

ДИАГНОСТИКА ПРОФЕССИОНАЛЬНОЙ КОМПЕТЕНТНОСТИ СТУДЕНТОВ ИТ-НАПРАВЛЕНИЙ НА ОСНОВЕ ДАННЫХ ЦИФРОВОГО СЛЕДА

И. Г. Захарова^{1,2}, Ю. В. Боганюк^{1,2}, М. С. Воробьева^{1,2}, Е. А. Павлова^{1,2}

¹ Тюменский государственный университет
625003, Россия, г. Тюмень, ул. Володарского, д. 6

² Научно-технический университет «Сириус»
354340, Россия, г. Сочи, Олимпийский проспект, д. 1

Аннотация

В статье предложен подход к диагностике уровня профессиональной компетентности выпускников ИТ-направлений, основанный на анализе данных цифрового следа студента и содержания соответствующей образовательной программы. Описаны способы извлечения показателей профессионального уровня студента из данных цифрового следа — текстов рабочих программ по освоённым дисциплинам и выпускных квалификационных работ. Показана методика соотнесения этих показателей с формализованными требованиями работодателей, отраженными в текстах вакансий в области информационных технологий. Приведен пример реализации предложенного подхода в Институте математики и компьютерных наук Тюменского государственного университета. Диагностика выполнялась на основании набора данных, включавшего тексты рабочих программ для ИТ-направлений бакалавриата и специалитета, 542 выпускные квалификационные работы по этим направлениям, 879 описаний требований к вакансиям и информации о трудоустройстве выпускников. Представленный подход позволяет оценить актуальность образовательной программы в целом и уровень профессиональной подготовки каждого студента на основе объективных данных. Полученные результаты были использованы для актуализации содержания базовых курсов и включения в учебный план новых элективных дисциплин.

Ключевые слова: цифровой след, извлечение данных, высшее образование, профессиональная компетентность, рынок труда, образовательный результат, анализ текстов.

DOI: 10.32517/0234-0453-2020-35-4-4-11

Для цитирования:

Захарова И. Г., Боганюк Ю. В., Воробьева М. С., Павлова Е. А. Диагностика профессиональной компетентности студентов ИТ-направлений на основе данных цифрового следа // Информатика и образование. 2020. № 4. С. 4–11.

Статья поступила в редакцию: 16 февраля 2020 года.

Статья принята к печати: 19 мая 2020 года.

Финансирование

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028.

Сведения об авторах

Захарова Ирина Гелиевна, доктор пед. наук, профессор, профессор кафедры программного обеспечения, Институт математики и компьютерных наук, Тюменский государственный университет, Россия; i.g.zakharova@utmn.ru; ORCID: 0000-0002-4211-7675

Боганюк Юлия Викторовна, аспирант кафедры программного обеспечения, Институт математики и компьютерных наук, Тюменский государственный университет, Россия; y.v.boganyuk@utmn.ru; ORCID: 0000-0002-3663-0173

Воробьева Марина Сергеевна, канд. тех. наук, доцент, зав. кафедрой программного обеспечения, Институт математики и компьютерных наук, Тюменский государственный университет, Россия; m.s.vorobyova@utmn.ru; ORCID: 0000-0002-1508-4089

Павлова Елена Александровна, доцент кафедры программного обеспечения, Институт математики и компьютерных наук, Тюменский государственный университет; e.a.pavlova@utmn.ru; ORCID: 0000-0002-0408-9888

1. Введение

Диагностика профессиональной подготовки студентов ИТ-направлений приобретает особую актуальность в связи с высокой потребностью в кадрах для цифровой экономики. Динамичность современного рынка труда определяет необходимость своевременной адаптации и даже коренного пересмотра структуры образовательных программ и содержания учебных дисциплин. И в качестве обратной связи здесь выступает ключевой показатель — соответствие подготовки выпускника требованиям работодателей. При этом необходимо учитывать тот факт, что на профессиональное развитие студента значительное влияние оказывает не только то, как организована образовательная деятельность в стенах

вуза — в области информационных технологий очень распространены онлайн-курсы различного объема и уровня сложности. При этом именно по причине востребованности ИТ-специалистов многие студенты уже на младших курсах совмещают учебу и работу.

Все эти факторы (формальное образование и неформальное самообразование, практическая деятельность в процессе обучения в вузе), безусловно, влияют на развитие профессиональной подготовки выпускника, и зачастую не вполне очевидно, что же является главным с точки зрения соответствия уровня подготовки актуальным требованиям работодателей. В то же время сопоставление содержания образовательных программ, образовательных результатов и запрашиваемых работодателями компетенций могло бы прояснить эту картину. Решение такой

многосторонней задачи требует соответствующей комплексной диагностики содержания образовательного процесса, его результатов с точки зрения актуальных целевых показателей. Традиционное для проведения процедуры аккредитации образовательных программ беглое (как правило) изучение программ дисциплин, содержания выпускных квалификационных работ и отчетов о трудоустройстве выпускников сложно считать способом объективной оценки сформированности профессиональной компетентности. При этом здесь сложно вычленишь и то, изучение каких учебных дисциплин, общение с какими преподавателями оказали свое влияние на профессиональное развитие студента, и то, какую роль сыграли самообразование и практическая деятельность.

Различные стороны образовательной деятельности студента, в том числе развитие его профессиональной компетентности, можно проанализировать, опираясь на данные цифрового следа. *Под цифровым следом студента мы понимаем постоянно пополняемый набор данных, включающий как значения традиционных показателей (текущая, промежуточная и итоговая аттестация, посещение занятий и своевременность сдачи различных заданий и т. д.), так и тексты, созданные самими студентами.* В их числе рефераты и обзоры литературы, курсовые работы по дисциплинам и по направлению, отчеты по практикам различных видов, описания учебных и научных проектов, выпускные квалификационные работы, эссе и мотивационные письма на конкурсы. Именно из этих текстов (по сути, контента расширенного электронного портфолио) современные методы анализа данных позволяют извлечь объективную информацию для диагностики профессиональной компетентности выпускника и выявления факторов, которые повлияли на ее формирование [1].

Цель статьи — показать возможность диагностики уровня профессиональной компетентности выпускников ИТ-направлений бакалавриата, сопоставляя отдельные данные цифрового следа студента, содержание соответствующей образовательной программы и актуальные запросы рынка труда.

2. Обзор литературы

Многообразие данных, имеющих прямое или косвенное отношение к образовательному процессу вуза, меняет традиционные представления о диагностике качества образования. Эта тенденция находит свое отражение в работах отечественных и зарубежных исследователей.

Эффективность современных технологий сбора и интеллектуального анализа данных определяет возможность развития и внедрения новых подходов к управлению образовательным процессом, основанному на всесторонней и объективной информации [2, 3]. Привлечение новых источников данных крайне важно не только для оценки достижений студента, но и для выявления факторов, влияющих на эти до-

стижения, для объективного анализа актуальности и качества образовательных программ.

В контексте оценки профессионального развития акцент закономерно делается на использовании данных портфолио студента [4]. При этом в условиях цифровизации образования особое значение для управления образовательным процессом приобретает возможность создания электронного портфолио. Важно, что исследователи не только рассматривают структуру электронного портфолио, но и обращают внимание на соотнесение его материалов с требованиями соответствующих профессиональных стандартов [5], подчеркивают важность электронного портфолио как для профессионального развития студента, так и для обратной связи с рынком труда [6–10].

Развитие методов интеллектуального анализа текстов нашло свое отражение и в том, что наряду с обобщенными показателями достижений студентов исследователи все чаще обращаются к конкретным текстам. Например, в [11–13] показано, что тексты постов студентов в социальной сети «ВКонтакте» могут служить основой для выявления их профессиональных предпочтений. При этом подчеркивается (см., например, [14, 15]), что результаты такого анализа напрямую связаны с выбором определенных признаков, которые, в частности, позволяют учесть особенности направления подготовки (ИТ, естественно-научные и гуманитарные области и др.).

Следующим этапом (после определения значимых для решаемой задачи признаков текста) является применение современных методов машинного обучения. Эти методы позволяют не только обеспечивать автоматическую оценку письменной работы (реферата, эссе) [16–18], но и выявлять по выделенным признакам особенности личности, например, креативность обучающихся [19–21], саморефлексию [22]. Важно, что, несмотря на сложность анализа текстов на русском языке, на примере ИТ-образования показана результативность методов машинного обучения для анализа текстов профессиональных стандартов и работ студентов с целью формирования содержания учебных дисциплин [23, 24].

В большей степени тексты работ студентов в качестве источника данных рассматриваются в зарубежных исследованиях, о чем свидетельствует количество соответствующих публикаций. Это можно объяснить тем, что для отечественных исследователей образовательного процесса задачи анализа текстов существенно сложнее — и из-за особенностей русского языка, и по причине меньшей регламентированности структуры и содержания работ студентов по сравнению с требованиями в зарубежных университетах.

3. Методология, материалы и методы исследования

Предлагаемый подход основан на предположении о том, что можно соотнести уровень профессиональной компетентности студента с требованиями рынка труда на основании формализации соответствующих

критериев и сопоставления атрибутов двух моделей: «Цифровой след студента» (ЦСС) и «Актуальные профессиональные компетенции» (АПК). Для проверки результативности этого подхода мы ограничились минимальными наборами атрибутов для этих моделей. Характеристики степени владения актуальными для работодателей ключевыми навыками (атрибуты ЦСС) извлекались из текстов работ студентов и содержания успешно освоенных учебных курсов, а требования рынка труда (атрибуты АПК) — из текстов вакансий в области ИТ.

Основным материалом для анализа послужил корпус документов, включающий тексты выпускных квалификационных работ (ВКР) студентов ИТ-направлений бакалавриата и специалитета Института математики и компьютерных наук Тюменского государственного университета. **Всего было проанализировано 542 текста ВКР, защищенных в 2016–2018 годах.**

Каждый текст ВКР был разделен на две основные части, используемые в анализе: 1) непосредственно основная часть и 2) приложение, содержащее исходный программный код. Также в качестве материала для анализа использовались справки о внедрении программных продуктов студентов на предприятии. Такие справки были приложены к 174 (32,1 %) ВКР.

В качестве источника, определяющего актуальные требования рынка труда для ИТ-специалистов, использовался набор текстов вакансий, размещенных в открытом доступе на сайте HeadHunter (<https://hh.ru>) по специализации «Информационные технологии, интернет, телеком» в городах Тюмень, Екатеринбург, Новосибирск. Был выполнен предварительный анализ этих данных для отбора вакансий, которые соответствуют направлениям подготовки студентов в Институте математики и компьютерных наук ТюмГУ: веб-разработчик, разработчики на определенном языке программирования (C#, C++, Java, Python и др.), 1С-программист, системный администратор, специалист по информационной безопасности и разработчик баз данных.

В итоге **в набор вакансий было включено 879 текстов.** В рамках предложенного подхода для определения актуальных запросов рынка труда к специалистам (модель АПК) из текстов вакансий необходимо извлечь названия требуемых технологий и уровень владения ими. Под технологиями здесь обобщенно понимаются названия операционных систем, языков программирования, библиотек и т. п.

Например, для текста вакансии «Программист C# (Middle)» со списком требований: «Уверенное знание платформы Microsoft .NET (Framework 4.6 и выше), языка C#; понимание шаблона MVVM, опыт работы с WPF; понимание DI, REST API; опыт работы с Git» нужно извлечь следующие значения (слова и словосочетания) параметров модели АПК:

- требуемые технологии: .Net, C#, MVVM, WPF, DI, REST API, Git;
- требуемый уровень: Middle, уверенное знание, понимание, опыт работы.

Для соотнесения с моделью ЦСС по результатам анализа текстов вакансий было выделено **три уровня владения технологиями** (по сути, навыками работы с ними), ключевыми для работодателя:

- 1) имеет представление;
- 2) имеет опыт;
- 3) уверенно владеет.

В таблице 1 приведен принцип сопоставления требуемого уровня с извлекаемыми характеристиками вакансий.

Таблица 1

Описание уровней владения на основе ключевых навыков в текстах вакансий

№ п/п	Уровень владения навыками	Описания в текстах вакансий
1	Имеет представление	Junior, базовое знание, знание, понимание, имеет(ся) представление, начальные знания, готовность работать
2	Имеет опыт	Middle, уверенное знание, опыт работы, глубокие знания, хорошие знания
3	Уверенно владеет	Middle, Senior, владение, уверенное знание, уверенное владение, большой опыт работы, участие в проектах, участие в проектировании, участие в разработке, проектирование, внедрение, опыт коммерческой разработки

В рамках предлагаемого подхода для определения уровня «имеет представление» целесообразно дополнительно привлечь тексты рабочих программ по дисциплинам. Мы предположили, что этот уровень в модели ЦСС обеспечивается упоминанием требуемых для данной вакансии технологий в тексте рабочей программы одной или нескольких дисциплин, успешно освоенных студентом (без учета соответствующих оценок).

Однако для выявления опыта работы с требуемыми технологиями и тем более для выявления уверенного владения ими необходим более глубокий анализ, материалом для которого и послужили тексты ВКР. Достижение уровня «имеет опыт» определялось по наличию названия соответствующей технологии в описании практической реализации ВКР. В данном случае важно отметить, что упоминание в обзоре или списке литературы не являлось достаточным показателем для данного уровня.

Внедрение программных продуктов (по меньшей мере, в опытную эксплуатацию) в модели ЦСС обеспечивало уровень «уверенно владеет», так как предполагается, что реализация проекта по заказу предприятия требует высокого уровня профессиональных навыков. О внедрении свидетельствовало наличие соответствующей справки с предприятия и упоминание об этом в заключительной части («Заключение», «Выводы», «Основные результаты») текста ВКР.

Для области разработки программных продуктов особо важны **дополнительные атрибуты модели ЦСС, которые отражают качество программного кода, представленного в ВКР**. В предположении функциональной пригодности [25] кода для решения задачи, сформулированной в ВКР, мы также оценивали качество документирования, структурирования и связи с основным текстом работы. Последний показатель отражает наличие в теоретической части (при описании методов и алгоритмов решения задачи) ссылок на представленные в практической части работы отдельные фрагменты кода, который реализует эти методы и алгоритмы. Оценка выполнялась автоматически на основе анализа текстов комментариев в программе и самого кода (разбивка на модули, ограничения на количество операторов в методах или функциях и т. п.). Использовалась трехбалльная шкала (0, 1, 2). Для уровня «имеет представление» этот показатель не учитывался, а для уровней «имеет опыт» и «уверенно владеет» требовалось получить минимум 1 и 2 балла соответственно.

Предварительная обработка текстов ВКР, справок о внедрении, рабочих программ и вакансий включала удаление стоп-слов и приведение слов к нормальной форме (удаление окончаний и суффиксов).

Последующий анализ текстов был направлен на извлечение ключевых слов (параметров моделей ЦСС и АПК), а также на определение качества программного кода. С этой целью авторами была разработана программа на языке Python с использованием библиотеки Natural Language Tool Kit [26].

4. Основные результаты

Анализ исходных данных показал, что тексты 516 ВКР (95,2 %) и 100 % рабочих программ по дисциплинам (для всех ИТ-направлений) содержат упоминания тех или иных технологий, присутствующих в текстах вакансий и, соответственно, востребованных на рынке труда. В остальных же 26 работах (4,8 %) студенты использовали или узкоспециализированные технологии (например, язык программирования Lua), или, наоборот, уже потерявшие актуальность (язык программирования Pascal) и поэтому не нашедшие отражения в запросах рынка труда.

Ожидаемым результатом стало то, что *у работодателей наиболее востребован опыт разработки программ на основе веб-технологий*, который был указан в 203 (23,1 %) из числа рассмотренных текстов вакансий. Поэтому далее приводятся результаты именно для данной категории.

Соответствующие дисциплины (в качестве обязательных или элективных) под разными названиями присутствуют в учебных планах всех ИТ-направлений и читаются, как правило, специалистами-практиками в области веб-разработки. Это фактически обеспечивает студентам, прослушавшим данные курсы (487 человек, 89,9 % от общего числа рассмотренных ВКР), достижение как минимум уровня «имеет представление» для языков программирования и инструментальных средств, наиболее распространенных в данной области.

По результатам сопоставления требуемого уровня владения веб-технологиями с показателями, извлеченными из текстов ВКР и рабочих программ:

- у 336 студентов (69,0 %) был определен уровень «имеет представление»;
- у 93 (19,1 %) — уровень «имеет опыт»;
- у 58 (11,9 %) — уровень «уверенно владеет».

То есть *в трети работ студентов так или иначе использовались веб-технологии*. При этом качество программного кода также было на достаточном уровне — средний балл за качество составил 1,63, нулевые оценки отсутствовали.

Для проверки результативности предложенного подхода к сопоставлению уровня профессиональной подготовки студентов с актуальными требованиями рынка труда **были привлечены дополнительные данные. А именно подтвержденная информация о фактическом трудоустройстве выпускников непосредственно после окончания обучения в вузе**. Было установлено, что число выпускников ИТ-направлений, работающих программистами в области веб-технологий, составляет 179 человек (36,8 % от общего числа студентов, ВКР которых использовались для построения модели ЦСС). Результаты сопоставления реально занимаемых позиций и диагностированного уровня подготовки приведены в таблице 2.

Таким образом, для студентов, представивших в своих ВКР разработки в области веб-технологий,

Таблица 2

Фактическое трудоустройство в сопоставлении с диагностированным уровнем подготовки

№ п/п	Фактическое трудоустройство	Диагностированный уровень		
		имеет представление	имеет опыт	уверенно владеет
1	Junior — 97 чел. (54,2 %)	82 чел. (84,5 %)	15 чел. (15,5 %)	0
2	Middle — 52 чел. (29,1 %)	15 чел. (28,8 %)	15 чел. (28,8 %)	22 чел. (42,4 %)
3	Прочие — 30 чел. (16,7 %)	23 чел. (76,7 %)	7 чел. (23,3 %)	0
	Итого — 179 чел.	120 чел. (67,0 %)	37 чел. (20,7 %)	22 чел. (12,3 %)
	Всего — 487 чел.	336 чел. (69,0 %)	93 чел. (19,1 %)	58 чел. (11,9 %)

данные о трудоустройстве в целом соответствуют диагностированному уровню. Результативность подхода подтверждает сравнение строк «Итого» и «Всего»: для группы выпускников с подтвержденным трудоустройством в данной области (строка «Итого») разные уровни подготовки представлены практически теми же пропорциями, что и по всей выборке (строка «Всего»).

Результаты свидетельствуют о том, что *успешного освоения образовательных программ ИТ-направлений университета достаточно для работы на младших позициях этого сегмента рынка труда*. Однако данные по уровню Middle показывают, что почти половина выпускников (42,8 %) уже работали по специальности, выполняли ВКР по заказу предприятий и предоставили справки о внедрении результатов. Поэтому роль вузовского образования здесь требует уточнения. При этом позиции Senior не занимает сразу после окончания вуза ни один из выпускников, несмотря на достаточную для этого квалификацию. Это объясняется отсутствием достаточного стажа и преобладанием вакансий уровня Junior.

Дополнительный анализ данных цифрового следа студентов объясняет тот факт, что позиции Middle занимают 28,6 % выпускников с уровнем «имеет представление». Соответствующие технологии изучались этими студентами в соответствии с учебным планом, но не отражены в ВКР. Характерно, что у этой группы выпускников высокий средний балл (более 4,5). Они выполняли ВКР по темам, отличным от веб-разработки, но реально уже работали в этой области во время учебы в вузе. Они по своей инициативе выполняли выпускные работы с выраженным научно-исследовательским уклоном и в настоящее время совмещают работу с обучением в магистратуре или аспирантуре. В группу «Прочие» попали выпускники, которые трудоустроены после окончания вуза не по специальности, но на условиях свободной занятости (фриланс) они также занимаются веб-разработкой.

5. Заключение

В целом в результате проведенного анализа можно сделать вывод о результативности предложенного подхода для определения качества подготовки выпускников ИТ-направлений. По результатам исследования разработан инструмент для автоматизированного сопоставления результатов освоения образовательных программ и их содержания с требованиями рынка труда. С его помощью мы смогли объективно оценить структуру и содержание образовательных программ, внести обоснованные изменения в учебные планы.

Конечно, для рассмотренной профессиональной сферы соответствующие требования можно достаточно просто формализовать. Мы использовали упрощенные модели, в которых не отражены требования к так называемым *гибким навыкам* (англ. soft skills) — коммуникабельности, готовности

к командной работе и пр. Однако распространение групповых проектов в рамках не только ВКР, но и курсовых работ, а также учебных дисциплин открывает возможности для привлечения дополнительных данных, позволяющих проанализировать и этот аспект.

Для объективности оценки уровня профессиональной компетентности выпускника мы сознательно не учитывали его успеваемость в процессе обучения, в том числе результаты итоговой аттестации — защиту ВКР. В то же время исследование вопроса корреляции успеваемости и профессионального развития на основе анализа конкретных результатов образовательной деятельности представляет большой интерес.

В дальнейшем предполагается развить предложенный подход для сопровождения индивидуальных образовательных траекторий на основе динамической оценки развития профессиональных и личностных качеств студента с использованием полных данных цифрового следа.

Список использованных источников

1. *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
2. *Baker R. S., Inventado P. S.* Educational data mining and learning analytics // Learning Analytics. New York: Springer, 2014. P. 61–75. DOI: 10.1007/978-1-4614-3305-7_4
3. *Siemens G., Baker R. S. J. d.* Learning analytics and educational data mining: towards communication and collaboration // LAK '12. Proc. 2nd Int. Conf. on Learning Analytics and Knowledge. New York: ACM, 2012. P. 252–254. DOI: 10.1145/2330601.2330661
4. *Зеер Э. Ф., Степанова Л. Н.* Портфолио как инструментальное средство самооценивания учебно-профессиональных достижений студентов // Образование и наука. 2018. Т. 20. № 6. С. 139–157. DOI: 10.17853/1994-5639-2018-6-139-157
5. *Бектемесов М. А., Гриншкун В. В., Скиба М. А., Турганбаева А. Р.* Определение структуры, видов решаемых задач и направлений эффективного использования электронного портфолио студентов вузов // Вестник Российского университета дружбы народов. Серия: Информатизация образования. 2017. Т. 14. № 4. С. 406–417. DOI: 10.22363/2312-8631-2017-14-4-406-417
6. *Добудько Т. В., Горбатов С. В., Добудько А. В., Пугач О. И.* Саморазвитие студентов сквозь призму электронного портфолио // Самарский научный вестник. 2019. Т. 8. № 3. С. 275–279. <https://journals.eco-vector.com/2309-4370/article/view/34389>
7. *Рочев К. В., Моданов А. В., Коршунов Г. В.* Реализация личного кабинета работодателя в информационной системе оценки деятельности студентов // Информатика и образование. 2019. № 5. С. 54–63. DOI: 10.32517/0234-0453-2019-34-5-54-63
8. *Шефер О. Р., Носова Л. С., Лебедева Т. Н.* Электронное портфолио в системе подготовки студентов бакалавриата к будущей профессиональной деятельности // Информатика и образование. 2019. № 2. С. 56–62. DOI: 10.32517/0234-0453-2019-34-2-56-62
9. *Ahmed E., Ward R.* Analysis of factors influencing acceptance of personal, academic and professional development e-portfolios // Computers in Human Behavior. 2016. Vol. 63. P. 152–161. DOI: 10.1016/j.chb.2016.05.043

10. *Gikandi J. W.* Promoting competence-based learning and assessment through innovative use of electronic portfolios // *Handbook of Research on Promoting Higher-Order Skills and Global Competencies in Life and Work*. Hershey: IGI Global, 2019. P. 181–208. DOI: 10.4018/978-1-5225-6331-0.ch012
11. *Можаяева Г. В., Слободская А. В., Фещенко А. В.* Информационный потенциал социальных сетей для выявления образовательных потребностей школьников // *Открытое и дистанционное образование*. 2017. № 3. С. 25–30. DOI: 10.17223/16095944/67/4
12. *Степаненко А. А., Шилляев 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. DOI: 10.17223/19986645/52/8
13. *Feshchenko A., Goiko V., Stepanenko A.* Recruiting university entrants via social networks // *EDULEARN17. Proc. 9th Int. Conf. on Education and New Learning Technologies*. Barcelona: IATED, 2017. P. 6077–6082. DOI: 10.21125/edulearn.2017.2375
14. *Напреенко Г. В.* Идентификация текста по его авторской принадлежности на лексическом уровне (формально-количественная модель) // *Вестник Томского государственного университета. Филология*. 2014. № 379. С. 17–23. DOI: 10.17223/15617793/379/3
15. *Резанова З. И., Романов А. С., Мещеряков Р. В.* О выборе признаков текста, релевантных в авторской экспертной деятельности // *Вестник Томского государственного университета. Филология*. 2013. № 6. С. 38–52. DOI: 10.17223/19986645/26/4
16. *Knight S., Buckingham Shum S., Ryan P., Sándor Á., Wang X.* Designing academic writing analytics for civil law student self-assessment // *International Journal of Artificial Intelligence in Education*. 2018. Vol. 28. P. 1–28. DOI: 10.1007/s40593-016-0121-0
17. *Woods B., Adamson D., Miel S., Mayfield E.* Formative essay feedback using predictive scoring models // *KDD '17. Proc. 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. New York: ACM, 2017. P. 2071–2080. DOI: 10.1145/3097983.3098160
18. *Kyle K., Crossley S., Berger C.* The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0 // *Behavior Research Methods*. 2018. Vol. 50. P. 1030–1046. DOI: 10.3758/s13428-017-0924-4
19. *Crossley S. A., Muldner K., McNamara D. S.* Idea generation in student writing: computational assessments and links to successful writing // *Written Communication*. 2016. Vol. 33. Is. 3. P. 328–354. DOI: 10.1177/0741088316650178
20. *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. Vol. 23. P. 35–59. DOI: 10.1016/j.asw.2014.09.002
21. *Skalicky S., Crossley S. A., McNamara D. S., Muldner K.* Identifying creativity during problem solving using linguistic features // *Creativity Research Journal*. 2017. Vol. 29. Is. 4. P. 343–353. DOI: 10.1080/10400419.2017.1376490
22. *Kovanović V., Joksimović S., Mirriahi N., Blaine E., Gašević D., Siemens G., Dawson S.* Understand students' self-reflections through learning analytics // *LAK '18. Proc. 8th Int. Conf. on Learning Analytics and Knowledge*. New York: ACM, 2018. P. 389–398. DOI: 10.1145/3170358.3170374
23. *Ботов Д. С., Мельников А. В.* Концепция автоматизированного формирования актуализированного набора учебных курсов с учетом требований рынка труда на основе интеллектуального анализа результатов обучения // *Инновации в информационных технологиях и образовании: Сборник трудов III Международной научно-практической конференции*. М.: ИТО, 2014. С. 238–244.
24. *Ботов Д. С.* Интеллектуальная поддержка формирования образовательных программ на основе нейросетевых моделей языка с учетом требований рынка труда // *Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника»*. 2019. № 1. С. 5–19. DOI: 10.14529/ctcr190101
25. ISO/IEC 25010:2011 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. <https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en>
26. *Bird S., Klein E., Loper E.* Natural language processing with Python. Sebastopol: O'Reilly Media, 2009.

DIAGNOSTICS OF PROFESSIONAL COMPETENCE OF IT STUDENTS BASED ON DIGITAL FOOTPRINT DATA

I. G. Zakharova^{1,2}, Yu. V. Boganyuk^{1,2}, M. S. Vorobyova^{1,2}, E. A. Pavlova^{1,2}

¹ *University of Tyumen*

625003, Russia, Tyumen, ul. Volodarskogo, 6

² *Sirius University of Science and Technology*

354340, Russia, Sochi, Olimpijskij prospekt, 1

Abstract

The article goal is to demonstrate the possibilities of the approach to diagnosing the level of IT graduates' professional competence, based on the analysis of the student's digital footprint and the content of the corresponding educational program. We describe methods for extracting student professional level indicators from digital footprint text data — courses' descriptions and graduation qualification works. We show methods of comparing these indicators with the formalized requirements of employers, reflected in the texts of vacancies in the field of information technology. The proposed approach was applied at the Institute of Mathematics and Computer Science of the University of Tyumen. We performed diagnostics using a data set that included texts of courses' descriptions for IT areas of undergraduate studies, 542 graduation qualification works in these areas, 879 descriptions of job requirements and information on graduate employment. The presented approach allows us to evaluate the relevance of the educational program as a whole and the level of professional competence of each student based on objective data. The results were used to update the content of some major courses and to include new elective courses in the curriculum.

Keywords: digital footprint, data extraction, higher education, professional competence, labor market, educational result, text analysis.

DOI: 10.32517/0234-0453-2020-35-4-4-11

For citation:

Zakharova I. G., Boganyuk Yu. V., Vorobyova M. S., Pavlova E. A. Diagnostika professional'noj kompetentnosti studentov IT-na-pravlenij na osnove dannykh tsifrovogo sleda [Diagnostics of professional competence of IT students based on digital footprint data]. *Informatika i obrazovanie — Informatics and Education*, 2020, no. 4, p. 4–11. (In Russian.)

Received: February 16, 2020.

Accepted: May 19, 2020.

Acknowledgments

The reported study was funded by RFBR, project number 19-37-51028.

About the authors

Irina G. Zakharova, Doctor of Sciences (Education), Professor, Professor at the Software Department, Institute of Mathematics and Computer Science, University of Tyumen, Russia; i.g.zakharova@utmn.ru; ORCID: 0000-0002-4211-7675

Yulia V. Boganyuk, Postgraduate Student at the Software Department, Institute of Mathematics and Computer Science, University of Tyumen, Russia; y.v.boganyuk@utmn.ru; ORCID: 0000-0002-3663-0173

Marina S. Vorobyova, Candidate of Sciences (Engineering), Docent, Head of the Software Department, Institute of Mathematics and Computer Science, University of Tyumen, Russia; m.s.vorobyova@utmn.ru; ORCID: 0000-0002-1508-4089

Elena A. Pavlova, Associate Professor at the Software Department, Institute of Mathematics and Computer Science, University of Tyumen, Russia; e.a.pavlova@utmn.ru; ORCID: 0000-0002-0408-9888

References

1. Zakharova I. G. Metody mashinnogo obucheniya dlya informatsionnogo obespecheniya upravleniya professional'nym razvitiem studentov [Machine learning methods of providing informational management support for students' professional development]. *Obrazovanie i nauka — The Education and Science Journal*, 2018, vol. 20, no. 9, p. 91–114. (In Russian.) DOI: 10.17853/1994-5639-2018-9-91-114
2. Baker R. S., Inventado P. S. Educational data mining and learning analytics. *Learning Analytics*. New York, Springer, 2014, p. 61–75. DOI: 10.1007/978-1-4614-3305-7_4
3. Siemens G., Baker R. S. J. d. Learning analytics and educational data mining: towards communication and collaboration. *LAK '12. Proc. 2nd Int. Conf. on Learning Analytics and Knowledge*. New York, ACM, 2012, p. 252–254. DOI: 10.1145/2330601.2330661
4. Zeyer E. F., Stepanova L. N. Portfolio kak instrumental'noe sredstvo samoocenivaniya uchebno-professional'nykh dostizhenij studentov [Portfolio as an instrumental means of self-evaluation of educational and professional achievements of students]. *Obrazovanie i nauka — The Education and Science Journal*, 2018, vol. 20, no. 6, p. 139–157. (In Russian.) DOI: 10.17853/1994-5639-2018-6-139-157
5. Bektemesov M. A., Grinshkun V. V., Skiba M. A., Turganbaeva A. R. Opredelenie struktury, vidov reshaemykh zadach i napravlenij ehffektivnogo ispol'zovaniya ehlektronnogo portfolio studentov vuzov [The definition of structure, types of tasks and directions for electronic portfolio students effective use]. *Vestnik Rossiiskogo universiteta druzhby narodov. Seriya: Informatizatsiya obrazovaniya — Bulletin of People's Friendship University of Russia. Series: Informatization of Education*, 2017, vol. 14, no. 4, p. 406–417. (In Russian.) DOI: 10.22363/2312-8631-2017-14-4-406-417
6. Dobudko T. V., Gorbato S. V., Dobudko A. V., Pugach O. I. Samorazvitie studentov skvoz' prizmu ehlektronnogo portfolio [Students' self-development through the prism of electronic portfolio]. *Samarskij nauchnyj vestnik — Samara Journal of Science*, 2019, vol. 8, no. 3, p. 275–279. (In Russian.) Available at: <https://journals.eco-vector.com/2309-4370/article/view/34389>
7. Rochev K. V., Modanov A. V., Korshunov G. V. Realizatsiya lichnogo kabineta rabotodatelya v informatsionnoj sisteme otsenki deyatel'nosti studentov [Implementation of the personal account of the employer in the information system for assessing students' activity]. *Informatika i obrazovanie — Informatics and Education*, 2019, no. 5, p. 54–63. (In Russian.) DOI: 10.32517/0234-0453-2019-34-5-54-63
8. Shefer O. R., Nosova L. S., Lebedeva T. N. Ehlektronnoe portfolio v sisteme podgotovki studentov bakalavriata k budushhej professional'noj deyatel'nosti [E-portfolio in the system of preparation of students of baccalaureate for future professional activity]. *Informatika i obrazovanie — Informatics and Education*, 2019, no. 2, p. 56–62. (In Russian.) DOI: 10.32517/0234-0453-2019-34-2-56-62
9. Ahmed E., Ward R. Analysis of factors influencing acceptance of personal, academic and professional development e-portfolios. *Computers in Human Behavior*, 2016, vol. 63, p. 152–161. DOI: 10.1016/j.chb.2016.05.043
10. Gikandi J. W. Promoting competence-based learning and assessment through innovative use of electronic portfolios. *Handbook of Research on Promoting Higher-Order Skills and Global Competencies in Life and Work*. Hershey, IGI Global, 2019, p. 181–208. DOI: 10.4018/978-1-5225-6331-0.ch012
11. Mozhaeva G. V., Slobodskaya A. V., Feshchenko A. V. Informatsionnyj potentsial sotsial'nykh setej dlya vyyavleniya obrazovatel'nykh potrebnostej shkol'nikov [Informational potential of social networks for revealing pupils educational needs]. *Otkrytoe i distantsionnoe obrazovanie — Open and Distance Education*, 2017, no. 3, p. 25–30. (In Russian.) DOI: 10.17223/16095944/67/4
12. Stepanenko A. A., Shilyaev K. S., Rezanova Z. I. Atributsiya professional'nykh interesov pol'zovatelej sotsial'noj seti "VKontakte" na osnove tekstov tematicheskikh grupp i personal'nykh stranits [Attribution of professional interests of social network users based on subject-oriented groups and personal pages]. *Vestnik Tomskogo gosudarstvennogo universiteta — Tomsk State University Journal of Philology*, 2018, no. 52, p. 130–144. (In Russian.) DOI: 10.17223/19986645/52/8
13. Feshchenko A., Goiko V., Stepanenko A. Recruiting university entrants via social networks. *EDULEARN17. Proc. 9th Int. Conf. on Education and New Learning Technologies*. Barcelona, IATED, 2017, p. 6077–6082. DOI: 10.21125/edulearn.2017.2375
14. Napreenko G. V. Identifikatsiya teksta po ego avtorskoj prinadlezhnosti na leksicheskom urovne (formal'no-kolichestvennaya model') [Authorship identification of the text on the lexical level (formal-quantitative model)]. *Vestnik Tomskogo gosudarstvennogo universiteta — Tomsk State University Journal*, 2014, no. 379, p. 17–23. (In Russian.) DOI: 10.17223/15617793/379/3
15. Rezanova Z. I., Romanov A. S., Meshcheryakov R. V. O vybere priznakov teksta, relevantnykh v avtorovedcheskoj ehkspertnoj deyatel'nosti [Selecting text features relevant for authorship attribution]. *Vestnik Tomskogo gosudarstvennogo universiteta — Tomsk State University Journal of Philology*, 2013, no. 6, p. 38–52. (In Russian.) DOI: 10.17223/19986645/26/4
16. Knight S., Buckingham Shum S., Ryan P., Sándor Á., Wang X. Designing academic writing analytics for civil law student self-assessment. *International Journal of Artificial Intelligence in Education*, 2018, vol. 28, p. 1–28. DOI: 10.1007/s40593-016-0121-0
17. Woods B., Adamson D., Miel S., Mayfield E. Formative essay feedback using predictive scoring models. *KDD '17*.

Proc. 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. New York, ACM, 2017, p. 2071–2080. DOI: 10.1145/3097983.3098160

18. Kyle K., Crossley S., Berger C. The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, 2018, vol. 50, p. 1030–1046. DOI: 10.3758/s13428-017-0924-4

19. Crossley S. A., Muldner K., McNamara D. S. Idea generation in student writing: computational assessments and links to successful writing. *Written Communication*, 2016, vol. 33, is. 3, p. 328–354. DOI: 10.1177/0741088316650178

20. 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, vol. 23, p. 35–59. DOI: 10.1016/j.asw.2014.09.002

21. Skalicky S., Crossley S. A., McNamara D. S., Muldner K. Identifying creativity during problem solving using linguistic features. *Creativity Research Journal*, 2017, vol. 29, is. 4, p. 343–353. DOI: 10.1080/10400419.2017.1376490

22. Kovanović V., Joksimović S., Mirriahi N., Blaine E., Gašević D., Siemens G., Dawson S. Understand students' self-reflections through learning analytics. *LAK '18. Proc. 8th Int. Conf. on Learning Analytics and Knowledge*. New York, ACM, 2018, p. 389–398. DOI: 10.1145/3170358.3170374

23. Botov D. S., Melnikov A. V. Kontseptsiya avtomatizirovannogo formirovaniya aktualizirovannogo nabora

uchebnykh kursov s uchetom trebovaniy rynka truda na osnove intellektual'nogo analiza rezul'tatov obucheniya [The concept of automated formation of an updated set of training courses, taking into account the requirements of the labor market based on the intellectual analysis of learning outcomes]. *Innovatsii v informatsionnykh tekhnologiyakh i obrazovanii: Sbornik trudov III Mezhdunarodnoj nauchno-prakticheskoy konferentsii [Innovations in information technology and education. Proc. 3d Int. Scientific-Practical Conf.]*. Moscow, ITO, 2014, p. 238–244. (In Russian.)

24. Botov D. S. Intellektual'naya podderzhka formirovaniya obrazovatel'nykh programm na osnove nejrosetevykh modelej yazyka s uchetom trebovaniy rynka truda [Intelligent support development of educational programs based on the neural language models taking into account of the labor market requirements]. *Vestnik YUUrGU. Seriya "Komp'yuternye tekhnologii, upravlenie, radioelektronika" — Bulletin of SUSU. Series "Computer Technologies, Automatic Control, Radioelectronics"*, 2019, no. 1, p. 5–19. (In Russian.) DOI: 10.14529/ctcr190101

25. ISO/IEC 25010:2011 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. <https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en>

26. Bird S., Klein E., Loper E. Natural language processing with Python. Sebastopol: O'Reilly Media, 2009.

НОВОСТИ

Microsoft создала один из пяти самых мощных суперкомпьютеров в мире для решения задач искусственного интеллекта

Microsoft разработала новый суперкомпьютер на базе облака Azure в сотрудничестве и эксклюзивно для OpenAI, некоммерческой исследовательской организации, основателями которой являются Илон Маск и Сэм Альтман. Он стал одним из пяти самых мощных суперкомпьютеров в мире.

Суперкомпьютер создан специально для тренировки следующего поколения массивных ИИ-моделей, способных выполнять более широкий спектр функций и, соответственно, требующих больших вычислительных ресурсов. Об этом компания сообщила в рамках своей крупнейшей технологической конференции Microsoft Build 2020. Новость стала следующим этапом анонсированного в прошлом году сотрудничества с OpenAI.

Исторически эксперты по машинному обучению разрабатывали небольшие ИИ-модели, которые использовали наборы размеченных данных, чтобы научиться выполнять какую-то одну отдельную задачу — распознать объект, перевести текст и т. д. Сообщество специалистов по искусственному интеллекту доказало, что некоторые из этих задач лучше выполняются одной большой моделью, обученной, к примеру, на миллиардах страниц публично доступных текстов. В результате такая модель настолько тонко понимает нюансы речи, грамматику, концепции и контекст, что способна отлично выполнять несколько задач: подво-

дить итоги длинных выступлений, модерировать чат в онлайн-играх и даже генерировать программный код по результатам поиска на GitHub. Microsoft, к примеру, разработала собственное семейство подобных моделей — Microsoft Turing models, — и самую крупную из существующих на данный момент языковую модель с 17 миллиардами параметров Turing Natural Language Generation (T-NLG).

Создание и тренировка таких массивных моделей требуют наличия прогрессивной супервычислительной инфраструктуры или кластеров современного оборудования, соединенных с помощью сетей высокой пропускной способности. Суперкомпьютер, разработанный для OpenAI, — это единая система с более чем 285 000 процессорных ядер (CPU core), 10 000 графических процессоров (GPU) и скоростью сетевого подключения 400 гигабит в секунду для каждого GPU-сервера. Благодаря размещению в облаке Azure суперкомпьютер обладает преимуществами современной облачной инфраструктуры, включая быстрое развертывание, надежность датацентров и доступ к сервисам Azure.

Разработка нового суперкомпьютера для OpenAI — первый шаг на пути к тому, чтобы сделать следующее поколение массивных ИИ-моделей и инфраструктуру, необходимую для их обучения, доступными широкому кругу организаций и разработчиков во всем мире.

(По материалам CNews)