

Турнир блиц-проектов Soft-Парад — 2019

Исходные данные

Файл `c:\sp\movies.csv` содержит сведения о фильмах в формате CSV:

идентификатор, название (год), жанр1|...|жанрN

Если в названии есть запятая, то название с годом заключаются в кавычки.

Файл `c:\sp\ratings.csv` содержит оценки фильмов пользователями также в формате CSV:

идентификатор пользователя, идентификатор фильма, рейтинг от 0.0 до 5.0 с точностью до 0.5,

Unix-время выставления оценки

Задания

- 1) Поиск фильмов.** Найти фильмы, содержащие все указанные пользователем слова в названии в виде подстроки (в любом порядке, независимо от регистра букв). Отсортировать результаты по идентификатору фильма и вывести название и год первых 3 фильмов. *(5 баллов)*
- 2) Поиск режиссёра.** Для указанного фильма найти в Интернет и вывести его режиссёра. Пользователем указывается идентификатор фильма из файла `movies.csv`. *(10 баллов)*
- 3) ТОП фильмов.** Для указанного жанра найти ТОП-3 фильмов на основе средней оценки пользователей. Учитывать только фильмы, у которых не менее 50 оценок. В случае равной оценки выводить фильмы по порядку идентификаторов. Выводить название и год. *(6 баллов)*
- 4) Рекомендация пользователя.** Найти пользователя, наиболее похожего на указанного пользователя, определённого на основе схожести их оценок на фильмы. Рассматривать только тех пользователей, которые просмотрели как минимум 5 одних и тех же фильмов. Формула схожести двух пользователей:
$$\frac{\text{количество совпадающих оценок}}{\text{количество фильмов (которые смотрели оба пользователя)}}$$
В случае совпадения схожести вывести пользователя с наименьшим идентификатором. *(10 баллов)*
- 5) Диаграмма активности пользователей.** Построить на одном изображении линейный график активности пользователей в указанном диапазоне годов (по локальному времени). Активность пользователей считать как количество оценок фильмам в год. Оси координат изображать необязательно. *(8 баллов)*

Ограничение на время запуска программы — 30 сек.

Ограничение на время выполнения каждого теста — 15 сек.