

ОБРАБОТКА ЗАПРОСА НА ЕСТЕСТВЕННОМ ЯЗЫКЕ ДЛЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ МООС НА ОСНОВЕ КОНТЕНТНОЙ ФИЛЬТРАЦИИ

Аннотация. В статье описан алгоритм для реализации рекомендательной системы курсов для интегратора МООС рунета. Предложены решения проблемы приведения запросов на естественном языке к виду, с которым могут работать методы построения рекомендаций, и проблемы сравнения запросов, включающих в себя синонимы. Описан подход, с помощью которого можно получить наиболее подходящие рекомендации курсов.

Ключевые слова: МООС, интегратор, рекомендательная система, фильтрация, взвешивание слов, степень схожести, синонимы.

Введение

Прототип интегратора [1] умеет находить в своей базе данных массовые открытые онлайн курсы с выбранными характеристиками и/или по слову, присутствующему в названии [1]. Поиск нужного курса начинается с формулировки того, чему хочет научиться пользователь. Потенциальный обучающийся часто не может точно указать предметную область курса, название и другие его характеристики. Чтобы найти нужный курс, пользователю нужно перебирать сочетания характеристик и запросов для поисковой строки, т.к. в названии могут присутствовать синонимы. Это займёт много времени. Можно сделать вывод, что для поиска МООС необходима рекомендательная система.

Рекомендательная система – это система, предоставляющая пользователям рекомендации по объектам, которые могли бы их заинтересовать. Используя возможности рекомендательных систем,

пользователи тратят значительно меньше времени на поиск нужного объекта [2]. Была поставлена задача создания рекомендательной системы для существующего MOOC интегратора.

1. Обзор существующих решений

Существует несколько проектов рекомендательных систем для MOOC. Были проанализированы методы, предложенные в этих проектах [3-5].

Один из интересных методов – персонализация рекомендаций на основе данных из социальной сети LinkedIn [3]. Это сеть для установления деловых контактов, в ней создаются группы по интересам, размещаются вакансии и резюме, есть возможность публиковать данные о конференциях, читаемых книгах и т.п. Эти данные можно использовать для рекомендации курсов, соответствующих профессиональным интересам пользователя. Предлагается следующий подход: рекомендательная система основывается на схожести предпочтений пользователей. Каждому пользователю присваиваются наборы профессиональных сфер, и чем больше у пользователей одинаковых интересов, тем вероятнее, что их интересуют одинаковые курсы [3].

Аналогичный подход предложен и в другой работе, но основная идея метода деления на кластеры заключается в таком подборе хеш-функций для некоторых измерений, чтобы похожие объекты с высокой степенью вероятности попадали в один кластер. В качестве векторов для пользователей используются оценки, данные ими определённым курсам [4].

Альтернативный метод предложен в работе [5]. Система рекомендует не курсы, а показывает список других, «похожих» на пользователя студентов. Системой предоставляется возможность переписки с этими студентами для обсуждения интересующей темы и взаимной рекомендации курсов [5].

В всех рассмотренных научных работах предложены методы, основанные на сходстве информации о пользователях – коллаборативной фильтрации. Существуют различные варианты реализаций этих методов и

вектор характеристик, тогда часть будет отброшена по несоответствию, например, предметной области.

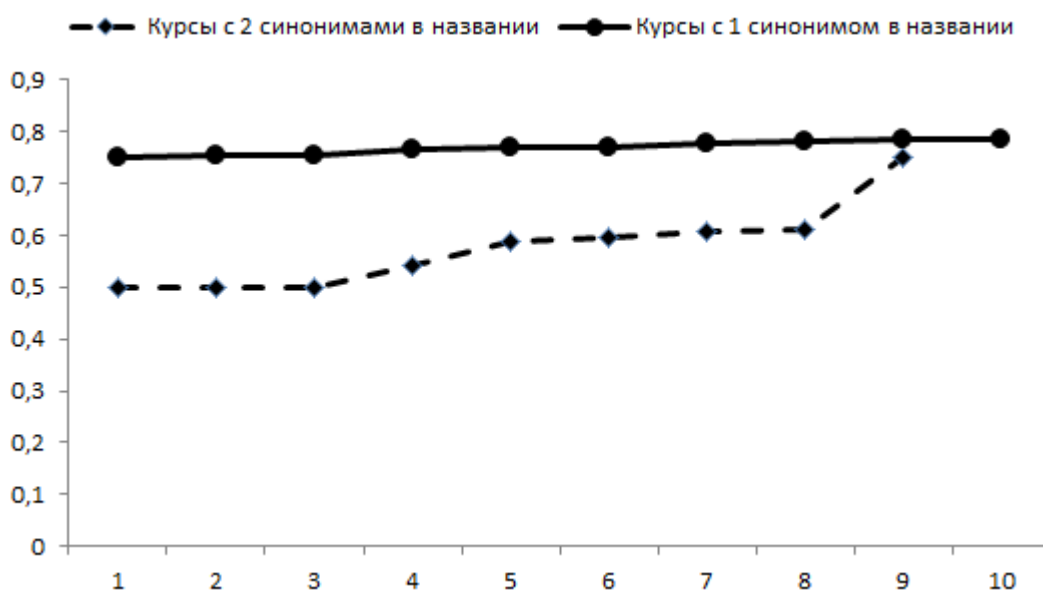


Рис. 2. График степеней схожести курсов с разным кол-вом синонимов

Но для поиска пользователь не обязан задавать характеристики. Следовательно, необходимо улучшить алгоритм выбора курсов, близких по названию. Проблему можно решить достаточно просто: список результатов с посчитанными значениями степени схожести нужно также делить на подсписки, сортировать по убыванию значений и при выдаче результатов брать оттуда только первые топ-N, которых не хватает до какого-то определенного количества рекомендаций. В этом случае, курсы с большим количеством синонимов точно окажутся выше в списке результатов, и нет необходимости считать степени схожести с коэффициентом.

СПИСОК ЛИТЕРАТУРЫ

1. Стольникова А.А., Дацун Н.Н. Проектирование MOOC-агрегатора для рунета. Математическое и информационное моделирование сборник научных трудов. Тюмень, 2017. С. 440-449.

2. Прохоров И.В., Мысев А.Э. Подходы к построению мультикритериальных рекомендательных систем, использующих неявные

оценки. Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. 2014. № 1. С. 33-36.

3. A New MOOCs' Recommendation Framework based on LinkedIn Data / K. Dai [et al.] 2017 Lecture Notes in Educational Technology. (9789811024184), p. 19-22.

4. Collaborative Filtering Recommendation for MOOC Application / Y. Pang [et al.] Computer Applications in Engineering Education 25(1), p. 120-128.

5. Who Wants to Chat on a MOOC? Lessons from a Peer Recommender System / F. Bouchet [et al.] Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10254 LNCS, p. 150-159.

6. Латкин И.И. Коллаборативная фильтрация в рекомендательных системах. SVD-разложение. Молодежный научно-технический вестник. 2015. № 7. С. 15-23.

7. Селивёрстов Е.В. Повышение качества рекомендательных систем за счет учета структуры документов. Nauka-Rastudent.ru. 2014. № 4 (04). С. 13-27.

8. Селяев А.Г. Взвешивание терминов в процессах индексирования электронных информационных ресурсов. Автоматизация процессов управления. 2007. № 2. С. 92-96.

9. Porter M.F. An algorithm for suffix stripping. Program electronic library and information systems. 1980. 14(3).

10. Модель определения нормальной формы слова для казахского языка. Вестник Новосибирского государственного университета. Серия: Информационные технологии. / А.М. Федотов [и др.] 2015. Т. 13. № 1. С.107-116.

11. George A. Miller. WordNet: A Lexical Database for English. Communications of the ACM. Vol. 38, No. 11. 1995, p. 39-41.

12. Александрова З.Е. Словарь синонимов русского языка: Практический справочник: Ок. 11 000 синоним. рядов. М.: Рус. яз., 2001. 568 с.