

УДК 004.942

## РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ КУРСОВ НА ОСНОВЕ ПРОФЕССИОНАЛЬНЫХ ИНТЕРЕСОВ И ДАННЫХ ПСИХОЛОГИЧЕСКИХ ТЕСТИРОВАНИЙ ПОЛЬЗОВАТЕЛЯ

**Смирнов И. З.** (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»), **Лаптев А. А.** (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»), **Лаушкина А. А.** (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

**Научный руководитель - доктор технических наук, доцент Басов О. О.** (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

**Аннотация:** Работа посвящена проблеме холодного старта при разработке рекомендательных систем. Объектом исследования в данной работе выступают эмоциональное состояние человека и его предпочтения в области профессиональной деятельности. Предлагается модель рекомендательной системы, которая способна, исходя из интересов, данных профиля и эмоционального состояния пользователя, советовать курсы и мастер классы, которые могли бы способствовать профессиональному и личностному развитию.

**Введение.** Область рекомендаций усложняется по мере расширения сферы ее применения, с каждым новым вызовом и нюансом использования. По мере увеличения объемов данных и усложнения задач появляются новые подходы. Рекомендательные системы – это программы, которые пытаются предсказать, какие объекты будут интересны пользователю, имея определенную информацию о его профиле.

Условно алгоритмы рекомендательных систем можно разделить на 3 крупных блока:

1. Рекомендательные системы, основанные на коллаборативной фильтрации;
2. Персонализированные системы;
3. Рекомендательные системы, основанные на контенте.

Одной из главных проблем при разработке рекомендательных систем является “холодный старт” - отсутствие или малый объем релевантных данных о пользовательских предпочтениях. В данной работе предлагается решение данной проблемы с помощью обхода “холодного старта” на основе психологических тестирований, а также на основе метрических данных по взаимодействию с рекомендованным контентом (время просмотра, лайк/больше не показывать контент).

**Основная часть.** В качестве исходных данных пользователю предлагается выбрать несколько профессиональных интересов и пройти психологическое тестирование (BIG5, цветовой тест Люшера). Выбор курсов производится:

- 1) по видам сфер
- 2) по их направлениям (уточняются теги).

Так же данные, полученные при прохождении тестирований и оценке эмоционального состояния, позволяют оценить предпочтительную сложность курсов и то, насколько рекомендации могут отклоняться от профессиональных интересов. Если, к примеру, человек утомлен, то курсы тоже могут отличаться от специализации с целью переключения на другие задачи, или для возможности отвлечься. Данная практика позволяет минимизировать “холодный старт” – когда в системе еще недостаточно данных о новом пользователе для адекватных рекомендаций. Изначальные интересы пользователя имеют самый большой весовой коэффициент среди всех входных данных разрабатываемой системы, т.е. они больше всего влияют на весовые коэффициенты тегов. Если пользователю понравился/не понравился

курс (лайк/сохранить курс/больше не показывать), то средний вес тегов, которые принадлежали данному курсу, увеличиваются и наоборот.

При выводе курсов в рекомендациях дается краткое описание и ключевые слова отзывов. Для повышения качества рекомендаций необходимо учитывать среднее время данной информации, а также среднюю скорость чтения пользователя и количество символов в заголовке. Для нормализации текста описания курса используется библиотека `ru morphology`, которая позволяет работать с русскоязычными текстами. Для получения ключевых слов, рассматриваемых рекомендательной системой, нормализованные текстовые данные переводятся в векторы при помощи Tf-Idf векторизатора. Таким образом каждый курс описывается длительностью, уровнем сложности, специализацией и набором ключевых слов.

На данный момент в базу курсов входят данные с образовательных платформ `skillbox.ru` и `stepik.org`: обе платформы имеют широкий выбор направлений.

Курсы имеют информацию о сложности и длительности прохождения. Эти параметры важно учитывать так же при рекомендации курсов, поскольку их длительность и сложность не должна отвлекать от основной деятельности.

**Выводы.** В одном из подходов для рекомендации курса стоит рассчитывать косинусное расстояние. При расчете возможна модификация результата вычисления расстояния в зависимости от сходства не текстовых признаков.

Рассмотренная альтернатива косинусному расстоянию при построении рекомендаций - кластеризация. В этом случае стоит рассматривать только текстовые данные. При подборе курсов следует учитывать их актуальность, что можно понять из отзывов других пользователей. При малой информации о пользователе так же возможна рекомендация курсов аналогов исходя из указанной информации и реакции на представленные рекомендации. Учет данных психологических тестирований также позволяет брать в расчет отзывы и не выводить негативно окрашенные курсы в превью в некоторых случаях.

Смирнов И.З. (автор)

Подпись

Лаптев А.А. (автор)

Подпись

Лаушкина А.А. (автор)

Подпись

Басов О.О. (научный руководитель)

Подпись