Написать Рекомендательную систему

Сервис создает рекомендации курсов в траектории для профессии, которую выбрал пользователь. Курсы рекомендуются как для навыков, которыми пользователь не владеет (чтобы обучиться новому), так и для навыков, которые пользователь уже указал как освоенные (чтобы улучшить старые).

Шаг 1

Набор навыков можно получить **явно** - пользователь самостоятельно, в процессе анкетирования, определяет для себя набор навыков, которыми он владеет.

Более интересно, когда набор навыков подбирается **неявно** - сама система должна выделить недостающие навыки для какой либо профессии и предложить курсы для их освоения.

Шаг 2

По окончании первого шага, пользователь будет иметь список навыков для его карьеры. Необходимо для каждого навыка вывести подходящий список курсов. Фильтрация будет производиться по названию курса (высокий приоритет) и описанию курса (низкий приоритет).

Также необходимо поработать с маской, по которой курс будет проходить валидацию. Необходимо, чтобы маска затрагивала исключительно сам навыки и НЕ входил в часть слова [%Git% (Digital)]. Для этого можно добавить к маске ограничения пробелов или иных знаков препинания.

В случае, если навыки сильно отличаются друг от друга, но имеют схожие курсы (SQL и СУБД) необходимо добавить ключи к таким навыкам, чтобы система рекомендовала к ним одинаковые курсы.

Шаг 3

Для ранжирования курсов в ленте рекомендаций гораздо лучше использовать маркер популярности, чем алфавитную сортировку. Для этого необходимо либо с помощью парсера вычленить оценку с сервиса, на котором размещен курс и добавить ее к курсу для ранжирования в системе. Если оценки одного и того же курса есть на нескольких площадках, то взять среднюю оценку или выбрать эталонную площадку. Парсинг можно производить раз в месяц.

Второй вариант реализации ранжирования путем популярности исключает парсинг и заключается в задаче регрессии по предсказанию популярности курса. В этом случае необходимо вычленить основные метрики курса (длина видео, количество текста, количество ссылок на внешние ресурсы, цитируемость) и по ним дать оценку популярности курса для дальнейшего ранжирования.