

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

УДК 371.267 013

DOI: 10.17853/1994-5639-2018-9-91-114

## МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ УПРАВЛЕНИЯ ПРОФЕССИОНАЛЬНЫМ РАЗВИТИЕМ СТУДЕНТОВ

**И. Г. Захарова**

*Тюменский государственный университет, Тюмень, Россия.*

*E-mail: i.g.zakharova@utmn.ru*

**Аннотация.** Введение. Профессиональное развитие студентов требует эффективного взаимодействия обучающихся с преподавателями, научными работниками и администрацией вуза, а также с другими обучающимися, представителями профессионального сообщества и агентами рынка труда. Результативность таких коммуникаций обусловлена их информационным обеспечением, оперативно предоставляющим достоверные сведения всем субъектам образовательного процесса.

*Цель* изложенного в статье исследования – изучение потенциала методов машинного обучения для действенного управления образовательным процессом на примере реализации компонента информационного обеспечения, предназначенного для диагностики и прогнозирования профессионального развития обучающихся посредством автоматического анализа их текстов.

*Методология и методы.* Теоретической базой исследования послужило моделирование профессионализации студентов с помощью компьютерного анализа информативности и профессиональной актуальности текстов их письменных работ. Для определения специфики профессионального развития разных групп будущих специалистов был задействован метод кластеризации k-средних. При построении модели классификации обучающихся применялся статистический метод распознавания Байеса.

*Результаты и научная новизна.* С тем чтобы оценить общий и специальный словарный запас учащихся, была выполнена программная обработка текстов, написанных на разных этапах образовательного процесса в вузе. Выявлены закономерности и особенности усвоения студентами общенаучной

и профессиональной терминологии. Исходя из этого сформированы группы учащихся, каждой из которых свойственны определенные тренды образовательного поведения. Продemonстрировано, как такая дифференциация, основанная на комплексе предварительно отобранных динамических показателей, характеризующих изменения профессионального лексикона, расширяет возможности диагностики и прогнозирования профессионального роста учащихся. Оговаривается, что эффективность подобных интеллектуальных систем определяется не только постоянным пополнением базы данных, от объема которых зависит точность модели классификации обучающихся и, соответственно, прогноза их профессионализации. Не менее важную роль играет совершенствование базы знаний, содержащей критерии профессионального развития, и соблюдение требования актуальности базовых словарей, поддержка которой может и должна осуществляться с участием представителей профессионального сообщества.

*Практическая значимость.* Предлагаемое информационное обеспечение управления профессиональным развитием учащихся может использоваться как для принятия оперативных решений, так и при разработке содержания и технологий образования. Благодаря такой системе студенты могут оценить результаты собственной подготовки, сопоставляя свои достижения с более ранними работами, работами однокурсников, целевыми показателями употребления общенаучной и профессиональной терминологии. Преподавателям данный компонент информационного обеспечения позволяет проводить мониторинг терминологической содержательности созданных студентами текстов и легко определять их авторство по частотности общего словаря обучающегося и характеру изменений используемой лексики. Представители рынка труда наряду с доступом к информации о текущей успеваемости студента могут определить его перспективы как будущего работника. Для руководителей образовательных программ и администрации вуза открываются новые возможности получать объективную информацию о содержательности тех или иных дисциплин, судя по тому как их изучение отражается на общем и профессиональном уровнях развития студентов.

**Ключевые слова:** профессиональное развитие, образовательный процесс, диагностика, прогноз, информационное обеспечение, машинное обучение, анализ текстов.

**Благодарности.** Исследование выполнено при финансовой поддержке гранта РФФИ, проект № 18-013-00106.

**Для цитирования:** Захарова И. Г. Методы машинного обучения для информационного обеспечения управления профессиональным развитием студентов // Образование и наука. 2018. Т. 20. № 9. С. 91–114. DOI: 10.17853/1994-5639-2018-9-91-114

## MACHINE LEARNING METHODS OF PROVIDING INFORMATIONAL MANAGEMENT SUPPORT FOR STUDENTS' PROFESSIONAL DEVELOPMENT

I. G. Zakharova

*Tyumen State University, Tyumen, Russia.*

*E-mail: i.g.zakharova@utmn.ru*

**Abstract.** *Introduction.* Professional development of students requires effective interaction with teachers, scientists, university administrators, students, representatives of professional community and labour market. The effectiveness of this interaction resulted from its information support, based on reliable information, promptly provided to all the members of learning process.

The *aim* of this paper was to study the machine learning methods potential for the effective management of learning process by the example of implementing information support component designed to diagnose and predict the professional development of students based on automatic text analysis.

*Methodology and research methods.* The theoretical basis of the research involved modelling of students' professional development using the analysis of textual informative and professional relevance in written works of students. To identify the characteristics of professional development, a computer cluster analysis of texts was carried out using the K-means method of clustering. The Bayes method was used to construct a model for classifying students from the standpoint of identified features.

*Results and scientific novelty.* A computer analysis of texts relating to different stages of learning for the evaluation of general and special vocabulary was performed. Regularities in the dynamics of students' use of general scientific and professional terminology were revealed. Accordingly, the groups with certain trends of educational behaviour of students were formed. It was shown how this differentiation, based on the complex of previously selected dynamic indicators characterising the changes of professional vocabulary, expands the possibilities for diagnostics and forecasting of professional growth of students. The author notes that the efficiency of similar intellectual systems is determined not only by the continued database up-dating, i.e. the amount of data in turn influence the accuracy of model of students' classification and, consequently, the forecast of students' professional development. Equally important is the improvement of knowledge base, which contains the criteria of professional development and complies with the requirement of basic dictionaries relevance. In addition, supportive procedures should be carried out with participating of the representatives of professional community.

*Practical significance.* The information support provided for the management of professional development of students can be used both for operational

decision making and developing content and technologies for educational process. This means students can evaluate the dynamics of own performance in comparison with earlier works, classmates' work, target indicators of the use of general scientific and professional terminology. This information management component allows teachers to monitor the content of texts and easily determine the authorship of content of learner's general frequency vocabulary and the dynamics of its change. The representatives of labour market along with access to information on the current progress of a student can define his or her prospects as a future worker. Heads of educational programmes, university administrators receive objective information about the content of disciplines as their study is reflected in the students' professional development.

**Keywords:** professional development, educational process, diagnostics, forecasting, information support, machine learning, text analysis.

**Acknowledgements.** The research project was performed with the financial support of the Russian Foundation for Basic Research, grant № 18-013-00106.

**For citation:** Zakharova I. G. Machine learning methods of providing informational management support for students' professional development. *The Education and Science Journal*. 2018; 9 (20): 91–114. DOI: 10.17853/1994-5639-2018-9-91-114

## Введение

Современное высшее образование, демонстрируя тренд к расширению – увеличению охвата студентов с различными способностями, мотивацией, профессиональными целями, становится все более многообразным [1]. Движение от типовых учебных планов и «усредненности» содержания и технологий образования к реализации индивидуальных образовательных траекторий (ИОТ) предполагает свободный выбор обучающимися не только дисциплин, но и способов их изучения – традиционных, онлайн, смешанных, формальных и неформальных.

Известная доля самостоятельности обучающихся в планировании обучения влечет дополнительную ответственность всех субъектов образовательного процесса за качество его результатов, профессиональное развитие и самоопределение студентов. Возникает естественный вопрос о том, при каких условиях образование, получаемое в соответствии с избранной ИОТ, обеспечит становление выпускника вуза как профессионала. Закономерный ответ – помимо учебно-методических и материально-технических ресурсов студенту необходима поддержка на уровне профессионально расширенной образовательной среды: взаимодействие с преподавателями, научными работниками и администрацией вуза, с другими

обучающимися, а также с профессиональным сообществом и представителями рынка труда. При этом успешность и степень результативности такого взаимодействия обуславливаются его информационным обеспечением, т. е. получением полной, достоверной и актуальной информации, доступной всем субъектам образовательного процесса для обратной связи, адекватной конкретной ситуации [2].

Основная задача эффективной поддержки управления динамическими процессами в различных системах (в том числе в образовании) заключается в обеспечении оперативного получения информации, необходимой для принятия соответствующих управленческих решений [3]. На основе данных, отражающих как потоки информации о субъектах образовательного процесса и целевых показателях профессионального развития, так и разнообразные информационные связи между этими потоками, можно строить интеллектуальные системы для поддержки управления процессом профессионального развития. Принципиально важным в данном случае представляется учет динамического характера данных. Действительно, многие компетенции осваиваются и, соответственно, диагностируются только в процессе практической работы (при прохождении производственной практики, выполнении практикоориентированных проектов и т. п.). При этом и само содержание профессиональных компетенций, и критерии, и целевые показатели профессионального развития в наше время очень динамичны. Это характерно не только для профессионального образования в области передовых (критических) технологий, но и для гуманитарных направлений – прежде всего, в контексте цифровизации.

Возвращаясь к задаче управления, заметим, что она сводится к циклическому решению следующих подзадач:

- 1) интерпретации данных с целью диагностики состояния компонентов системы;
- 2) прогнозу последствий для возможных вариантов развития этих состояний;
- 3) целенаправленному развитию с помощью управляющих воздействий.

В процессе профессионального развития перечисленные подзадачи прямо или опосредованно решаются всеми субъектами образовательного процесса. Это обстоятельство, и в первую очередь активность самих обучающихся в управлении своим образованием, требует развивать представление об информационном обеспечении. Так называемые «сырые» данные, полученные в результате измерения тех или иных показателей, могут быть

дополнены благодаря использованию методов искусственного интеллекта. Методы машинного обучения позволяют разрабатывать перманентно обучающиеся и уточняющиеся модели интерпретации (классификации, кластеризации) данных различного типа, как числовых, так и текстовых. В результате появляется возможность определять характерные особенности образовательного поведения и управлять профессиональным развитием.

Цель настоящего исследования заключается в иллюстрации данного подхода на примере реализации особого компонента информационного обеспечения для управления профессиональным развитием обучающихся на основе автоматического анализа текстов с помощью методов машинного обучения.

### **Обзор литературы**

Отечественные и зарубежные исследователи в последнее время акцентируют необходимость расширения инструментария и форм сопровождения образовательного процесса для обеспечения адекватной и оперативной оценки учебных достижений, личностного и профессионального развития как самими обучающимися, так и преподавателями.

Один из наиболее важных аспектов практической поддержки студентов в условиях реализации индивидуальных образовательных траекторий рассмотрен Э. Ф. Зеером и Е. Ю. Журловой, которые сконструировали модель сопровождения ИОТ и обосновали необходимость разработки навигационных средств в сочетании с инструментами оценки и самооценки результатов образовательной деятельности [4].

М. Ю. Катаев, А. М. Корилов и В. С. Мкртчян указывают на важность непрерывного (конечно, условно – в дополнение к традиционным текущему, промежуточному и итоговому контролю) сбора и обработки информации об учебной деятельности и предлагают концепцию автоматизированной информационной системы для непрерывного накопления и анализа всесторонних данных и их использования для персонализированной поддержки обучающихся [5].

В работах R. S. Baker, P. S. Inventado, G. Siemens подчеркивается, что сбор и анализ данных важен не только для интерпретации результатов освоения новых компетенций, но и для понимания того, как проходит образовательный процесс, какова его динамика для тех или иных категорий студентов [6, 7].

Привычные формальные показатели могут быть дополнены и иными данными. Так, Э. Ф. Зеер и Л. Н. Степанова демонстрируют широкие возможности портфолио студента как источника информации для мони-

торинга и активизации профессионального развития благодаря соотносению результатов собственной деятельности обучающихся с установленными критериями [8].

Ж. Beckmann и Р. Weber показывают действенность применения онлайн-форума сетевого курса не только для развития критического мышления, но и для его оценки [9].

И. Б. Смирнов, Е. В. Сивак и Я. Я. Козьмина предлагают использовать данные социальной сети «ВКонтакте» для анализа образовательного процесса, подчеркивая продуктивность интеграции этих данных с информацией на уровне образовательного учреждения [10].

В то же время помимо извлечения и анализа индикаторов активности студентов исследователи всё больше внимания уделяют анализу текстов, создаваемых в рамках образовательной деятельности, публикаций на учебных и профессиональных форумах, а также в группах социальных сетей.

В исследованиях А. А. Степаненко, К. С. Шиляева и З. И. Резановой, а также Г. В. Можяевой, А. В. Слободской, А. В. Фещенко и др. представлен компьютерный анализ текстов, размещенных учащимися в сети «ВКонтакте», который лег в основу построения модели для классификации профессиональных интересов [11–13]. При этом, как доказывают З. И. Резанова, А. С. Романов, Р. В. Мещеряков и Г. В. Напреенко, важны сами принципы выбора признаков текста, позволяющих как выполнять общую диагностику, так и учитывать специализацию и своеобразие текстов [14, 15].

Публикации S. A. Crossley, D. S. McNamara, K. S. Muldner, S. Skaliksky и др. иллюстрируют многогранные возможности применения методов искусственного интеллекта для построения когнитивных моделей и выявления креативности авторов текстов [16–18].

В работе V. Kovanovic и др. для оценки саморефлексии обучающихся предлагается использовать специальные индикаторы, выделенные в созданных ими текстах [19].

В качестве одного из наиболее важных направлений зарубежных исследований, результаты которых содержатся в работах K. Kyle, S. Crossley, C. Berger, необходимо отметить автоматический анализ и предварительное оценивание письменных работ (эссе) для обеспечения оперативной обратной связи и получения объективной информации о допущенных грамматических и орфографических ошибках, недочетах в общей организации текста и логике изложения, необоснованных утверждениях [20, 21].

В исследованиях M. Cantabella, E. D. de la Fuente, R. Martínez-España, L. Y. Li, C. C. Tsai, J. Liu, Z. Yang, X. Wang немаловажная роль отводится вы-

явлению определенных тенденций (шаблонов, паттернов) образовательного поведения, поскольку на их основе можно прогнозировать развитие обучающихся и реализовывать интеллектуальную обратную связь [22–24].

Таким образом, анализ литературы позволяет констатировать, что, несмотря на наличие целого комплекса инструментов для оценки качества образования, современные методы и технологии извлечения и анализа данных об образовательном процессе стимулируют появление новых подходов.

### **Материалы и методы**

Объектом нашего исследования являлся процесс профессионального развития бакалавров и магистрантов. Предмет исследования – моделирование профессионального развития студентов на основе анализа информативности и профессиональной актуальности созданных ими текстов отчетов и письменных работ. В этой связи критерием, характеризующим динамику развития профессиональных качеств, было выбрано расширение содержательности текстов в процессе обучения. Индикатором служило изменение как общего словарного запаса, так и актуальной профессиональной и общенаучной терминологии, используемых обучающимися. Для проверки гипотезы исследования о результативности применения методов машинного обучения для информационного обеспечения профессионального развития студентов применялись количественные и качественные методы.

Основным материалом для анализа на основе количественных методов послужил корпус документов, включающий тексты отчетов по практике (производственной и преддипломной) и научно-исследовательской работе (НИР), курсовых и выпускных квалификационных работ (ВКР), обзоров литературы и рецензий на отдельные статьи, выполненных студентами бакалавриата и магистратуры по направлению «Математическое обеспечение и администрирование информационных систем» (МОАИС) Института математики и компьютерных наук Тюменского государственного университета в 2013–2018 гг.

Модель прогноза профессионального развития строилась путем последовательного применения методов математической статистики и машинного обучения [25]. В первую очередь производились предварительная обработка и статистический анализ текстов для выделения признаков, характеризующих общий словарный запас и использование общенаучной и профессиональной терминологии. Затем на основе указанных признаков выполнялась кластеризация документов с помощью метода k-средних. В результате были выявлены группы обучающихся, для каждой из которых характерна особая

динамика изменений значений признаков. Это позволило выполнить разметку данных (дифференцировать обучающихся по определенным группам). В итоге с помощью метода Байеса была построена модель классификации для прогнозирования профессионального развития в соответствии с трендами, свойственными каждой группе.

Методы машинного обучения были реализованы с помощью свободно распространяемого программного обеспечения MyStem<sup>1</sup>, Natural Language Tool Kit [26] и SciKit Learn<sup>2</sup>.

С целью уточнения корректности результатов применения указанных методов и объективной оценки профессионального развития были проведены беседы с обучающимися, преподавателями, представителями рынка труда и проанализированы внешние отзывы о прохождении производственной практики и рецензии на ВКР, а также открытые данные, размещенные на персональных страницах обучающихся в социальной сети «ВКонтакте», отзывы с места постоянной работы студентов и выпускников.

### **Результаты исследования**

Исследование включало следующие основные этапы:

- 1) отбор и предварительная обработка данных;
- 2) выявление скрытых закономерностей для выделения особенностей профессионального развития;
- 3) построение прогнозной модели профессионального развития.

Рассмотрим их подробнее.

**Отбор и предварительная обработка данных.** На данном этапе исследования нами были отобраны и приведены к текстовому формату различные документы, подготовленные обучающимися и отражающие результаты выполнения ими профессионально ориентированных самостоятельных проектов по направлению подготовки МОАИС (табл. 1). Последовательность перечисления документов в таблице соответствует хронологии их представления на кафедру.

Дополнительно был сформирован набор документов, включающий следующие тексты:

- 1) подготовленные нами лекции по профессионально ориентированным дисциплинам («Языки программирования», «Алгоритмы и технологии параллельного программирования», «Современные технологии про-

---

<sup>1</sup> MyStem – Технологии Яндекса [Электрон. ресурс]. Режим доступа: <https://tech.yandex.ru/mystem/doc/> (дата обращения 20.08.2018).

<sup>2</sup> Documentation SciKit Learn – Machine Learning in Python Available from: <http://scikit-learn.org/stable/documentation.html> (дата обращения 20.08.2018).

граммирования», «Открытые технологии разработки программного обеспечения», «Разработка мобильных приложений», «Методология научных исследований» и др.);

2) рабочие программы для бакалавриата и магистратуры по направлению МОАИС по дисциплинам, производственной и преддипломной практике, курсовым работам и ВКР, научно-исследовательской работе, а также программы тренингов и онлайн-курсов (доступны на сайте образовательных программ Тюменского университета [or.utmn.ru](http://or.utmn.ru));

3) находящиеся в открытом доступе научные статьи, анонсы новых программных продуктов и технические руководства по современным технологиям, актуальным для указанных направлений подготовки (алгоритмы, языки и технологии программирования, мобильные и web-приложения, машинное обучение, компьютерная графика, базы данных и т. д.).

Таблица 1

Структура основного корпуса документов

Table 1

The structure of the text corpus

Уровень подготовки	Документы		
	Вид (семестр)	Код	Количество
Бакалавриат	Курсовая работа (4-й)	БК4	153
	Курсовая работа (6-й)	БК6	132
	Отчет по производственной практике (6-й)	БО6	132
	Рецензия на научную статью (8-й)	БР8	87
	Отчет по преддипломной практике (8-й)	БО8	82
	ВКР (8-й)	БВ8	82
Магистратура	Отчет по НИР (1-й)	МО1	63
	Отчет по НИР (2-й)	МО2	63
	Рецензия на научную статью (3-й)	МР3	61
	Обзор литературы по теме ВКР (3-й)	МЛ3	61
	Отчет по НИР (3-й)	МО3	61
	Отчет по преддипломной практике (4-й)	МО4	57
	ВКР (4-й)	МВ4	57

Нами была разработана компьютерная программа на языке Python, с помощью которой была выполнена предварительная обработка текста (удаление слов, относящихся к служебным частям речи (стоп-слова); выделение основы каждого слова), а затем на основании этого набора данных автоматически сформирован общий частотный словарь. Его образовали как отдельные слова, так и сочетания из двух слов, например «программирован(ие)» и «языки программирован(ия)», входящие в общий текст не ме-

нее двух раз. Объем этого словаря составил 3672 слова и словосочетания. Из него уже в «полуавтоматическом» режиме мы выделили два специализированных словаря – общенаучных и профессиональных терминов.

Безусловно, данные словари носят вспомогательный характер и ограничены по объему в соответствии с их назначением (по сравнению, например, с объемом словаря науки, включающем более 2600 терминов [27]). Кроме того, словарь общенаучных терминов дополнительно был сокращен с учетом специфики предметной области. Например, такие традиционно соответствующие общенаучной лексике термины, как «функция», «параметр», «метод», «класс», «объект», относятся к профессиональным терминам в области разработки компьютерных программ, а термины «признак», «прогноз», «классификация» являются профессиональными терминами для специалистов по компьютерному анализу данных и машинному обучению. Слова с подобными двойными значениями были отнесены к профессиональной лексике, поскольку тематика ряда курсовых работ, ВКР и НИР касалась данной области. Объем словаря общенаучных терминов составил 262 термина. В словарь профессиональных терминов наряду с русскими словами вошли англоязычные названия языков программирования, специализированных библиотек, программных продуктов и технологий (например, Java, C++, Python, Linux, Windows, OpenGL и др.). Объем данного словаря – 479 терминов.

Учитывая, что лексика студентов в процессе освоения новых компетенций расширяется, мы рассматривали эти словари как определенный ориентир для оценки динамики профессионального развития на основании мониторинга изменений словарного состава текстов обучающихся. Кроме использования выделенных общенаучных и профессиональных терминов, оценивались изменения общего словесного запаса.

**Выявление скрытых закономерностей для выделения особенностей профессионального развития.** На основе предварительно обработанных данных было выполнено сопоставление характеристик текстов, подготовленных студентами на разных этапах обучения, с составом словарей общенаучной и профессиональной лексики (далее «базовые словари»), а также дана оценка общего словарного запаса. С этой целью с помощью упомянутой выше компьютерной программы для каждого из текстов были сформированы наборы использованных в нем слов и словосочетаний, которые были сопоставлены с содержанием базовых словарей. Сводные результаты статистической обработки этих данных в соответствии с ранее приведенными кодами документов (см. табл. 1) представлены в табл. 2. Общий характер динамики использования общенаучной и профессиональной терминологии характеризуют средние значения ( $\mu$ ) и стандартные отклонения ( $\sigma$ ) доли

используемых терминов соответствующего базового словаря. Изменение в процессе обучения словарного запаса в целом иллюстрируется аналогичным образом на основе сопоставления с общим словарем.

Таблица 2

Сводные результаты статистической обработки текстов

Table 2

Summary results of statistical processing of the texts

Код	Общий словарный запас			Общенаучная лексика			Профессиональная лексика		
	$\mu$	$\sigma$	$\sigma/\mu, \%$	$\mu$	$\sigma$	$\sigma/\mu, \%$	$\mu$	$\sigma$	$\sigma/\mu, \%$
<i>Бакалавриат</i>									
БК4	0,15	0,006	4,0	0,09	0,004	4,4	0,11	0,004	3,6
БК6	0,19	0,015	7,9	0,11	0,006	5,5	0,21	0,02	9,5
БО6	0,17	0,009	5,3	0,1	0,005	5,0	0,22	0,03	13,6
БР8	0,17	0,008	4,7	0,13	0,007	5,4	0,19	0,015	7,9
БО8	0,2	0,012	6,0	0,14	0,01	7,1	0,27	0,03	11,1
БВ8	0,21	0,011	5,2	0,16	0,008	5,0	0,27	0,02	7,4
<i>Магистратура</i>									
МО1	0,19	0,005	2,6	0,14	0,007	5,0	0,26	0,012	4,6
МО2	0,22	0,004	1,8	0,17	0,006	3,5	0,27	0,011	4,1
МР3	0,21	0,004	1,9	0,19	0,002	1,1	0,19	0,009	4,7
МЛ3	0,25	0,006	2,4	0,26	0,01	3,8	0,28	0,015	5,4
МО3	0,27	0,007	2,6	0,26	0,009	3,5	0,3	0,012	4,0
МО4	0,27	0,01	3,7	0,26	0,015	5,8	0,3	0,014	4,7
МВ4	0,28	0,009	3,2	0,27	0,02	7,4	0,31	0,014	4,5

Характер изменения показателей, представленных в табл. 2, в целом подтверждает допустимость их использования для оценки динамики профессионального развития студентов направления МОАИС.

Как показывают приведенные данные, в процессе получения образования студенты демонстрируют в своих работах развитие общего и специального (общенаучного и профессионального) словарного запаса. При этом можно заметить определенные различия в приведенных значениях показателей для студентов бакалавриата и магистрантов, которые, впрочем, вполне очевидны.

В процессе обучения в бакалавриате студенты IT-направлений в большей степени ориентированы на приобретение профессиональных навыков, чем на участие в исследовательских проектах. Это объясняет отличия в динамике расширения профессиональной и общенаучной лексики. Иную логику можно видеть в изменениях аналогичных показателей у студентов магистратуры, где развитие общенаучного словарного запаса происходит более

интенсивно по сравнению с профессиональным. Данный факт обусловлен не только обязательностью НИР магистрантов, отчеты по которой предоставляются в каждом семестре. Опросы студентов и анализ дополнительных материалов показали результативность инициативной работы над заявками и докладами для участия в научных конференциях и конкурсах («Умник», «Стипендиальная программа Благотворительного фонда В. Потанина», программы академической мобильности и др.) и, наконец, нацеленность студентов на продолжение обучения в аспирантуре. Профессиональная лексика меняется в меньшей степени, поскольку большинство студентов к этому времени уже определило узкую специализацию в области IT.

Подтверждение общей тенденции профессионального развития тем не менее не столь актуально для информационного обеспечения управления этим процессом по сравнению с необходимостью выявления особых путей этого развития. О наличии указанных особенностей свидетельствуют большие значения стандартных отклонений для показателей, относящихся к 3-му и 4-му курсам бакалавриата, а также к последнему этапу обучения в магистратуре.

#### **Построение прогнозной модели профессионального развития.**

Заключительный этап исследования был связан с изучением возможностей прогнозирования профессионального развития на основе выделенных нами признаков.

Единая универсальная модель не сможет учесть объективно существующих отличий или будет весьма приблизительной, поэтому возникла задача выявления групп обучающихся, которым свойственна особая динамика изменения показателей. С этой целью предварительно были отобраны тексты определенных студентов, прошедших обучение в бакалавриате (82 чел.) и магистратуре (57 чел.). По этим текстам для каждого обучающегося вычислялись дополнительные характеристики – хронологически изменяющиеся показатели поэтапного накопления общего словарного запаса, а также общенаучных и профессиональных терминов. В качестве ориентира при вычислении относительных значений (долей), как и на предыдущем этапе, использовались базовые словари.

В итоге для каждого студента были получены наборы значений трех признаков, характеризующих в избранном нами аспекте поэтапное профессиональное развитие (в бакалавриате – 6 этапов, в магистратуре – 7). Согласно этим данным с помощью разработанной нами компьютерной программы была выполнена кластеризация для выделения групп студентов с общим характером хронологического изменения показателей.

Результаты вычислений показали, что и для бакалавриата, и для магистратуры можно достаточно точно выделить по три кластера, каждый

из которых характеризуется собственной тенденцией изменения соответствующих показателей.

На рис. 1а – 1с (бакалавриат), 2а – 2с (магистратура) показана динамика изменения общего, общенаучного и профессионального словарного запаса, использованного при выполнении соответствующих работ (табл. 1). Показатели определялись на основании средних по кластерам значений доли слов из базовых словарей (соответственно общего, общенаучного и профессионального).

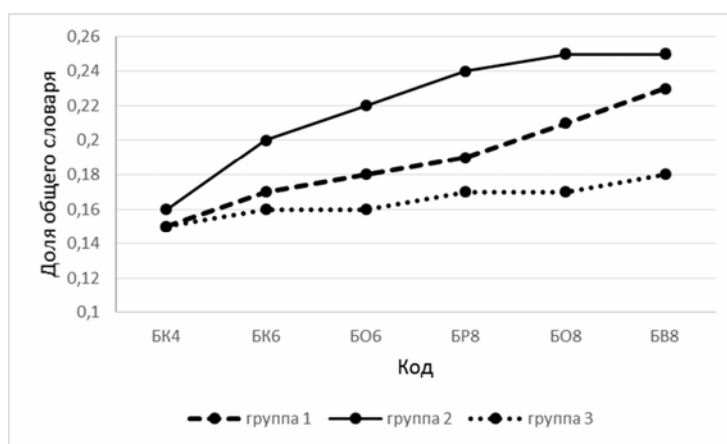


Рис. 1а. Динамика изменения общего словарного запаса  
Fig. 1a. Dynamics of the general lexicon changing

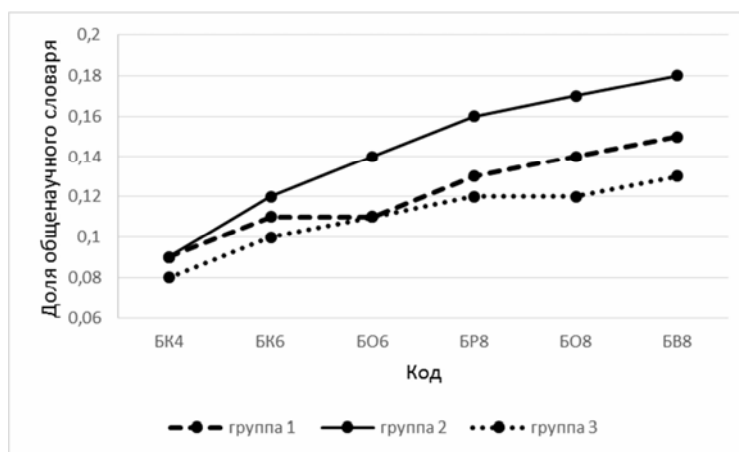


Рис. 1б. Динамика использования общенаучной лексики  
Fig. 1b. The dynamics of the use of general scientific lexicon

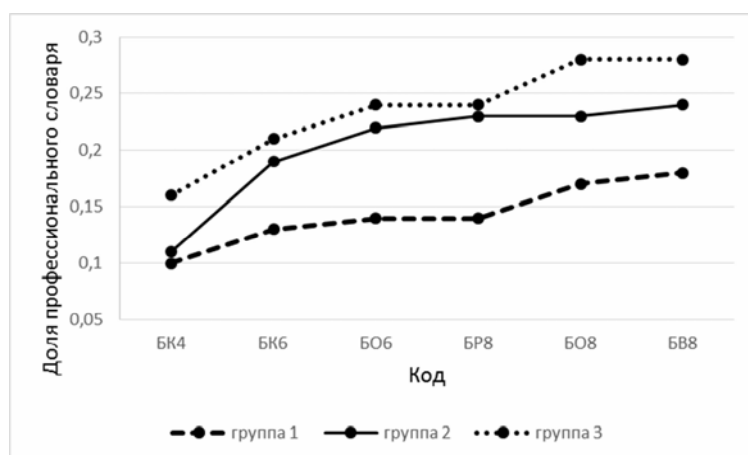


Рис. 1с. Динамика использования профессиональной лексики  
Fig. 1c. The dynamics of the use of professional lexicon

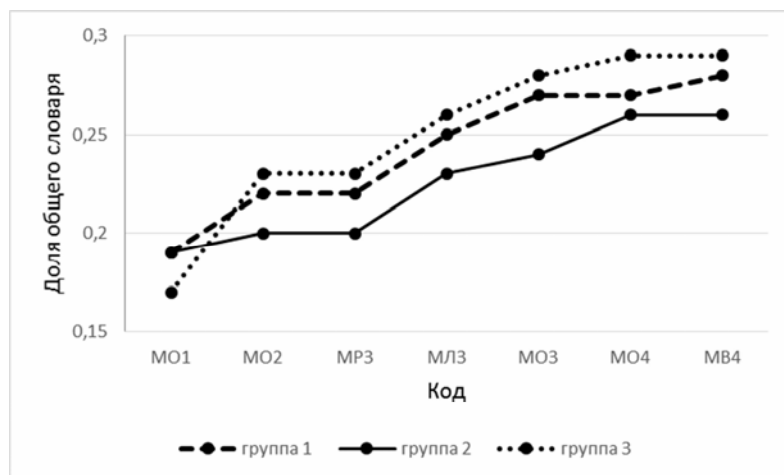


Рис. 2а. Динамика изменения общего словарного запаса  
Fig. 2a. Dynamics of the general lexicon changing

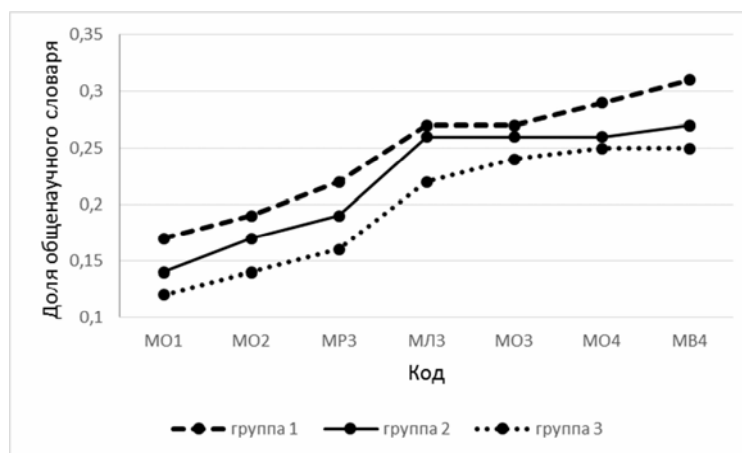


Рис. 2b. Динамика использования общенаучной лексики  
Fig. 2b. The dynamics of the use of general scientific lexicon

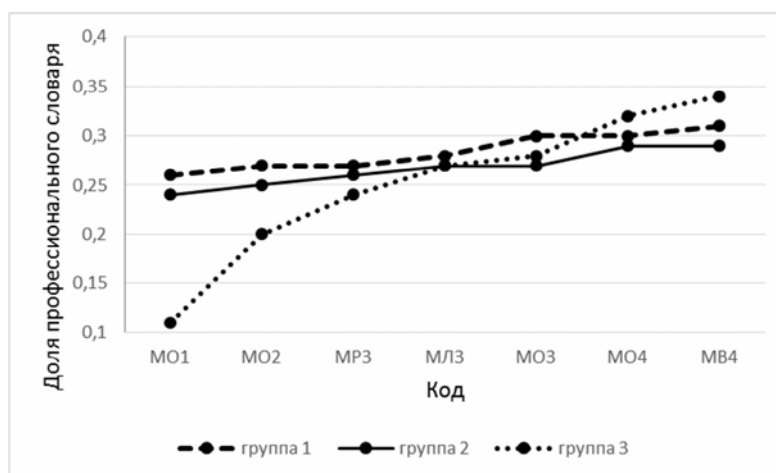


Рис. 2с. Динамика использования профессиональной лексики  
Fig. 2c. The dynamics of the use of professional lexicon

Анализ результатов учебной и внеучебной деятельности обучающихся (участие в олимпиадах, конкурсах, конференциях), особенностей их трудоустройства, беседы с руководителями производственной практики и представителями организаций, в которых работают студенты и выпускники направления МОАИС, позволили объяснить выявленные тенденции

и дать качественное описание каждой из групп студентов, отнесенных к выделенным кластерам.

*Бакалавриат.*

Группа 1 (29 чел., 35,4% от общего числа испытуемых) – студенты, нацеленные только на учебу строго в рамках курсов образовательной программы. Имеют, как правило, средние показатели успеваемости, опыт практической деятельности – только в период практики. Отказываются от предложений работы по специальности.

Группа 2 (30 чел., 36,6%) – студенты, ориентированные на профессиональную карьеру, со средними и высокими показателями успеваемости, совмещающие учебу с работой по специальности в средней или крупной компании не менее года на 3–4-м курсах. Многие слушают дополнительные онлайн-курсы, участвуют в профессиональных тренингах.

Группа 3 (23 чел., 28%) – студенты с различными показателями успеваемости, совмещающие учебу с работой по специальности в небольших компьютерных фирмах или в качестве фрилансеров. Многие стихийно занимаются профессиональным самообразованием, используя технические руководства и участвуя в онлайн-сообществах. Основная цель – зарабатывание денег.

*Магистратура* (практически все студенты работают по специальности).

Группа 1 (19 чел., 33,3% от общего числа испытуемых) – студенты с высокими показателями успеваемости, имеющие опыт участия в исследовательских проектах в период обучения в бакалавриате и ориентированные на дальнейшее обучение в аспирантуре и научно-педагогическую деятельность.

Группа 2 (22 чел., 38,6%) – студенты со средними и высокими показателями успеваемости, работают более двух лет в крупной организации и нацелены на профессиональную карьеру в области проектного менеджмента.

Группа 3 (16 чел., 28,1%) – студенты с различными показателями успеваемости, работают в соответствии с направлениями подготовки уровня бакалавриата, существенно отличающимися от направлений подготовки МОАИС (экономика, педагогическое образование, строительство и др.). Выбор магистерской программы обосновывают желанием приобрести новые компетенции для решения прикладных задач.

Представленные зависимости в совокупности определяют прогнозную модель профессионального развития на основании принадлежности обучающегося к той или иной группе. Исходя из этих данных была по-

строена модель классификации для последующего отнесения других студентов к той или иной группе. Перекрестное тестирование подтвердило работоспособность модели (точность составила 73,4%).

Следует заметить, что для оперативного управления образовательным процессом очень важно отслеживание характера изменений используемых словарей для каждого обучающегося. Введенные нами динамические показатели дают дополнительную информацию для более точного определения авторства текста по сравнению с использованием только словарного запаса.

### **Обсуждение и заключение**

В проведенном исследовании мы опирались на эмпирическую базу, связанную с конкретным направлением подготовки бакалавров и магистров. Безусловно, наряду с инвариантными, не зависящими от направления подготовки компонентами, при проектировании информационного обеспечения необходимо учитывать особенности образовательных программ разных специальностей. В частности, диагностика и прогнозирование профессионального развития будущих педагогов представляет собой одни из наиболее сложных процессов. Наряду с развитием компетентности учителя (преподавателя) в своей предметной области (математика, физика, информатика и др.) следует диагностировать его готовность к осознанному выполнению функций воспитателя, тьютора, модератора, исследователя, менеджера и т. д. Для решения этих задач в рамках предложенного подхода потребуется, как минимум, ввести более сложную структуру базового профессионального словаря.

Возможно, при постоянном контакте преподавателей и представителей рынка труда, тесном общении тех и других со студентами, тщательном анализе и самоанализе образовательных достижений и проблем, коллективных обсуждениях результатов различных самостоятельных проектов описанные нами модели не дадут ничего нового никому из субъектов образовательного процесса. Однако реальность обычно отличается от идеальной картины.

На наш взгляд, предлагаемый подход может стать основой автоматизированной системы информационного обеспечения управления профессиональным развитием студентов вуза и использоваться как для принятия оперативных решений, так и при разработке содержания и технологий образования.

Студент, предоставляя такой системе информацию в виде текстов документов, подготовленных по результатам выполнения курсовых работ,

прохождения практики и т. д., мог бы отслеживать и рефлексировать то, как эти тексты (на уровне словарного запаса) соотносятся с его более ранними работами, работами его однокурсников, целевыми показателями использования общенаучной и профессиональной терминологии.

Преподаватели, в свою очередь, могли бы не только выполнять мониторинг терминологической содержательности текстов, но и легко определять авторство по содержанию общего частотного словаря обучающегося и динамике его изменения. В рамках онлайн-обучения такое информационное обеспечение ориентировало бы преподавателей и тьюторов относительно особенностей индивидуальной образовательной траектории и обеспечивало бы адекватную обратную связь.

Представители рынка труда наряду с текущими результатами студента могли бы видеть потенциал его профессионального развития.

Наконец, руководители образовательных программ и администрация вуза имели бы возможность получать объективную информацию о содержательности тех или иных обязательных и элективных курсов благодаря показателям того, как их изучение отражается на профессиональном и общем развитии студентов.

Безусловно, эффективность подобных интеллектуальных систем определяется не только постоянным пополнением базы данных, от объема которых зависит точность модели классификации и, соответственно, прогноза. Не менее важную роль играет совершенствование базы знаний, содержащей правила (критерии) профессионального развития. Кроме того, результативность представленного подхода обусловлена поддержкой актуальности базовых словарей, которая может и должна обеспечиваться участием в такой работе представителей профессионального сообщества.

### **Список использованных источников**

1. Тайхлер У. Многообразие и диверсификация высшего образования: тенденции, вызовы и варианты политики // Вопросы образования. 2015. № 1. С. 14–38.
2. Захарова И. Г. Big Data и управление образовательным процессом // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. 2017. № 1. С. 210–219.
3. Шориков А. Ф. Методология моделирования многоуровневых систем: иерархия и динамика // Прикладная информатика. 2006. № 1. С. 136–141.
4. Зеер Э. Ф., Журлова Е. Ю. Навигационные средства как инструменты сопровождения освоения компетенций в условиях реализации индивидуальной образовательной траектории // Образование и наука. 2017. № 3. С. 77–93 [Электрон. ресурс]. Режим доступа: <https://doi.org/10.17853/1994-5639-2017-3-77-93> (дата обращения: 25.08.2018)

5. Катаев М. Ю., Корилов А. М., Мкртчян В. С. Концепция и структура автоматизированной системы мониторинга качества обучения студентов // Образование и наука. 2017. № 10. С. 30–46 [Электрон. ресурс]. Режим доступа: <https://doi.org/10.17853/1994-5639-2017-10-30-46> (дата обращения: 25.08.2018)
6. Siemens G., Baker R. S. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration // Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. April 29 – May 02. 2012. Vancouver, BC, Canada. New York: ACM, 2012. P. 252–254.
7. Baker R. S., Inventado P. S. Educational Data Mining and Learning Analytics // Learning Analytics: from Research to Practice. Berlin: Springer, 2014. P. 61–75.
8. Зеер Э. Ф., Степанова Л. Н. Портфолио как инструментальное средство самооценивания учебно-профессиональных достижений студентов // Образование и наука. 2018. № 6. С. 139–157 [Электрон. ресурс]. Режим доступа: <https://doi.org/10.17853/1994-5639-2018-6-139-157> (дата обращения: 25.08.2018)
9. Beckmann J., Weber P. Cognitive Presence in Virtual Collaborative Learning: Assessing and Improving Critical Thinking in Online Discussion Forums // Interactive Technology and Smart Education. 2016. № 1. P. 52–70.
10. Смирнов И. Б., Сивак Е. В., Козьмина Я. Я. В поисках утраченных профилей: достоверность данных «ВКонтакте» и их значение для исследований образования // Вопросы образования. 2016. № 4. С. 106–122.
11. Степаненко А. А., Шиляев К. С., Резанова З. И. Атрибуция профессиональных интересов пользователей социальной сети «ВКонтакте» на основе текстов тематических групп и персональных страниц // Вестник Томского государственного университета. Филология. 2018. № 52. С. 130–144.
12. Можаяева Г. В., Слободская А. В., Феценко А. В. Информационный потенциал социальных сетей для выявления образовательных потребностей школьников // Открытое и дистанционное образование. 2017. № 3. С. 25–29.
13. Feshchenko A., Goiko V., Stepanenko A. Recruiting University Entrants via Social Networks // Proceedings 9th International Conference on Education and New Learning Technologies. July 03–05.2017. Barcelona, Spain. Valencia: IATED, 2017. P. 6077–6082.
14. Резанова З. И., Романов А. С., Мещеряков Р. В. О выборе признаков текста, релевантных в автороведческой экспертной деятельности // Вестник Томского государственного университета. Филология. 2013. № 6. С. 38–52.
15. Напреенко Г. В. Идентификация текста по его авторской принадлежности на лексическом уровне (формально-количественная модель) // Вестник Томского государственного университета. 2014. № 379. С. 17–23.
16. McNamara D. S., Crossley S. A., Roscoe R. D., Allen L. K., Dai J. A Hierarchical Classification Approach to Automated Essay Scoring // Assessing Writing. 2015. № 23. P. 35–59.
17. Skalicky S., Crossley S. A., McNamara D. S., Muldner K. Identifying Creativity During Problem Solving Using Linguistic Features // Creativity Research Journal. 2017. № 4. P. 343–353.

18. Crossley S. A., Muldner K., McNamara D. S. Idea Generation in Student Writing: Computational Assessments and Links to Successful Writing // *Written Communication*. 2016. № 3. P. 328–354.
19. Kovanović V., Joksimović S., Mirriahi N., Blaine E., Gašević D., Siemens G., Dawson S. Understand Students' Self-Reflections through Learning Analytics // *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. March 07–09. 2018. Sydney, NSW, Australia. New York: ACM, 2018. P. 389–398.
20. Kyle K., Crossley S., Berger C. The Tool for the Automatic Analysis of Lexical Sophistication (TAALES): Version 2.0 // *Behavior Research Methods*. 2018. № 3. P. 1030–1046.
21. Kyle K., Crossley S., Berger C. Formative Essay Feedback Using Predictive Scoring Models // *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. August 13–17. 2017. Halifax, NS, Canada. New York: ACM, 2017. P. 2071–2080.
22. Cantabella M. de la Fuente E. D., Martínez-España R., Ayuso B., Muñoz A. et al. Searching for Behavior Patterns of Students in Different Training Modalities through Learning Management Systems // *Proceedings of the International Conference on Intelligent Environments (IE)*. August 21–25. 2017. Seoul. South Korea. New York: IEEE, 2017. P. 44–51.
23. Li L. Y., Tsai C. C. Accessing Online Learning Material: Quantitative Behavior Patterns and Their Effects on Motivation and Learning Performance // *Computers & Education*. 2017. № 114. P. 286–297.
24. Liu J., Yang Z., Wang X., Zhang X., Feng J. An Early-Warning Method on e-Learning // *E-Learning, E-Education and Online Training. Lecture Notes of the Institute for Computer Sciences. Social Informatics and Telecommunications Engineering*. 2018. Vol. 243. Cham: Springer, 2018. P. 62–72.
25. Козьмо А. П., Ричарт В. Построение систем машинного обучения на языке Python. Москва: ДМК Пресс, 2016. 302 с.
26. Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Sebastopol. CA: O'Reilly Media. 2009. 504 с.
27. Сарыбеков М. Н., Сыдыкназаров М. К. Словарь науки. Общонаучные термины и определения, науковедческие понятия и категории. Алматы: ТРИУМФ-Т, 2008. 504 с.

## References

1. Teichler U. Diversity and diversification of higher education: Trends, challenges and policy options. *Voprosy obrazovaniya = Educational Studies*. 2015; 1: 14–38. (In Russ.)
2. Zakharova I. G. Big Data and educational process management. *Vestnik Tyumenskogo Gosudarstvennogo Universiteta. Gumanitarnye Issledovaniya. Humanitates = Tyumen State University Herald. Humanities Research. Humanitates*. 2017; 1: 210–219. (In Russ.)

3. Shorikov A. F. Methodology for modeling multi-level systems: hierarchy and dynamics. *Prikladnaya Informatika = Applied Informatics*. 2006; 1: 136–141. (In Russ.)
4. Zeer E. F., Zhurlova E. Y. Navigation aids as tools to support the development of competences in the conditions of realization of individual educational trajectory. *Obrazovanie i nauka = The Education and Science Journal* [Internet]. 2017 [cited 2018 Aug 25]; 19 (3): 77–93. Available from: <https://doi.org/10.17853/1994-5639-2017-3-77-93> (In Russ.)
5. Kataev M. Y., Korikov A. M., Mkrttchian V. S. Concept and structure of automated system for monitoring student learning quality. *Obrazovanie i nauka = The Education and Science Journal* [Internet]. 2017 [cited 2018 Aug 25]; 19 (10): 30–46. Available from: <https://doi.org/10.17853/1994-5639-2017-10-30-46> (In Russ.)
6. Siemens G., Baker R. S. Learning analytics and educational data mining: towards communication and collaboration. In: *Proceedings of the 2<sup>nd</sup> International Conference on Learning Analytics and Knowledge*; 2012 Apr 29 – May 02; Vancouver, BC, Canada. New York: ACM; 2012. p. 252–254.
7. Baker R. S., Inventado P. S. Educational data mining and learning analytics. In: J. A. Larusson & B. White (Eds). *Learning Analytics: From Research to Practice*. Berlin. Germany: Springer; 2014. p. 61–75.
8. Zeer E. F., Stepanova L. N. Portfolio as an instrumental means of self-evaluation of educational and professional achievements of students. *Obrazovanie i nauka = The Education and Science Journal* [Internet]. 2018 [cited 2018 Aug 25]; 20 (6): 139–157. Available from: <https://doi.org/10.17853/1994-5639-2018-6-139-157> (In Russ.)
9. Beckmann J., Weber P. Cognitive presence in virtual collaborative learning: assessing and improving critical thinking in online discussion forums. *Interactive Technology and Smart Education*. 2016; 1: 52–70.
10. Smirnov I. B., Sivak E. V., Koz'mina Y. A. In search of lost profiles: The reliability of VKontakte data and their significance for educational research. *Voprosy obrazovaniya = Educational Studies*. 2016; 4: 106–122. (In Russ.)
11. Stepanenko A. A., Shilyaev K. S., Rezanova Z. I. Attribution of professional interests of users of social network VKontakte on the basis of texts of thematic groups and personal pages. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Filologiya = Tomsk State University Journal of Philology*. 2018; 52: 130–144. (In Russ.)
12. Mozhaeva G. V., Slobodskaya A. V., Feshchenko A. V. Information potential of social networks for revealing educational needs of schoolchildren. *Otkrytoe i distantsionnoe obrazovanie = Open and Distance Education*. 2017; 3: 25–29. (In Russ.)
13. Feshchenko A., Goiko V., Stepanenko A. Recruiting university entrants via social networks. In: *Proceedings 9<sup>th</sup> International Conference on Education and New Learning Technologies*; 2017 July 03 – 05; Barcelona, Spain. Valencia: IATED; 2017. p. 6077–6082.
14. Rezanova Z. I., Romanov A. S., Meshcheryakov R. V. On the choice of signs of the text relevant in the author expert's activity. *Vestnik Tomskogo Gosu-*

darstvennogo Universiteta. *Filologiya* = Tomsk State University Journal of Philology. 2013; 6: 38–52 (In Russ.)

15. Napreenko G. V. Identification of the text by its author's affiliation at the lexical level (formal-quantitative model). *Vestnik Tomskogo Gosudarstvennogo Universiteta* = Tomsk State University Journal. 2014; 379: 17–23. (In Russ.)

16. McNamara D. S., Crossley S. A., Roscoe R. D., Allen L. K., Dai J. A hierarchical classification approach to automated essay scoring. *Assessing Writing*. 2015; 23: 35–59.

17. Skalicky S., Crossley S. A., McNamara D. S., Muldner K. Identifying creativity during problem solving using linguistic features. *Creativity Research Journal*. 2017; 4: 343–353.

18. Crossley S. A., Muldner K., McNamara D. S. Idea generation in student writing: computational assessments and links to successful writing. *Written Communication*. 2016; 3: 328–354.

19. Kovanović V., Joksimović S., Mirriahi N., Blaine E., Gašević D., Siemens G., et al. Understand students' self-reflections through learning analytics. In: *Proceedings of the 8<sup>th</sup> International Conference on Learning Analytics and Knowledge*; 2018 Mar 07 – 09; Sydney, NSW, Australia. New York: ACM; 2018. p. 389–398.

20. Kyle K., Crossley S., Berger C. The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*. 2018; 3: 1030–1046.

21. Kyle K., Crossley S., Berger C. Formative essay feedback using predictive scoring models. In: *Proceedings of the 23<sup>rd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2017 Aug 13 – 17; Halifax, NS, Canada. New York: ACM; 2017. p. 2071–2080.

22. Cantabella M. de la Fuente E. D., Martínez-España R., Ayuso B., Muñoz A., et al. Searching for behavior patterns of students in different training modalities through learning management systems. In: *Proceedings of the International Conference on Intelligent Environments (IE)*; 2017 Aug 21 – 25; Seoul, South Korea. New York: IEEE; 2017. p. 44–51.

23. Li L. Y., Tsai C. C. Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education*. 2017; 114: 286–297.

24. Liu J., Yang Z., Wang X., Zhang X., Feng J. An early-warning method on e-learning. *E-Learning, E-Education, and Online Training, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. V. 243. Cham: Springer; 2018. p. 62–72.

25. Koehl'o L. P., Richart V. Postroenie sistem mashinnogo obucheniya na yazyke Python = Building machine learning systems with Python. Moscow: Publishing House DMK Press; 2016. 302 p. (In Russ.)

26. Bird S., Klein E., Loper E. Natural language processing with Python: Analyzing text with the natural language toolkit. Sebastopol, CA: O'Reilly Media. Inc.; 2009. 504 p.

27. Sarybekov M. N., Sydyknazarov M. K. Slovar' nauki. Obshhenauchnye terminy i opredeleniya, nauchovedcheskie ponyatiya i kategorii = Dictionary of science. General scientific terms and definitions, science-related concepts and categories. Almaty: Publishing House TRIUMF-T; 2008. 504 p. (In Russ.)

***Информация об авторе:***

**Захарова Ирина Гелиевна** – доктор педагогических наук, профессор, профессор кафедры программного обеспечения Института математики и компьютерных наук Тюменского государственного университета; ORCID ID 0000-0002-4211-7675, Scopus Author ID 57062143300; Тюмень, Россия. E-mail: i.g.zakharova@utmn.ru

Статья поступила в редакцию 22.06.2018; принята в печать 17.10.2018.  
Автор прочитал и одобрил окончательный вариант рукописи.

***Information about the author:***

**Irina G. Zakharova** – Doctor of Pedagogical Sciences, Professor, Software Department, Institute of Mathematics and Computer Science, Tyumen State University; ORCID ID 0000-0002-4211-7675. Scopus Author ID – 57062143300; Tyumen, Russia. E-mail: i.g.zakharova@utmn.ru

Received 22.06.2018; accepted for publication 17.10.2018.  
The author has read and approved the final manuscript.