
Открытое и дистанционное образование

№ 4 (68)

Научно-методический журнал
Свидетельство о регистрации ПИ №77-12619 от 14 мая 2002 г.

2017 г.

СОДЕРЖАНИЕ

От редакции	3
Педагогика и психология открытого и дистанционного образования	
<i>Малахова О.Н., Жученко О.А.</i> Мониторинг психических состояний студентов в условиях дистанционной и аудиторной образовательной коммуникации	5
<i>Юхта Н.М.</i> Феномен современного дистанционного образования	13
Электронные средства учебного назначения	
<i>Кирьянова Л.В., Осипов Ю.В., Сафина Г.Л.</i> Анализ профилей в тестах по математике	17
Методологическое, научно-методическое и кадровое обеспечение информатизации образования	
<i>Захарова У.С., Можая Г.В., Бабанская О.М., Танасенко К.И.</i> Развитие онлайн-обучения в программе Томского регионального центра компетенций в области онлайн-обучения	25
<i>Маркова Л.А.</i> Информационно-образовательная среда методических школьных объединений как подсистема ИОС школы	32
<i>Морозова Ю.В., Уртамова И.А.</i> Методика анализа электронного учебного контента	38
<i>Гойко В.Л., Киселев П.Б., Мацута В.В., Суханова Е.А., Степаненко А.А., Фещенко А.В.</i> Методы и инструменты выявления перспективных абитуриентов в социальных сетях	45
Социально-гуманитарные проблемы информатизации образования	
<i>Яковлев В.Ф.</i> Эффективные меры противодействия использованию студентами услуг гострайтеров при дистанционном обучении	53
<i>Степаненко А.А., Фещенко А.В.</i> «Цифровой след» студента: поиск, анализ, интерпретация	58
Автоматизированные информационные системы в образовании и науке	
<i>Карнаухов В.М.</i> Адаптивный метод нечетких множеств	63
Наши авторы	70

Open and distance education

№ 4 (68)

Scientifically-methodical magazine
the Certificate of registration PI №77-12619 from May, 14th 2002

2017

CONTENT

Editorial Note	4
Pedagogics and psychology of open and distance education	
<i>Malakhova O.N., Zhuchenko O.A.</i> Monitoring of mental states of students in distance and classroom communication	5
<i>Yukhta N.M.</i> The phenomenon of the modern distance education	13
Electronic educational means	
<i>Kirianova L.V., Osipov Yu.V., Safina G.L.</i> Analysis of patterns in mathematical tests	17
Methodological, scientific and methodical and staff provision of educational informatization	
<i>Zakharova U.S., Mozhaeva G.V., Babanskaya O.M., Tanasenko K.I.</i> Development of online learning according to the program of tomsk regional center for competencies in the field of online learning	25
<i>Markova L.A.</i> Information and educational environment of methodical school associations as a subsystem of a school	32
<i>Morozova Y.V., Urtamova I.A.</i> Methodology of analysis of electronic educational content	38
<i>Gojko V.L., Kiselev P.B., Matsuta V.V., Sukhanova E.A., Stepanenko A.A., Feshchenko A.V.</i> Methods and tools to identify promising entrants in social networks	45
Social-humanitarian problems of educational informatization	
<i>Yakovlev V.F.</i> Effective preventive measures for using ghostwriters services by distance learning students	53
<i>Stepanenko A.A., Feshchenko A.V.</i> Digital footprint of the student: search, analysis, interpretation	58
The automated information systems in formation and a science	
<i>Karnaukhov V.M.</i> Adaptive method of fuzzy sets	63
Our authors	70

В.Л. Гойко¹, П.Б. Киселев², В.В. Мацута¹, Е.А. Суханова¹, А.А. Степаненко¹, А.В. Фещенко¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

² Психологический институт Российской академии образования, г. Москва, Россия

МЕТОДЫ И ИНСТРУМЕНТЫ ВЫЯВЛЕНИЯ ПЕРСПЕКТИВНЫХ АБИТУРИЕНТОВ В СОЦИАЛЬНЫХ СЕТЯХ

Представлен опыт Томского государственного университета по определению целевой модели абитуриента на основе данных из социальной сети «ВКонтакте», разработке инструментов выгрузки и интеллектуального анализа данных и выявлению одаренных старшеклассников для поступления в вуз. Результаты апробации модели показали возможность применения результатов исследования в проектировании системного подхода по выявлению и удержанию талантливой молодежи в регионе.

Ключевые слова: социальные сети, абитуриенты, анализ данных, одаренность, машинное обучение, мягкие навыки.

За последние десятилетия существенно изменились требования к содержанию и результатам высшего образования. В запросе динамично меняющегося рынка труда устойчиво звучат требования работодателей к выпускнику университета: готовность эффективно действовать в ситуации неопределенности, формулировать и решать проблемы, работать в команде, постоянно доращивать свои компетенции до уровня новых задач. Это означает, что еще в вузе выпускник должен иметь мотивацию к постоянному саморазвитию, сформировать готовность быстро перестраивать как себя, так и свою профессиональную деятельность. В мировом сообществе изменения требований к выпускнику университета обсуждаются вокруг понятия *soft skills* (мягкие навыки). Исследования, проведенные в 16 европейских странах, показали, что 93 % работодателей считают *soft skills* таким же важным качеством сотрудника, как и его профессиональные навыки. По мнению Д. Гоулмана, эффективность современной профессиональной деятельности во многом определена именно «мягкими» навыками, освоение и применение которых отличает успешных специалистов от неуспешных, эффективные организации от неэффективных. Иными словами, именно «мягкие» навыки ведут к жестким результатам. Именно поэтому «мягкие» навыки сегодня являются большим приоритетом в корпоративном обучении, по сравнению с «жесткими» [1]. Д. Гоулманом и его коллегами были проанализированы данные о почти 500 моделях развития лидерских навыков, реализованных в ряде международных компаний (включая IBM, Lucent, PepsiCo, British Airways, Credit Swiss First Boston), а также в учреждениях

здравоохранения, учебных заведениях, органах государственной власти и др. В результате все значимые для работодателей навыки были сгруппированы в три категории: специальные навыки (то, что называется *hard skills*), когнитивные способности, составляющие эмоционального интеллекта, и *soft skills* [2].

Изменения образовательной парадигмы университетов со знаниевой на компетентностную также обусловили новые требования к выпускнику школы. Вуз ожидает не только качества информационно-знаниевой компоненты, но и хорошего уровня компетенций работы с информацией, опыта проектно-исследовательской деятельности, а также готовности к построению и реализации собственной образовательной траектории. Особенно актуализировались подобные ожидания у вузов, получивших статус национальных исследовательских. Разрабатывая программу повышения международной конкурентоспособности, Томский государственный университет выделил базовые характеристики целевой модели выпускника: развитая личность, способная действовать поверх профессиональных границ и создающая новую технологическую и социальную реальность в области повышения качества жизни человека в постиндустриальном обществе.

ТГУ выбрал стратегическим ориентиром создание системы выявления талантливых школьников, их раннего привлечения в университетскую среду для формирования качеств «своего» абитуриента, т.е. мотивированного и подготовленного к обучению по внутренним корпоративным стандартам университета. Это потребовало от университета, помимо успешной практики партнерства

с системой общего образования, разработки комплекса методов и инструментов, позволяющего выявлять перспективных старшеклассников в социальных сетях и проводить с ними системную работу по включению в среду университета. Такой подход позволяет университету выходить за границы своего территориального присутствия и активно искать и привлекать таланты на территориях соседних регионов без существенного увеличения финансовых издержек.

При разработке методов поиска перспективных абитуриентов в социальных сетях определено четыре направления исследовательских работ и создан набор инструментов для выгрузки данных из «ВКонтакте». Под инструментами мы понимаем набор программных и технических средств для сбора, очистки и хранения данных, а под методами – алгоритмы их анализа и теоретические основания интерпретации.

В рамках инициативы Томского государственного университета создан консорциум университетов для совместных исследований в области анализа больших данных. Консорциум запустил проект «Университет открытых данных» – портал по работе с открытыми данными (<http://data.tsu.ru>), который объединил специалистов в области компьютерных технологий и представителей других наук: социологов, политологов, филологов, психологов, биологов, генетиков. В личном кабинете пользователя портала доступны инструменты для выгрузки с помощью API пользовательских данных из «ВКонтакте»: данные о стенах пользователя (текстовый контент стены, статистика по текстовому контенту с использованием лемматизатора *mystem*), данные из профиля пользователя (более 30 параметров), данные для построения графа связей для одного пользователя и группы пользователей, выгрузка статистики

по половому, возрастному составу сообщества и географии его участников.

Первое направление исследования было связано с проверкой гипотезы о взаимосвязи интересов школьников к контенту в социальных сетях и выбором программы профессиональной подготовки. Частотный анализ подписок на тематические сообщества «ВКонтакте» позволил моделировать профиль интересов школьника в социальной сети по простой классификации: гуманитарные, точные и естественные науки [3]. Сравнение профиля интересов 18 тыс. абитуриентов Томской области с выбранным при поступлении факультетом в 2017 г. показало, что не всегда выбор профиля обучения проецируется на образовательные интересы в социальных сетях (таблица). У всех абитуриентов, выбравших гуманитарные факультеты в профиле интересов «ВКонтакте», преобладают тематические сообщества, связанные с гуманитарными предметами. Для большинства абитуриентов, поступивших на факультеты с подготовкой по точным и естественным наукам, не приоритетны интересы к соответствующим темам в социальной сети. Более того, у значительной части из них (точные науки – 24 %, естественные – 77 %) отсутствуют подписки на контент, тематически связанный с направлением подготовки.

Полученные в результате сравнения данные позволили эмпирическим путем определить критерии поиска потенциальных абитуриентов по каждому из трех направлений подготовки: для выявления потенциальных гуманитариев – более 90 % подписок на сообщества с соответствующей тематикой, для представителей точных наук – более 50 %, для естественных – более 10 %. Для достижения точности поиска необходимо выполнение еще нескольких условий: доля сообществ образовательной тематики в профиле не менее

Сравнение образовательных интересов пользователей «ВКонтакте» с выбранным направлением подготовки в вузе

Выбранное направление подготовки	Относительное количество пользователей «ВКонтакте» из контрольной группы абитуриентов, %			
	Доля сообществ, соответствующих направлению подготовки (67–100 %)	Доля сообществ, соответствующих направлению подготовки (34–66 %)	Доля сообществ, соответствующих направлению подготовки (1–33 %)	Доля сообществ, соответствующих направлению подготовки (0 %)
Гуманитарные науки	99	1	0	0
Точные науки	8	15	53	24
Естественные науки	4	2	17	77

5 % от общего количества подписок, не менее двух сообществ, соответствующих по теме предполагаемому направлению подготовки. Данная модель прогнозирования выбора направления подготовки абитуриента была проверена на большой выборке пользователей «ВКонтакте» (126 тыс. потенциальных абитуриентов СФО). Точность прогноза составила 0,82 – для гуманитариев, 0,76 – для точных наук, 0,69 – для естественных.

Второе направление исследования посвящено апробации автоматических методов анализа и классификации текстов на основе специальных тематических тезаурусов для проверки гипотезы, согласно которой тексты персональной страницы могут выражать научные и профессиональные интересы абитуриента в рамках противопоставления «гуманитарный – естественнонаучный – математический (точные науки) профиль». При выдвижении данной гипотезы и при определении способов и этапов ее проверки мы опирались на ряд предшествующих социологических исследований интересов пользователей социальных сетей [4–7], на лингвистические работы, посвященные анализу дискурсивно-жанрового своеобразия анализируемых текстов и лингвистическому моделированию языковой дискурсивной личности [8]. Кроме того, был использован опыт применения междисциплинарного использования методов лингвистического анализа текста, автоматической обработки значительных текстовых массивов

и статистической проверки лингвистических гипотез [9, 10].

Для классификации текстов по принадлежности к гуманитарным, точным и естественным наукам нами использовались следующие виды классификаторов: линейный дискриминантный анализ (LDA), метод опорных векторов (SVM), логистическая регрессия (LR), деревья решений (Trees), случайный лес (RF). Наиболее успешным оказался классификатор – LR (логистическая регрессия), показывающий стабильный результат распределения стен пользователей с точностью 0,64. На рис. 1 представлены результаты использования данного классификатора.

График классификации показывает, что все тексты разделяются на три класса: гуманитарный, естественный и математический, что свидетельствует о корректности работы составленных нами тезаурусов. Однако интерес вызывают тексты, относящиеся к гуманитарному научному направлению. Полагаем, что это вызвано большим количеством лексических единиц в гуманитарном тезаурусе, которые также встречаются в постах (или стенах) пользователей социальных сетей двух других научных направлений. Подобный результат требует дальнейшего улучшения за счет более детального изучения атрибутов тезауруса (в частности, гуманитарного направления), поиска оптимальных моделей классификации и применения методов мультиколлинережности в нашем исследовании.

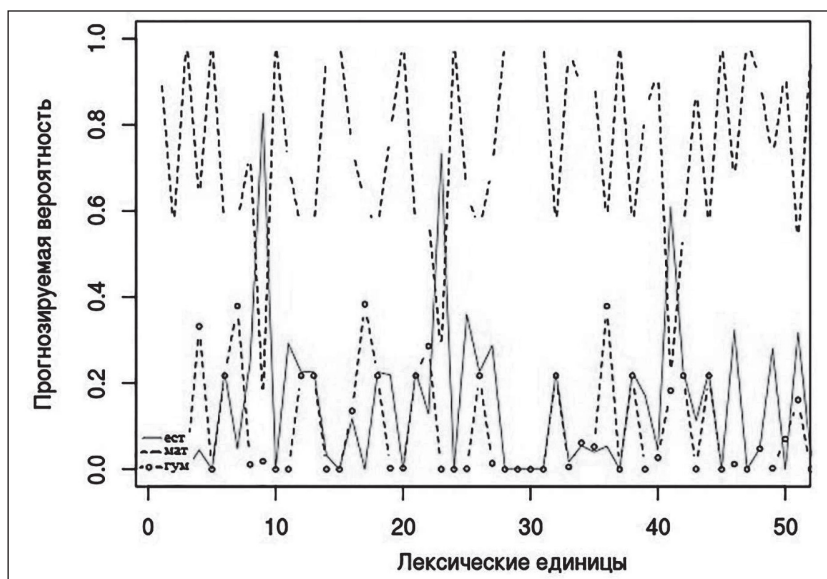


Рис. 1. Прогнозируемая вероятность лексических единиц по научным направлениям на основе логистической регрессии

Проведенный анализ свидетельствует, во-первых, о результативности автоматической классификации текстов тематически свободного общения в социальной сети «ВКонтакте» с использованием в качестве атрибутов выделенных ключевых слов текстов групп профессионально ориентированного общения, во-вторых, о меньшей степени релевантности в рассматриваемом аспекте данных гуманитарного блока профессионализации. В последнем случае необходимо работать в направлении как уточнения состава лингвистических маркеров, так и поисков более точных систем автоматической обработки текста. Однако уже на этом уровне можно говорить о возможности применения данной модели классификации к текстам других групп пользователей социальной сети «ВКонтакте», что предполагается осуществить на следующем этапе работы.

Методы классификации интересов старшеклассников через анализ их подписок и текстов на личной странице «ВКонтакте» могут быть объединены в рамках одной прогностической модели, дополняя и уточняя друг друга.

Третье направление исследования посвящено изучению потенциала социальных сетей для выявления важных для диагностики одаренности психологических качеств, интересов и профессиональных увлечений учащихся старших классов с помощью анализа данных тестирования и данных, представленных в профилях в социальной сети «ВКонтакте». Методологическими основаниями исследования явились современные научные представления об одаренности, представленные в работах отечественных и зарубежных ученых: концепции одаренности Ю.Д. Бабаевой, Д.Б. Богоявленской, Н.С. Лейтеса, А.М. Матюшкина, А.А. Мелик-Пашаева, Д.В. Ушакова, М.А. Холодной, В.Д. Шадрикова, Н.Б. Шумаковой, Е.И. Щеплановой, В.С. Юркевич и др., многомерные модели одаренности Дж. Рензулли, К. Хеллера и др.

В соответствии с указанными представлениями одаренность принято рассматривать как взаимодействие интеллекта, креативности и мотивационно-личностных особенностей. Одаренными считаются люди, обладающие этими качествами или способные их развить и использовать в любой потенциально ценной деятельности [11].

Общую выборку исследования составили 1692 человека – 969 девушек и 723 юноши – учащиеся старших классов средних общеобразовательных

учебных заведений г. Томска. Методы сбора данных: психологическое тестирование и анкетирование, анализ академической успеваемости, «ручной» анализ структуры и содержания профилей в социальной сети.

Метод психологического тестирования и анкетирования был реализован с помощью компьютеризированной методики «#Профорентация», предназначенной для учащихся старших классов образовательных учреждений разного вида [12]. Кроме основной цели определения общей направленности профессиональной ориентации, с помощью данной методики могут решаться задачи определения интеллектуального потенциала испытуемых, изучения их личностных свойств и качеств (эмоциональных, волевых, коммуникативных), профессиональных интересов и ценностей, а также карьерных ориентаций.

На основании результатов «ручного» анализа структуры и содержания профилей испытуемых в социальной сети можно говорить о некоторых психологических качествах старшеклассников, которые, однако, не могут рассматриваться в качестве значимых признаков одаренности. Для сравнения в исследовании были изучены структура и содержание профилей у испытуемых, не имеющих высоких значений по важным для определения одаренности субтестам ($n=240$, 140 девушек и 100 юношей), – значимых различий между ними и одаренными испытуемыми обнаружено не было.

Можно предположить, что признаки одаренности трудноуловимы в «цифровом следе» в социальной сети по ряду причин:

– специфика самой социальной сети «ВКонтакте», в которой доминирует контекст развлекательной деятельности, в то время как контексты для реализации других так называемых «полезных» видов деятельности практически отсутствуют. Если в Facebook пользователи приходят общаться на серьезные темы, обсудить бизнес-проблемы, то в «ВКонтакте» – поболтать с друзьями и расслабиться: послушать музыку, посмотреть видео, поиграть в игры. Возможно, поэтому она так популярна среди студентов и школьников (средний возраст ее пользователей от 13 до 34 лет);

– специфика содержания социальной сети «ВКонтакте», в которой доминирует графический контент, а не текстовый, – методы психолингвистики в этом случае не применимы, поскольку

вместо авторского текста публикуются изображения;

– слишком большое количество признаков для проведения «ручного» анализа. Например, у одного пользователя может быть только 100 подписок на группы и интересные страницы, и если анализировать еще и его тексты, связи, реакции и пр., таких признаков у одного пользователя может «накопиться» до нескольких сотен.

Поэтому для выявления одаренных испытуемых в социальной сети необходимо применять методы, позволяющие проводить масштабные исследования как по количеству испытуемых, так и по количеству необходимых для анализа признаков. Поэтому «ручной» анализ структуры и содержания профилей испытуемых в социальной сети был дополнен методом машинного обучения.

В исследовании J. W. Pennebaker [13] было показано, что именно методы машинного обучения помогают понять взаимосвязь между данными о человеке и психологическими качествами. Это исследование было посвящено использованию метода опорных векторов для прогнозирования личностных качеств опросника «Большая пятерка» по авторским текстам. Лучшим результатом стала точность 58 % для шкалы «Нейротизм». В другом исследовании взаимосвязи текстов блогеров и личностных качеств опросника «Большая пятерка» [14], выполненном S. Nowson, J. Oberlander, A. J. Gill, показана значимая корреляция между характеристиками текста и результатами по шкалам «Открытость», «Доброжелательность». Широкую известность получили работы M. Kosinski и D. Stillwell при участии специалиста по машинному обучению T. Graepel. В рамках реализованного ими проекта Mypersonality были собраны данные социальных сетей и результаты заполнения различных психодиагностических опросников более чем семи миллионов пользователей [15].

Аналогично перечисленным работам в представляемом исследовании задача, решаемая моделями машинного обучения, заключалась в бинарной классификации испытуемых ($n = 1\ 640$). К отрицательному классу относятся испытуемые, показавшие результаты ниже среднего для данного психологического качества, к положительному классу – испытуемые с результатами выше среднего по данному психологическому качеству. Для определения среднего использовались нормы, полученные при тестировании более шести тысяч

старшекласников. Таким образом, с применением метода машинного обучения сравнивались одаренные и «неодаренные» старшекласники.

Для решения задачи бинарной классификации использовались следующие модели: метод опорных векторов, случайные леса и градиентный бустинг. Сравнение моделей путем расчета площади под ROC-кривой показало, что наиболее эффективной для решения задачи бинарной классификации является модель на основе метода опорных векторов. Точность прогностической модели, основанной на этом методе, для выявления старшекласников с высоким уровнем когнитивных и психологических качеств составила: интеллект 0,7; креативность 0,7; мотивационно-личностные особенности 0,72. Таким образом, применение метода машинного обучения позволяет со средней точностью 0,7 идентифицировать испытуемых с высоким уровнем развития когнитивных и психологических качеств, выявлять корреляции между испытуемыми с различной степенью выраженности признаков одаренности и сообществами социальной сети.

В рамках четвертого направления исследования проверено предположение о наличии особенностей в структуре социальных связей у одаренных старшекласников с помощью метода анализа социальных сетей (social network analysis). Был проанализирован социальный граф для школьников с высокими и низкими результатами теста на определение уровня развития интеллекта. Метод не проверялся по креативным и мотивационно-личностным характеристикам. В результате не удалось выявить существенных различий в количестве друзей у школьников с результатами теста на определение уровня развития интеллекта выше и ниже среднего значения: среднее количество связей у 148 и 151 соответственно. Метрики социального графа, построенного для каждой из двух групп, также не показали существенных различий. Визуализация сети связей школьников-интеллектуалов (рис. 2) показала наличие единой сети виртуальной связи, несмотря на то, что в ней представлены учащиеся из 54 образовательных учреждений нескольких населенных пунктов Томской области. Только 15 % школьников оказались никак не связанными с общей сетью знакомств. Природу формирования такой сети еще предстоит описать и исследовать. Это позволит выявить причины и условия «зна-

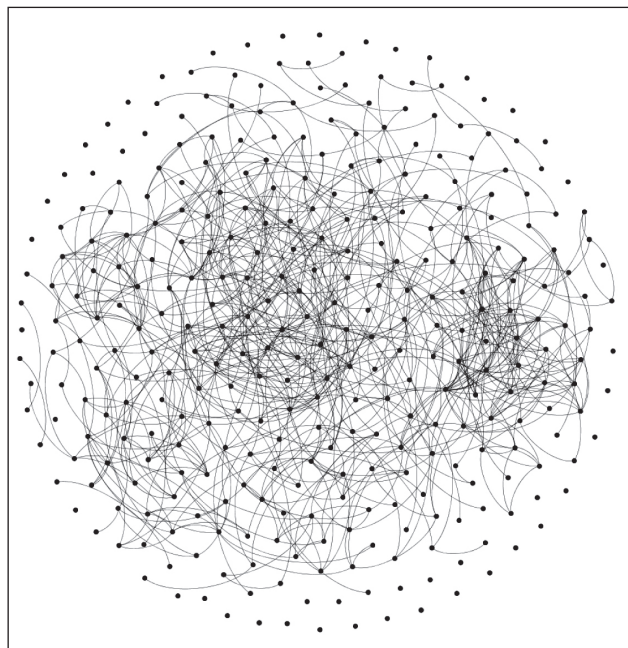


Рис. 2. Граф социальных связей школьников с результатами теста на определение уровня развития интеллекта выше среднего значения

комства» и поддержания связей между абитуриентами с признаками одаренности и использовать эти данные для разработки механизмов привлечения талантливой молодежи в вуз. Также анализ социальной сети школьников-интеллектуалов позволил выявить лидеров мнения – школьников с наибольшим количеством знакомств внутри сети, что позволит повысить эффективность коммуникации вуза с целевым абитуриентом.

Результаты всех четырех направлений представленного исследования показывают, что методика предсказания образовательных интересов и признаков одаренности по подпискам пользователей дала наилучшую результативность. На текущем этапе исследования на их основе возможно конструирование прогностической модели по выявлению перспективных абитуриентов через проекцию целевой модели выпускника ТГУ¹. В состав такой модели должны войти следующие критерии:

– высокий уровень интеллекта и увлеченности (заинтересованности) предметом как проекция качества «развитая личность» в целевой модели выпускника ТГУ;

– высокий уровень креативности – проекция «создания новой технологической и социальной реальности».

Оценка потенциала абитуриента по двум критериям этой модели позволит университетам выявлять старшеклассников, обладающих полным или частичным набором желаемых качеств, и организовывать их раннее привлечение в университетскую среду для формирования интереса к образовательным программам вуза и подготовки к поступлению и обучению.

Исследования, связанные с определением профессиональных интересов абитуриентов по их текстам на личной странице, и анализ сетевых структур абитуриентов также представляются перспективными и могут быть включены в модель поиска после их дополнительной доработки.

Перспективы исследования связаны с возможностями методов машинного обучения при проведении масштабного, оперативного, лонгитюдного и менее затратного по сравнению со стандартным бланковым тестирования (или серии тестирований) старшеклассников для определения их одаренности, психологических качеств и отслеживания динамики их развития. Дополнительно к определению одаренности разработанная модель прогнозирования может применяться для определения общекультурных компетенций (soft skills), значимых для успешной профессиональной реализации: лидерство, предпринимательство, толерантность и др.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-16-70004.

ЛИТЕРАТУРА

1. Гоулдман Д. Эмоциональный интеллект / Д. Гоулман. – М.: АСТ, 2009. – 480 с.
2. Гоулман Д., Бояцис Р., Макки Э. Эмоциональное лидерство. – М.: Альпина Бизнес Букс, 2008.
3. Можалева Г.В., Слободская А.В., Феценко А.В. Информационный потенциал социальных сетей для выявления образовательных потребностей школьников // Открытое и дистанционное образование. – 2017. – № 3(67). – С. 25–30. DOI: 10.17223/16095944/67/4.
4. Коршунов А.В. Задачи и методы определения атрибутов пользователей социальных сетей // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. – 2013. – С. 380–390.
5. Kim J. et al. Extracting User Interests on Facebook // International Journal of Distributed Sensor Networks. – 2014. – Vol. 10, № 6. – P. 1–5.

¹ Развитая личность, способная действовать поверх профессиональных границ и создающая новую технологическую и социальную реальность в области повышения качества жизни человека в постиндустриальном обществе.

6. *Ahmed A. et al.* Scalable distributed inference of dynamic user interests for behavioral targeting // Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2011. – P. 114–122.

7. *Michelson M., Macskassy S.A.* Discovering users' topics of interest on twitter: a first look // Proceedings of the fourth workshop on Analytics for noisy unstructured text data. – ACM, 2010. – P. 73–80.

8. *Резанова З.И., Скрипко Ю.К.* Личность в среде дискурса: языковая репрезентация социально-психологических типов (на материале дискурса виртуальных фан-сообществ музыкальной направленности) // Вестник Том. гос. ун-та. Филология. – 2016. – № 3 (41). DOI: 10.17223/19986645/41/4.

9. *Резанова З.И., Романов А.С., Мещеряков Р.В.* Задачи авторской атрибуции текста в аспекте гендерной принадлежности (к проблеме междисциплинарного взаимодействия лингвистики и информатики) // Вестник Том. гос. ун-та. – 2013. – № 370. – С. 24–28.

10. *Резанова З.И., Романов А.С., Мещеряков Р.В.* О выборе признаков текста, релевантных в автороведческой экспертной деятельности // Вестник Том. гос. ун-та. Филология. – 2013. – № 6 (26). – С. 38–52.

11. *Щебланова Е.И.* Одаренность как психологическая система: структура и динамика в школьном возрасте: дис. ... д-ра психол. наук / Е.И. Щебланова. – М., 2006. – 311 с.

12. *Богдановская И.М.* Компьютерная психодиагностика в профориентационной работе со старшеклассниками / И.М. Богдановская, П.Б. Киселев, А.Н. Кошелева, В.А. Рубан // Психологические проблемы образования и воспитания в современной России: матер. IV конференции психологов образования Сибири. – Иркутск, 2016. – С. 296–302.

13. *Pennebaker J.W.* Linguistic styles: language use as an individual difference / J.W. Pennebaker, L.A. King // Journal of personality and social psychology. – 1999. – Vol. 77, № 6. – P. 1296.

14. *Nowson S.* Weblogs, genres and individual differences / S. Nowson, J. Oberlander, A.J. Gill // Proceedings of the 27th Annual Conference of the Cognitive Science Society. – 2005. – Vol. 1666. – P. 1671.

15. *myPersonality Project* [Электронный ресурс]. – URL: http://mypersonality.org/wiki/doku.php#mypersonality_in_numbers (дата обращения: 04.12.2017).

Gojko V.L.¹, Kiselev P.B.², Matsuta V.V.¹,
Sukhanova E.A.¹, Stepanenko A.A.¹,
Feshchenko A.V.¹

¹ National research Tomsk state university,
Tomsk, Russia

² Psychological Institute, Russian Academy of
Education, Moscow, Russia

METHODS AND TOOLS TO IDENTIFY
PROMISING ENTRANTS IN SOCIAL
NETWORKS

Keywords: social networking, entrants, data analysis, talent, machine learning, soft skills.

Tomsk state University created the Consortium of universities for joint research in the field of

big data Analytics. The consortium launched the project “University of open data” (<http://data.tsu.ru>), which brought together specialists in the field of computer technology and other Sciences – sociologists, political scientists, philologists, psychologists, biologists, geneticists. Tools available for uploading with the API custom data from “Vkontakte” for user of the project . Also we used the program “1C:School Psychodiagnostics” as a tool for collecting, cleaning, structuring and storing data in the study.

The first area of research was related to testing of the hypothesis on the relationship between the interests of the entrants in social networks and choice of training programmes. We developed a model predicting choice of the direction of preparation of the applicant was tested on a large sample of the users “Vkontakte” (126 000). The accuracy of prediction equaled to 0.82 for the humanities, of 0.76 for the mathematical, of 0.69 for the natural sciences.

The second direction of studies is connected with the automatic validation of methods of analysis and text classification to determine the scientific and professional interests of the applicant. The results of the study confirmed the possibility of the classification of texts in the HUMANITIES, natural and mathematical science with precision of 0.64.

The third direction of research is devoted to studying the potential of social networks to identify key talent to diagnose psychological qualities of entrants through analysis of their subscriptions in a social network. To solve the problem of binary classification used support vector machines. The accuracy of a predictive model based on this method were as follows: 0.7 for identification of children with high intelligence; 0.7 for creative; 0.72 for children with high personal motivation.

In the fourth direction was analyzed the social graph (SNA) for entrants with high and low test results to determine the level of intellectual development. The result failed to reveal significant differences in the number of friends pupils with low and high level of intelligence. Visualization of the network of intellectuals has shown the existence of a single network virtual connections, despite the fact that it includes entrants from 54 educational institutions of several cities in Tomsk region. Also the analysis of social networks of intellectuals helped to identify opinion leaders - entrants with

the greatest number of people in your network that will improve the effectiveness of communication of the University with the task applicant.

At the current stage of research it is possible to design a predictive model to identify promising entrants through the projection of the target model of the graduate of TSU. The composition of the model should include the following criteria:

- a high level of intelligence and interest to science as the projection quality of a «developed personality» of the target model of a graduate of TSU;

- high level of creativity as the projection «create new technological and social reality.»

Assessment of entrant's potential according to two criteria of this model will allow universities to identify high school entrants with a full or partial set of desired qualities, and to organize their early engagement in the University environment for the formation of interest to the educational program of the University and prepare for entry and training.

REFERENCES

1. *Gouldman D.* Jemocional'nyj intellekt / D. Gouldman. – M.: AST, 2009. – 480 s.
2. *Goulman D., Bojaxis R., Makki Je.* Jemocional'noe liderstvo. – M.: Al'pina Biznes Buks, 2008.
3. *Mozhaeva G.V., Slobodskaja A.V., Feshhenko A.V.* Informacionnyj potencial social'nyh setej dlja vyjavlenija obrazovatel'nyh potrebnostej shkol'nikov // Otkrytoe i distancionnoe obrazovanie. – 2017. – № 3(67). – S. 25–30. DOI: 10.17223/16095944/67/4.
4. *Korshunov A.V.* Zadachi i metody opredelenija atributov pol'zovatelej social'nyh setej // Jelektronnye biblioteki: perspektivnye metody i tehnologii, jelektronnye kollekcii. – 2013. – S. 380–390.
5. *Kim J. et al.* Extracting User Interests on Facebook // International Journal of Distributed Sensor Networks. – 2014. – Vol. 10, № 6. – P. 1–5.
6. *Ahmed A. et al.* Scalable distributed inference of dynamic user interests for behavioral targeting // Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2011. – P. 114–122.
7. *Michelson M., Macskassy S.A.* Discovering users' topics of interest on twitter: a first look // Proceedings of the fourth workshop on Analytics for noisy unstructured text data. – ACM, 2010. – P. 73–80.
8. *Rezanova Z.I., Skripko Ju.K.* Lichnost' v srede diskursa: jazykovaja reprezentacija social'no-psihologicheskikh tipov (na materiale diskursa virtual'nyh fan-soobshhestv muzykal'noj napravlenosti) // Vestnik Tom. gos. un-ta. Filologija. – 2016. – № 3 (41). DOI: 10.17223/19986645/41/4.
9. *Rezanova Z.I., Romanov A.S., Meshherjakov R.V.* Zadachi avtorskoj atribucii teksta v aspekte gendernoj prinadlezhnosti (k probleme mezhdisciplinarnogo vzaimodejstvija lingvistiki i informatiki) // Vestnik Tom. gos. un-ta. – 2013. – № 370. – S. 24–28.
10. *Rezanova Z.I., Romanov A.S., Meshherjakov R.V.* O vybore priznakov teksta, relevantnyh v avtorovedcheskoj jekspertnoj dejatel'nosti // Vestnik Tom. gos. un-ta. Filologija. – 2013. – № 6 (26). – S. 38–52.
11. *Shheblanova E.I.* Odarennost' kak psihologicheskaja sistema: struktura i dinamika v shkol'nom vozraste: dis. ... d-ra psihol. nauk / E.I. Shheblanova. – M., 2006. – 311 s.
12. *Bogdanovskaja I.M.* Komp'juternaja psihodiagnostika v proforientacionnoj rabote so starsheklassnikami / I.M. Bogdanovskaja, P.B. Kiselev, A.N. Kosheleva, V.A. Ruban // Psihologicheskie problemy obrazovaniya i vospitaniya v sovremennoj Rossii: mater. IV konferencii psihologov obrazovaniya Sibiri. – Irkutsk, 2016. – S. 296–302.
13. *Pennebaker J.W.* Linguistic styles: language use as an individual difference / J.W. Pennebaker, L.A. King // Journal of personality and social psychology. – 1999. – Vol. 77, № 6. – P. 1296.
14. *Nowson S.* Weblogs, genres and individual differences / S. Nowson, J. Oberlander, A.J. Gill // Proceedings of the 27th Annual Conference of the Cognitive Science Society. – 2005. – Vol. 1666. – P. 1671.
15. *myPersonality Project* [Jelektronnyj resurs]. – URL: http://mypersonality.org/wiki/doku.php#mypersonality_in_numbers (data obrashhenija: 04.12.2017).

Открытое и дистанционное образование

Научно-методический журнал
№ 4(68) 2017 г.

Редактор
В.Г. Лихачева

Компьютерная верстка
В.Б. Малиновский

Подписано в печать 21.12.2017 г. Формат 84x108¹/₁₆.
Бумага офсетная № 1. Печать офсетная. П. л. 5,0. Усл. п. л. 7,0. Уч.-изд. л. 7,4.
Тираж 500 экз. Заказ 389.

ООО «Издательство ТГУ», 634029, г. Томск, ул. Никитина, 4.
ООО «Новые Печатные Технологии», 634040, г. Томск, ул. Высоцкого, 28, стр. 1