

Тестовое задание

Задание:

- 1) Скачать файлы для парсинга с

<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>

Будут использоваться 2 файла:

- IMDb names.csv
- IMDb movies.csv

- 2) Из файла "IMDb names.csv" нужны данные:

- imdb_name_id
- name
- height
- bio
- date_of_birth
- place_of_birth
- children

Данные из этого файла записывать в таблицу casts.

Дополнительно должны быть колонки:

- is_usa - если дата рождения USA
- is_europe - если дата рождения страна из Европы

- 3) Из файла "IMDb movies.csv" нужны данные:

- imdb_title_id
- title
- year

- genre
- duration
- country
- language
- director
- writer
- actors
- description
- avg_vote
- votes
- reviews_from_users
- reviews_from_critics

Дополнительно должны быть колонки:

- is_usa - если страна создания фильма USA
- is_europe - если страна создания фильма из Европы
- is_top - если рейтинг больше 8.0 (включительно)

Данные из этого файла записывать таблицу movies и дополнительно записывать в две таблицы old_movies и new_movies.

old_movies = [id, movie_id, year] - до 1980 года

new_movies = [id, movie_id, year] - после 1980 года включительно

В отдельные таблицы нужно записывать:

- genres = [id, name]
- countries = [id, name, is_usa, is_europe].
- language = [id, name]

Поэтому должны быть проставлены связи

- movies <-> genres
- movies <-> countries
- movies <-> casts (director, writer, actors - это все casts)

4) Сделать отдельную страничку, где будет отображаться дополнительная информация:

- показать топ5 лучших по рейтингу фильмов (avg_vote) за каждое десятилетие в США и Европе по отдельности с количеством голосов (votes) больше 500.
- в таблице должны быть поля: Years (например 1920s), Rating (avg_votes), Movie name, Year of movie, Count of actors, Genre

5) На главной странице (дашборде) должна быть таблица с статистикой по каждому парсингу

- дата и время (когда прошел парсинг),
- сколько данных в каждой файле парсилось, сколько данных записалось в каждой таблице
- сколько времени происходил парсинг (в human readable формате - минуты, секунды)

Условия:

- Нужно использовать паттерн репозиторий. Например - <https://github.com/andersao/l5-repository>
- Используем последнюю версию Laravel и PHP 7
- Парсинг должен работать в очередях
- Повторный вызов парсинга не должен создавать дубликатов в базе

Доп. информация:

- список страниц из Европы можно взять, например, тут <https://github.com/aiturner/acetate/blob/master/places/Countries-Europe.csv>

