

М.С. Воробьева^{1,2}, Ю.В. Боганюк¹, Р.М. Якубов¹

¹ Тюменский государственный университет, г. Тюмень

² Научно-технический университет «Сириус», г. Сочи

УДК 004.912

РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НАВЫКОВ ДЛЯ СТУДЕНТА НА ОСНОВЕ ТЕКСТОВ ЕГО РАБОТ

Аннотация. В статье рассматривается метод построения рекомендаций по изучению новых технологий для студентов ИТ-направлений подготовки. В основе лежит анализ текстов студенческих работ и метод поиска ассоциативных правил, построенных с использованием данных, извлеченных из текстов вакансий. В результате разработано веб-приложение, реализующее функционал рекомендательной системы навыков.

Ключевые слова: анализ данных, поиск ассоциативных правил, рекомендательная система.

В настоящее время остро стоит вопрос о трудоустройстве молодежи, особенно выпускников вузов. По данным выборочного исследования Росстата около 20% студентов ИТ-специальностей, окончивших вузы в 2017-2019 годах трудоустроены не по специальности. Так студент-программист при поиске работы часто сталкивается с проблемой несоответствия имеющихся навыков или их отсутствия по сравнению с теми, которые указаны в интересующих его вакансиях.

Данную проблему можно решить путем выявления интересующей студента специализации и построения рекомендаций по изучению технологий, которые чаще всего используются в данной области. Технологии, которыми владеет студент, можно выявить из текстов его работ, а подходящие для изучения новые навыки можно определить путем поиска ассоциативных правил, на основе данных, полученных при анализе требований соответствующих специализаций на рынке труда.

В качестве источников данных для анализа актуальных требований рынка труда и профессиональной подготовки студентов были использованы тексты вакансий с перечислением требуемых навыков, а также тексты выпускных квалификационных работ (ВКР) ИТ-направлений подготовки студентов Института математики и компьютерных наук Тюменского государственного университета.

Анализ ВКР заключается в поиске ключевых слов – технологии, которые студент использовал в своей разработке. Предварительно текст подвергается обработке: приведение к нижнему регистру, удаление стоп-слов и токенизация по предложениям. Поиск использованных технологий в тексте осуществляется на основе подготовленного корпуса ключевых слов, специфичных для ИТ сферы. Таким образом, текст ВКР характеризуется совокупностью приобретенных студентом навыков.

Фундаментальной частью разработанной системы является модифицированный подход к формированию рекомендаций на основе задачи поиска ассоциативных правил, предложенный в исследовании [1]. В результате периодического анализа автоматически обновляемых данных о вакансиях, с помощью алгоритма FP-Growth формируется множество правил, каждое из которых представляет совокупность:

$$S = \{P, A, c\},$$

где P – множество компетенций,

A – множество ассоциативных компетенций к P ,

c – значение достоверности, что A соответствует P .

Модификация подхода заключается в формировании специального векторного пространства навыков, к которому применяется алгоритм KNN, для отбора рекомендаций для студента. Рассмотрим формальную постановку задачи:

Дано:

Множество ассоциативных правил $L = \{S_i \mid i = \overline{1, n}, n = |L|\}$

Множество навыков студента M .

Требуется составить множество рекомендаций:

$$R = \{r \mid r - \text{рекомендованная к изучению технология}\}.$$

Алгоритм получения рекомендаций состоит из 4 этапов:

Этап 1. Отбираем множества ассоциативных правил:

$$L' = \{S \mid M \subseteq P, |M \cap P| = \max |M \cap P|, P \in S, S \in L\}$$

Этап 2. Формируем упорядоченное множество всех навыков:

$$T = P \cup A \cup M, P \in S, A \in S, S \in L'$$

Формируем векторное пространство навыков V , где

$$v = (v_0, v_1, \dots, v_m),$$
$$v_i = \begin{cases} 1, & \text{если } t_i \in P \cup M, \\ c, & \text{если } t_i \in A \setminus M, c \in S, \\ 0, & \text{если } t_i \notin P \cup M, \end{cases}$$

$$m = |T|, t_i \in T, v \in V, S \in L', P \in L', A \in L', |V| = |L'| + 1$$

Этап 3. Вычисляем расстояние от v_0 до остальных векторов и выбираем k ближайших в множество V' .

Этап 4. Формируем множество $T' = \{t_i \mid t_i \in T, v_j^i \neq 0, v_j \in V'\}$.

Конструируем множество рекомендаций $R = T' \setminus M$.

Функционал рекомендательной системы. В качестве основного функционала рекомендательной системы выступает возможность получения списка рекомендаций по изучению новых технологий и освоения новых навыков со стороны студента на основе текстов его работ, предварительно загруженных в систему. Так, пользователь может выбрать одну из своих работ и получить список использованных им технологий, а также список рекомендованных к изучению технологий, которые наиболее близко подходят к уже приобретенным им компетенциям.

Пример реализации. В качестве подтверждения результатов можно привести примеры рекомендаций, полученных в ходе проведения исследования, в котором были проанализированы тексты 7877 вакансий ИТ-специализации. В результате анализа был получен список из 4770 ассоциативных правил, каждое

из которых содержит наиболее встречающиеся (более чем в 70% вакансий) множества компетенций специалистов.

Предположим, что пользователь выбрал текст курсовой работы по дисциплине «Структуры и алгоритмы компьютерной обработки данных», в результате анализа которого был получен список использованных технологий: ACID, Akka, C, C++, CI, concurrent programming, Elixir, Erlang, Git, GitLab, GitLab CI, Hadoop, JVM, Linux, NoSQL, Rust, Scala, UNIX, YAML. Большинство из этих технологий присущи DevOps-инженерам, поэтому пользователь получит рекомендации по изучению новых технологий, специфичных для данной специальности: SQL, CD, Docker, PostgreSQL (см. рис. 1).

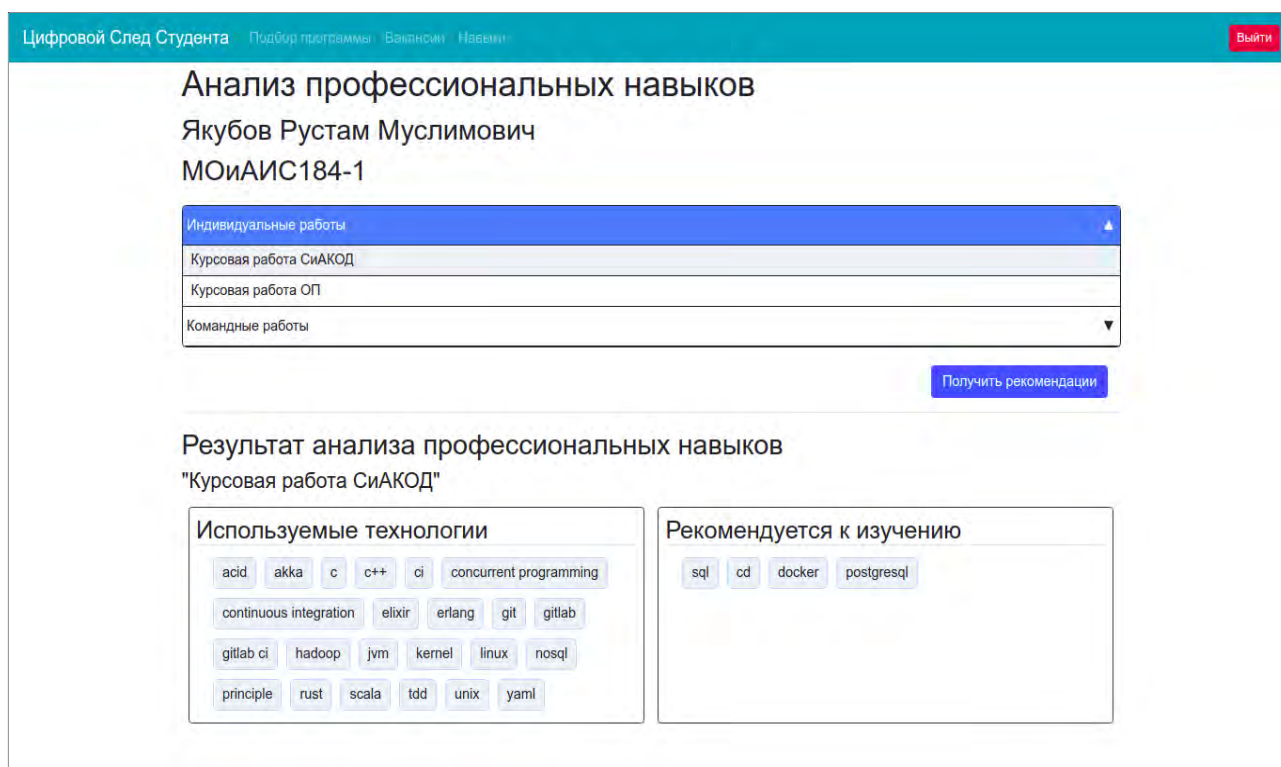


Рис. 1. Интерфейс рекомендательной системы

В результате была разработана рекомендательная система навыков для студента, основанная на анализе текстов его работ и актуальных требований рынка труда ИТ-специалистов. Анализ результатов показал эффективность предложенного подхода к формированию рекомендаций.

Благодарности

Статья подготовлена в рамках разработки образовательного кейса для НТУ Сириус при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028.

СПИСОК ЛИТЕРАТУРЫ

1. Боганюк Ю.В., Воробьев А.М. Рекомендательная система для проектирования образовательных программ в условиях индивидуальных образовательных траекторий [Электронный ресурс] / Ю.В. Боганюк, А.М. Воробьев. – Электрон. журн. – 2020. – С. 42-45. – Режим доступа: <https://elibrary.ru/item.asp?id=44034423&pff=1> (дата обращения: 07.04.2021).

2. Jiawei H., Jian P., Yiwen Y., Ruying M. Mining Frequent Patterns Without Candidate Generation [Электронный ресурс] // Proceedings of 2000 ACM SIGMOD International Conference on Management Data. – 2000. – Т. SIGMOD '00. – С. 1-12. – Режим доступа: <https://dl.acm.org/doi/10.1145/342009.335372> (дата обращения: 11.04.2021).