Стек технологий и дополнительных библиотек:

1. PHP 7.2;
2. MySQL 5.7;
3. Библиотеки simple\_html\_dom и curl\_query;

Особенности проекта:

Так как фильтр <http://filmix.cc/persons> (далее сайт\*) работает через POST метод невозможно взять начальную точку парсинга со страниц отсортированных по алфавиту. Так же невозможно начать со страницы <http://filmix.cc/persons/page/1> чтобы в полной мери использовать шаблон «/page/$i» того начальная точка страница со звёздами.

Описание архитектуры, методов и пр:

Начальная точка входа index.php.

Имеется основных два класса (Parser, DB):

1. Class Parser класс работает только со страницами и их элементами имеет три метода:

array\_formation() – формирует массив ссылок. Так как на сайте стоит по умолчанию фильтр по рейтингу, то новые актёры могут смещаться в зависимости от изменения их рейтинга. Для того чтобы не упустить новых актёров формируем массив страниц.

start\_parse(array $urls) – принимает массив ссылок. С помощью библиотек curl\_query.php и simple\_html\_dom.php и их методов парсим чтобы получить ссылки на открытые страницы актёров. Так же с помощью метода check\_url() класса DB проверяется является ли актёр уже записанный в базе данных. Если нет, то добавляем его в массив для дальнейшего парсинга. Так же проверяется количество элементов в массиве для того чтобы записать сто новых актёров.

Parse(array $pages) – принимает массив ссылок на открытые страницы актёров. Где уже парсится по основным критериям имя на русском, оригинальное имя, ссылка на фото, дата и место рождения, биографию, ссылку на страницу. Из полученных данных формируется массив для вызова метода set\_info класса DB.

1. DB класс работает с БД, имеет три метода:

getInstance() – статический метод для подключение в БД использовался паттерн “singleton”.

check\_url($link\_page) – метод, цель которого проверить принимает ссылку на персональную страницу актёра. При запросе в БД экранируются данные.

set\_info(array $data) – метод принимает в себя распарсенные данные по актёру чтобы записать их в базу данных. Также все данных экранируются.

Дополнительные методы:

check\_date() – метод, цель его проверять запускалась ли функция уже в этот день. Вместе с тем для решения задачи «при условии запуска 1 раз в сутки (или другой период)» лучше подойдет cron ведь он лучше, чем проверка из БД. Того я бы установил бы на сервер вот примерно вот такой cron: \* \* \*/1 \* \* curl <http://site.com/>. P.s. метод закомментирован так как считаю, что в реалии лучше сделать через крон.

Архитектура БД состоит всего из двух таблиц одной основной и одной дополнительной.

Основная “info\_stars” имеет такие поля как:

| [**id**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[2]=1) | [**link\_page**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[3]=1) | [**name\_rus**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[4]=1) | [**name\_original**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[5]=1) | [**foto\_link**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[6]=1) | [**birthday**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[7]=1) | [**birth\_place**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[8]=1) | [**biography**](http://127.0.0.1/openserver/phpmyadmin/index.php?db=parser&table=info_stars&target=sql.php&token=ecb3d8023ba0208727ca6892cb08ae4c#usesubform[9]=1) |
| --- | --- | --- | --- | --- | --- | --- | --- |

Дополнительная “date\_parsers”:

| **id** | **date\_pars** |
| --- | --- |

Это если всё-таки нужно выполнять работу парсера раз в сутки силами РНР.