.php | Pull information of olex license table and allow the user to sort by alphabet for each column |
| Search.php | Allow the user to search for olex\_license |
| Admin.php | Allow the administrator to add, update, insert licenses and obligations |
| License\_info.php | Compare the information from such 4 webpages as olex\_license.php, nexb\_license.php, spdx\_license.php, and tldr\_license.php. |
| Olex\_license.php | Show olex license information |
| Nexb\_license.php | Show nexb license information |
| Spdx\_license.php | Show spdx license information |
| Tldr\_license.php | Show tldr license information |
| Sql\_connect.php | Store the variables using for MySQL connection |
| Form.php | Show insert forms for licenses and obligations |
| Insert\_olex.php | Insert the license information which received from form.php into an olex\_license table |
| Insert\_nexb.php | Insert the license information which received from form.php into a nexb\_license table |
| Insert\_spdx.php | Insert the license information which received from form.php into a spdx\_license table |
| Insert\_tldr.php | Insert the license information which received from form.php into a tldr\_license table |
| Insert\_nexb\_ob.php | Insert the obligation information which received from form.php into a nexb\_requirements table |
| Insert\_olex\_ob.php | Insert the obligation information which received from form.php into an olex\_obligation table |
| License\_list\_edit.php | Show the list of licenses and obligations from different tables |
| Edit\_match.php | Show the information of a selected key and allow the administrator to change the information before update or delete it in a match\_license table. |
| Edit\_nexb\_ob.php | Show the information of a selected obligation and allow the administrator to change the information before update or delete it in nexb\_requirements table. |
| Edit\_nexb.php | Show the information of a selected license and allow the administrator to change the information before update or delete the license in nexb\_license table. |
| Edit\_olex.ob.php | Show the information of a selected obligation and allow the administrator to change the information before update or delete it in olex\_obligation table. |
| Edit\_olex.php | Show the information of a selected license and allow the administrator to change the information before update or delete the license in olex\_license table. |
| Edit\_spdx.php | Show the information of a selected license and allow the administrator to change the information before update or delete the license in spdx\_license table. |
| Edit\_tldr.php | Show the information of a selected license and allow the administrator to change the information before update or delete the license in tldr\_license table. |
| Update\_match.php | Update information of the selected key in match\_license table. |
| Update\_nexb\_ob.php | Update information of the selected nexb obligation in nexb\_requirements table. |
| Update\_nexb.php | Update information of the selected nexb license in nexb\_license table. |
| Update\_olex.php | Update information of the selected olex license in olex\_license table. |
| Update\_spdx.php | Update information of the selected spdx license in spdx\_license table. |
| Update\_tldr.php | Update information of the selected tldr license in tldr\_license table. |

# Implement back-end pages

In software engineering, front end and back end distinguish between the separation of concerns between the presentation layer (the front end) - which is the interface between the users - and the data access layer (the back end). For back-end pages, I started to connect PHP to MySQL and pulled the data into different pages such as insert, update, and delete licenses and obligation.

### Connect PHP to MySQL

The variables of connection from PHP to MySQL are in **sql\_connect.php**.

### Query code

### SELECT

The SQL SELECT statement returns a result set of records from one or more tables in order to present information of the license list, search function, and sortable table.

SYNTAX: SELECT *column* FROM *table* WHERE *condition* ORDER BY *column;*

**Here is the sample code of the SQL SELECT statement in license\_list.php.

### INSERT

The INSERT INTO statement is used to insert new records in a table.

SYNTAX: INSERT INTO table\_name (column1, column2, column3 ...)   
VALUES (value1, value2, value3...);

 Here is the sample code of the SQL INSERT statement in insert\_olex.php.

### UPDATE

The UPDATE statement is used to update records in a table.

SYNTAX: UPDATE table\_name

SET column1=value1, column2=value2,...

WHERE some\_column=some\_value;

 Here is the sample code of the SQL UPDATE statement in update\_olex.php.

### DELETE

The DELETE statement is used to delete rows in a table.

SYNTAX: DELETE FROM table\_name  
WHERE some\_column=some\_value;

 Here is the sample code of the SQL DELETE statement in delete\_olex.php.

# Develop front-end pages

## Install Bootstrap CDN and customize it

Bootstrap is a free and open-source front-end web framework for designing websites and web applications. You may download it from <http://getbootstrap.com/>.