

ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕВЫХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ИССЛЕДОВАНИЯ ВОСТРЕБОВАННОСТИ ПРОФЕССИОНАЛЬНЫХ КОМПЕТЕНЦИЙ ВЫСШЕГО ОБРАЗОВАНИЯ НА РЫНКЕ ТРУДА

Белов Сергей Дмитриевич¹, Зрелов Пётр Валентинович², Ильина Анна Владимировна³, Кореньков Владимир Васильевич⁴, Тарабрин Виталий Анатольевич⁵

¹Начальник сектора;
Объединенный институт ядерных исследований;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: belov@jinr.ru.

²Кандидат физико-математических наук, доцент;
Государственный университет «Дубна»;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
Начальник отдела;
Объединенный институт ядерных исследований;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: zrelov@jinr.ru.

³Ассистент;
Государственный университет «Дубна»;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
Стажер-исследователь;
Объединенный институт ядерных исследований;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: annailina@jinr.ru.

⁴Доктор технических наук, профессор;
Государственный университет «Дубна»;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
Научный руководитель ЛИТ;
Объединенный институт ядерных исследований;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: korenkov@jinr.ru.

⁵Инженер-программист;
Объединенный институт ядерных исследований;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: vitaliy.tarabrin@gmail.com

В работе представлены результаты развития методов сопоставления программ высшего образования и потребностей рынка труда. Данные методы разработаны авторами в рамках реализации аналитической платформы автоматизированного мониторинга рынка труда и интеллектуального анализа кадровых потребностей по номенклатуре специальностей вуза. Для анализа использовались тексты названий, уровней освоения и индикаторов достижения компетенций IT-профиля, а также векторные языковые модели с различными архитектурами и обучающими корпусами текстов. Приведено сравнение результатов и показана устойчивость методов сопоставления образовательных компетенций и потребностей рынка труда.

Ключевые слова: рынок труда, система высшего образования, обработка текста, семантический анализ, языковые модели, нейронные сети, аналитические системы.

Для цитирования:

Использование нейросетевых языковых моделей для исследования востребованности профессиональных компетенций высшего образования на рынке труда / С. Д. Белов, П. В. Зрелов,



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/deed.ru>

А. В. Ильина [и др.] // Системный анализ в науке и образовании: сетевое научное издание. 2023. № 3. С. 13-25. EDN: PBNLSC. URL : <https://sanse.ru/index.php/sanse/article/view/585>.

USE OF NEURAL NETWORK LANGUAGE MODELS TO STUDY THE DEMAND FOR PROFESSIONAL COMPETENCIES OF HIGHER EDUCATION IN THE LABOR MARKET

Belov Sergey D.¹, Zrelov Pert V.², Ilina Anna V.³, Korenkov Vladimir V.⁴, Tarabrin Vitalij A.⁵

¹Head of sector;

Joint Institute for Nuclear Research;

6 Joliot-Curie St, Dubna, Moscow Region, 141980, Russia;

e-mail: belov@jinr.ru.

²PhD in Physics and Mathematics, associate professor;

Dubna State University;

19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;

Head of division;

Joint Institute for Nuclear Research;

6 Joliot-Curie St, Dubna, Moscow Region, 141980, Russia;

e-mail: zrelov@jinr.ru.

³Assistant;

Dubna State University;

19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;

Research assistant;

Joint Institute for Nuclear Research;

6 Joliot-Curie St, Dubna, Moscow Region, 141980, Russia;

e-mail: annailina@jinr.ru.

⁴Grand PhD in Science in Engineering, professor;

Dubna State University;

19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;

Scientific director of MLIT;

Joint Institute for Nuclear Research;

6 Joliot-Curie St, Dubna, Moscow Region, 141980, Russia;

e-mail: korenkov@jinr.ru.

⁵Software engineer;

Joint Institute for Nuclear Research;

6 Joliot-Curie St, Dubna, Moscow Region, 141980, Russia;

e-mail: tarabrin@jinr.ru.

The paper presents the results of the development of methods for comparing higher education programs and the needs of the labor market. These methods were developed by the authors as part of the implementation of an analytical platform for automated monitoring of the labor market and intelligent analysis of personnel against university specialties. Texts of titles, levels of development and indicators of IT-profile competence achievement, as well as vector language models with various architectures and training text bodies were used for the analysis. The results are mutually compared and the stability of the methods of matching educational competencies and the needs of the labor market is shown.

Keywords: labor market, higher education, text processing, semantic analysis, language models, neural networks, analytical systems.

For citation:

Belov S. D., Zrelov P. V., Ilina A. V., Korenkov V. V., Tarabrin V. A. Use of neural network language models to study the demand for professional competencies of higher education in the labor market. *System analysis in science and education*, 2023;(3):13-25 (in Russ). EDN: PBNLSC. Available from: <https://sanse.ru/index.php/sanse/article/view/585>.

Введение

Исследованию различных аспектов рынка труда с помощью математических методов посвящено множество работ, а в последние годы все большее широкое применение находит общий набор «инструментальных» средств – машинное обучение, алгоритмы обработки естественного языка, семантические методы, онтологии [1–8]. Более заметной становится тенденция на создание программных комплексов и систем поддержки принятия решений, сквозной обработки и анализа данных и т.д. [3, 9].

С данными исследованиями связана задача выявления степени соответствия системы образования потребностям рынка труда. Взаимодействие рынка труда и системы профессионального образования является сложным, в идеале взаимно согласованным процессом, в результате которого рынок труда покрывает свои потребности в квалифицированной рабочей силе подготовленными сферой образования выпускниками. При этом система образования должна быстро реагировать на изменения экономической ситуации, требований к профессиональным навыкам и знаниям, а также учитывать прогнозные потребности рынка труда в специалистах.

В работах [1–3] представлена разработанная авторами аналитическая платформа, реализующая автоматизированный мониторинг и анализ рынка труда в Российской Федерации, а также мониторинг соответствия кадровых потребностей работодателей уровню подготовки специалистов.

Платформа основана на решениях и технологиях Больших данных и реализует полный цикл обработки данных – от сбора и хранения до семантического анализа и сервисов визуализации результатов и принятия решений. Информационной базой служат открытые источники – базы данных вакансий «HeadHunter»¹, «Работа России»² и «SuperJob»³, официальный перечень профессий Министерства труда и социальной защиты Российской Федерации⁴, стандарты высшего образования⁵.

В отличие от традиционных подходов сопоставления и анализа содержания образовательных программ, основанных на онтологических моделях и экспертных системах, в работах [1–3] анализ соответствия образовательных стандартов потребностям рынка труда впервые произведен на основе алгоритма перевода коротких предложений в векторное пространство с использованием нейросетевой модели *Word2Vec* [10]. Анализ потребностей рынка труда основан на изучении требований, предъявляемых к соискателю, к его навыкам, знаниям и умениям, отраженных в объявлениях о вакансиях, и на сопоставлении их с содержанием программ образовательных стандартов.

В исследованиях [1–3] сравнение происходит в рамках компетентностной модели системы образования. Сопоставление производится на основе предложенных в этих работах иерархических моделей рынка труда и образования с помощью автоматизированной информационно-аналитической системы. Аналитический блок системы построен с использованием технологий машинного обучения для векторного представления слов и коротких предложений. Система позволяет количественно оценивать потребность в конкретных профессиях для регионов, решать вопрос соответствия профессиональных стандартов реальным рыночным рабочим местам, диапазоны зарплат и уровень требований к соискателям вакансий.

В настоящей работе представлены результаты, связанные с развитием предложенных в [1–3] методов. Исходными данными со стороны системы высшего образования являются описания образовательных профессиональных компетенций. Однако в отличие от прежнего подхода используются не только тексты названий, но и тексты индикаторов достижения профессиональных компетенций, что потребовало определенных изменений и дополнений в методике. В рамках нового

¹ *HeadHunter* – один из крупнейших сайтов по поиску работы и сотрудников в мире (по рейтингу *Similarweb*). Интернет-ресурс содержит около 51 млн резюме, 1,032 тыс. вакансий в базе данных (<https://hh.ru>).

² Веб-портал «Работа России» – федеральная государственная информационная система Федеральной службы по труду и занятости (<https://trudvsem.ru>).

³ Официальный сайт *SuperJob* – IT-компания, осуществляющая разработку в области технологий найма и поиска работы (<https://www.superjob.ru>).

⁴ Официальный интернет-ресурс Министерства труда и социальной защиты Российской Федерации, предоставляющий информацию о перечне профессиональных стандартов (<https://profstandart.rosmintrud.ru/>).

⁵ Федеральные государственные образовательные стандарты (<http://fgos.ru>).

подхода произведено сравнение ряда векторных нейросетевых моделей естественного языка, имеющих различную архитектуру и обученных на различных текстовых корпусах.

Описание исходных данных

На официальном сайте Государственного университета «Дубна»⁶ доступны открытые текстовые материалы, содержащие информацию о профессиональной компетентности выпускников различных программ. Эта информация содержится в образовательных программах, которые регулируют структуру, содержание и качество образования в соответствии с федеральными образовательными стандартами.

Для целей исследования были взяты 14 профессиональных компетенций разных направлений и профилей из области информационных технологий уровня бакалавриата. В текстах образовательных программ Государственного университета «Дубна» каждая компетенция представлена⁷: 1) кодом и наименованием профессиональной компетенции, 2) кодом и наименованием индикатора достижения профессиональной компетенции, 3) основанием разработки профессиональной компетенции.

Поле «Код и наименование индикатора достижения профессиональной компетенции» представляет собой детализацию поля «Код и наименование профессиональной компетенции». Каждой компетенции для удобства идентификации был присвоен числовой индекс (см. табл. 1 в Приложении).

Подход к выбору исходных текстов данных со стороны рынка труда остался тем же, что и в работе [3]. Исследуемые данные были представлены тестовой выборкой объемом около ста тысяч объявлений о вакансиях в сфере «Информационные технологии, интернет, телеком», размещенных на платформе *HeadHunter* в период с января по февраль 2022 г. Для анализа использовались тексты раздела «Обязанности».

Нейросетевые языковые модели

В исследовании использовались нейросетевые модели *Word2Vec* [10], *FastText* [11] и *BERT* [12]: модель семейства *Word2Vec*, предварительно обученная на Национальном корпусе русского языка (НКРЯ) за 2017 г.⁸ и размещенная на сервисе *RusVectōrēs*⁹ под названием «*ruscorpora_upos_skipgram_300_10_2017*» в 2017 г. Размер вектора модели 300, контекстное окно – 10 слов [13].

Модель семейства *FastText*, предварительно обученная на русскоязычном веб-корпусе за 2018 г., содержащем около 10 миллионов слов, с размером вектора 300 и контекстным окном из 5 слов. Модель доступна на сервисе *RusVectōrēs* под названием «*araneum_none_fasttextskipgram_300_5_2018*» [13].

Модель семейства *BERT* под названием «*sbert_large_mt_nlu_ru*», применяемая *SberDevices*¹⁰ в решении широкого спектра задач обработки¹¹ и понимания¹² естественного языка, среди которых:

⁶ Университет «Дубна», раздел «Образование» (<https://uni-dubna.ru/sveden/education/programs>)

⁷ Набор полей, представляющих профессиональную компетенцию, зависит от типа компетентностной модели, выбранной вузом. Так, например, в образовательных программах Российского экономического университета им. Г. В. Плеханова в зависимости от года их разработки профессиональные компетенции представлены не только формулировками названий, но также формулировками индикаторов достижения или уровней освоения «Знать», «Уметь», «Владеть».

⁸ Размер корпуса – 250 миллионов слов.

⁹ Сервис *RusVectōrēs* содержит предварительно обученные дистрибутивно-семантические модели [13].

¹⁰ *SberDevices* — компания, входящая в экосистему Сбера и занимающаяся разработкой решений на базе искусственного интеллекта для таких сфер, как анализ речи, распознавание естественного языка, а также технологии лицевой и голосовой биометрии (<https://sberdevices.ru/>).

¹¹ *Natural Language Processing, NLP*.

¹² *Natural Language Understanding, NLU*.

распознавание намерений, выделение именованных сущностей, сентимент-анализ, поиск схожих запросов. Является улучшенной версией модели «*sbert_large_nlu_ru*», предварительно обученной на наборе текста из 16 миллиардов русскоязычных токенов, включая публичные и закрытые источники данных. Размер вектора – 1024, общее количество параметров – 427 миллионов [14].

Также, как и в работе [3], определялась семантическая близость текстов вакансий, представляемых полем «обязанности», и текстов наименований (названий) компетенций образовательных стандартов. В дополнение в качестве альтернативного подхода вместо названий компетенций использовались их индикаторы. В качестве метрики сходства использовалось косинусное расстояние между векторами в семантическом пространстве – чем оно больше, тем выше степень семантической близости текстов, и, следовательно, более актуальными представляются данные образовательных компетенций на рынке труда.

Общая схема подхода представлена на рис. 1.

Подготовка исходных данных

Необходимость реализации различных алгоритмов предварительной обработки входных текстовых данных обусловлена использованием моделей различных архитектур, обученных на разных текстовых корпусах.

Поскольку в исходных текстах нередко встречаются терминологические аббревиатуры (например, ИБ – информационная безопасность или АС – автоматизированная система), авторами составлен специализированный словарь, содержащий соответствия между аббревиатурами и их полными расшифровками, представленными не только в лемматизированной форме, но и в частеречной разметке.

Также в исходных текстах со стороны системы высшего образования нередко встречаются слова, не несущие в себе принципиально значимый смысл, отражающий семантическое ядро профессиональной компетенции. Список таких слов составлен авторами в результате отдельно проведенной работы по их выявлению. При обработке текстов со стороны системы образования данный список используется в качестве набора стоп-слов, подлежащих исключению из текста.

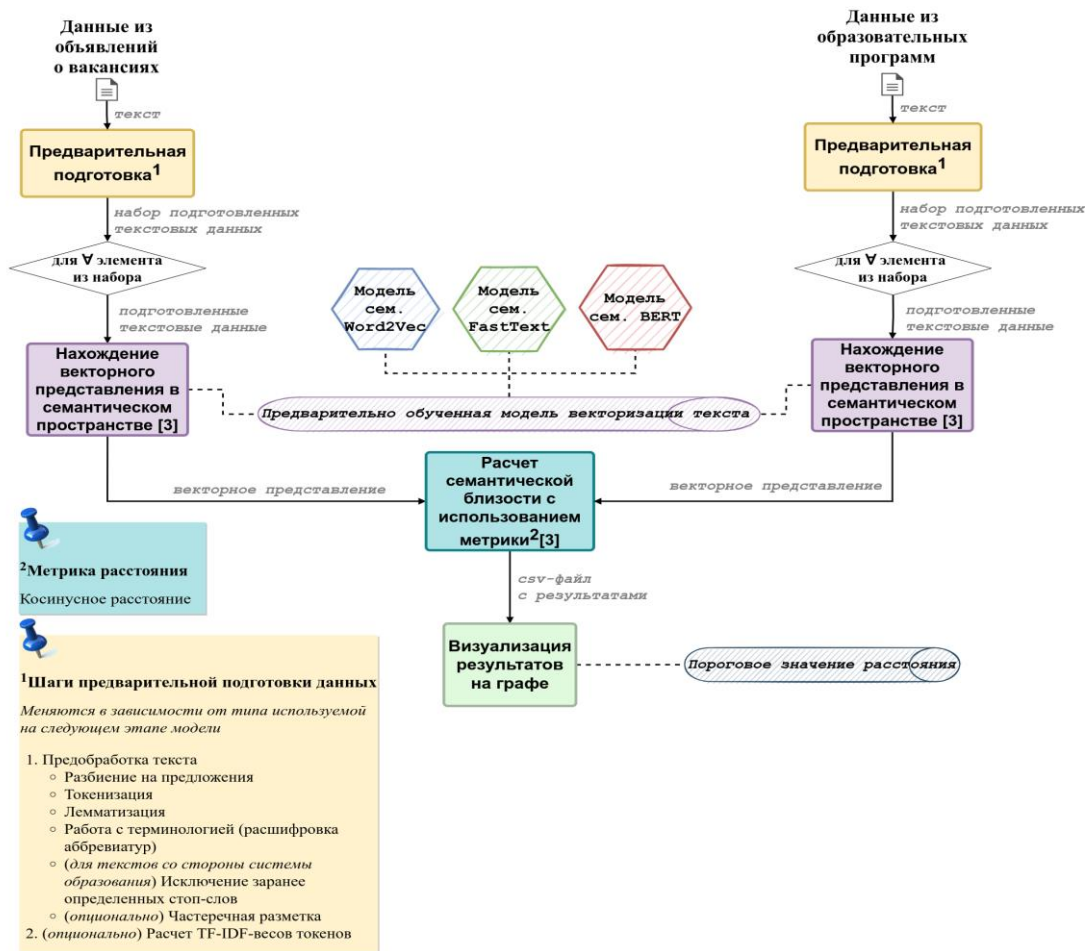


Рис. 1. Общая схема подхода

Для использования модели семейства *Word2Vec* исходные текстовые данные со стороны как системы образования, так и рынка труда были предварительно лемматизированы и размечены в соответствии с частями речи по стандарту *UPOS*¹³. Модель возвращает вектор размером 300 для каждого входного токена¹⁴.

Для использования модели семейства *FastText* исходные данные лемматизировались, но, в отличие от *Word2Vec*, уже без частеречной разметки. Алгоритм получения векторного представления текста аналогичен алгоритму для модели *Word2Vec*.

Разработанная *Google* в 2018 году модель *BERT*¹⁵ имеет архитектуру трансформера и состоит из двух основных компонентов – кодировщика и декодировщика. Кодировщик представляет собой последовательность слоев трансформеров, которые принимают входной текст и генерируют его векторное представление. Декодер используется для генерации выходного текста на основе этого векторного представления. Одним из ключевых аспектов является использование механизма понимания, который позволяет модели учитывать контекст слова при его обработке. Это помогает модели понимать значение слов в зависимости от их контекста и генерировать более точные результаты.

Для использования модели *BERT* не требовалось никакой предобработки.

Визуализация результатов

Одним из возможных подходов к визуализации результатов является создание взвешенного графа, отражающего связи между профессиональными компетенциями и потребностями рынка труда [1]. В таком графе образовательные компетенции и требования рынка труда представляются

¹³ *Universal POS tags* (<https://universaldependencies.org/u/pos/>).

¹⁴ Под токеном в данной работе понимается одно слово из входного текста (токенизация на уровне слов).

¹⁵ *Bidirectional Encoder Representations from Transformers*.

вершинами разных типов. Вершины, относящиеся к образовательным компетенциям, маркируются индексом компетенции. Ребра графа связывают только вершины разных типов и показывают наличие и степень связи между ними, определяемой косинусным расстоянием. Важным преимуществом этого подхода является возможность фильтрации связей между вершинами с помощью изменения порогового значения. Порог позволяет отбирать только такие ребра, вес которых превышает его значение.

В качестве инструмента визуализации и анализа графов использовано программное обеспечение *Gephi*¹⁶, имеющее не только широкий спектр алгоритмов укладки графов, но и обширный функционал, позволяющий форматировать внешний вид графа, применять фильтрацию связей и отражать статистику по графу. *Gephi* является бесплатным приложением с открытым исходным кодом¹⁷.

На рис. 2 представлен граф связей между профессиональными компетенциями и требованиями рынка труда, построенный с различными пороговыми значениями (для большей наглядности представлены связи только для 500 вакансий). Вершины, представляющие профессиональные компетенции и имеющие большую степень, расположены ближе к центру графа, чем аналогичные вершины с меньшей степенью. Степень вершины определяется числом ее ребер. Поэтому более востребованные на рынке компетенции, т. е. имеющие большое количество связей с вакансиями, имеют большую степень, равную количеству ребер у представляющих их вершин графа. При отсутствии порогового значения отображаются все имеющиеся между вершинами профессиональных компетенций и вакансий ребра (рис. 2а). При выборе порогового значения, равного 0.8 (рис. 2б), остаются только те ребра, которые связывают семантически наиболее близкие вершины (косинусное расстояние и, соответственно, вес ребра, равные 0.8 и выше). При выборе порогового значения, равного 0.87 (рис. 2в), граф отображает только самые сильные связи. Фильтрация связей на основе выбора порогового значения позволяет выявить профессиональные компетенции, формулировки которых имеют наибольшее отражение на рынке труда.

Выбор (подбор) порогового значения важен, если надо отобразить только определенную часть ребер графа, например, с целью сравнить результаты разных методов в одинаковых условиях. Так, если мы хотим сравнить результаты, зная, что конкретный вуз готовит специалистов только для определенной части рынка труда, «претендуя», например, лишь на 80% вакансий, то и все оценки следует производить с учетом этого фактора.

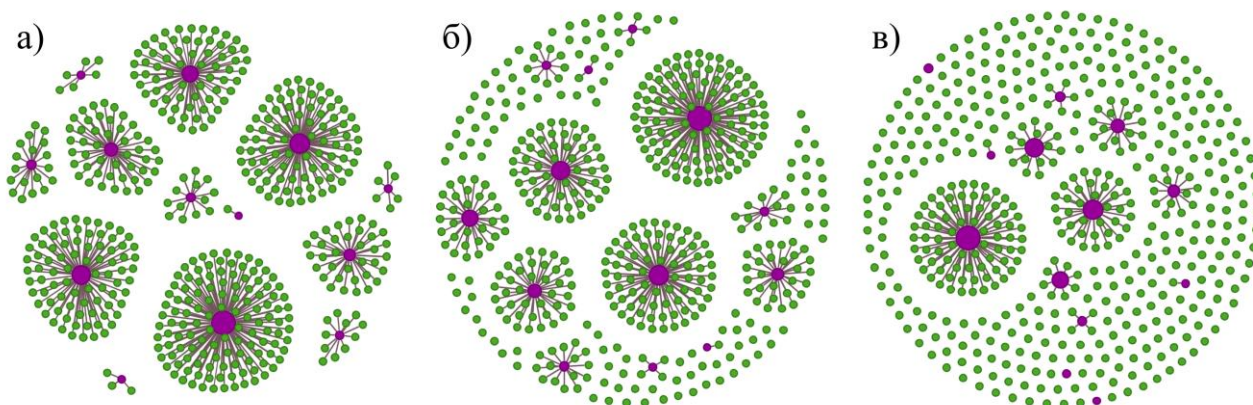


Рис. 2. Влияние выбора порогового значения на количество отображаемых ребер графа, представляющих связи между профессиональными компетенциями и требованиями рынка труда. Зеленым цветом обозначены вершины, представляющие вакансии. Розовым цветом обозначены вершины, представляющие профессиональные компетенции. Размер каждой вершины пропорционален ее степени

Для укладки графа использовался быстрый алгоритм *OpenOrd*, разработанный для очень больших графов и обеспечивающий среднюю степень точности [15]. *OpenOrd* может масштабироваться до более чем 1 млн. узлов, что делает его практически идеальным инструментом

¹⁶ *Gephi* (<https://gephi.org/>).

¹⁷ Исходный код доступен на платформе *GitHub* (<https://github.com/gephi/gephi>).

для больших графов [16]. Данный алгоритм предполагает использование неориентированных взвешенных графов и нацелен на лучшее выделение кластеров [16]. Для отображения итоговых графов, демонстрирующих результаты проведенных исследований, параметры алгоритма укладки подбирались эмпирически¹⁸ – фаза жидкости (*Liquid*) = 5%, фаза расширения (*Expansion*) = 30%.

Анализ результатов

На рис. 3 представлены графы соответствия профессиональных компетенций программ ИТ-направлений Государственного университета «Дубна» вакансиям рынка труда в сфере «Информационные технологии, интернет, телеком» (см. параграф «Описание исходных данных») для двух типов исходных данных и трех языковых моделей.

При сравнении работы алгоритмов пороговые значения для каждой модели выбирались таким образом, чтобы доля отобранных ребер была одинаковой и составляла 80%. Так, для модели *Word2Vec* порог был установлен равным 0.8 (как в случае использования только названий (наименований) компетенций, так и в случае использования индикаторов достижения). Для модели *FastText* порог равнялся 0.77 (в случае названия компетенций) и 0.78 (в случае индикаторов достижения). Для модели *BERT* – 0.72 и 0.74, соответственно.

Хорошо видно, что все 14 компетенций, независимо от типа их представления и используемой модели, отражены в центре графа, в области наиболее сильных связей.

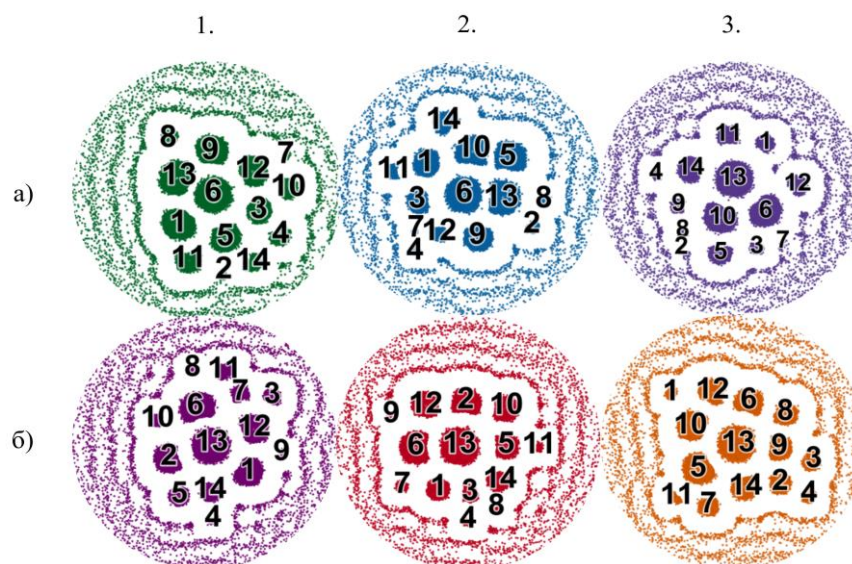


Рис. 3. Сравнение графов «Профессиональная компетенция - Вакансия», отображающих результаты исследований с использованием различных уровней детализации формулировок профессиональных компетенций – названий (а) и индикаторов достижения (б), а также различных моделей – *Word2Vec* (1), *FastText* (2), *BERT* (3) для набора профессиональных компетенций программ ИТ-направлений Государственного университета «Дубна»

На рис. 4 представлены результаты анализа графов, которые позволяют сделать ряд выводов о закономерностях, наблюдаемых во взаимосвязи системы высшего образования и рынка труда.

Независимо от типа исходных данных и выбранной модели наиболее частое отражение на рынке труда имеют компетенции с идентификаторами 6 и 13. Наименьшее отражение имеют компетенции под идентификаторами 4 и 7 (см. табл. 1 в Приложении).

Наибольшая востребованность компетенций наблюдается для вакансий со специализациями: веб-программирование для бизнеса, техническая поддержка, управление ИТ-проектами. Наибольшее количество связей относится к вакансиям с требованием опыта работы от 1 года до 3 лет. Около 10% связей относится к вакансиям без опыта работы (см. табл. 1). Наиболее часто наблюдаются связи для

¹⁸ Подробнее про описание фаз см. в [17].

вакансий с имеющими зарплатами в интервале от 100 000 руб. (8%) до 150 000 руб. (8%). Регионы востребованности компетенций: г. Москва (в среднем около 45%) и г. Санкт-Петербург (в среднем – 14%). Исследуемые компетенции оказались менее востребованы в других регионах (менее 3%).

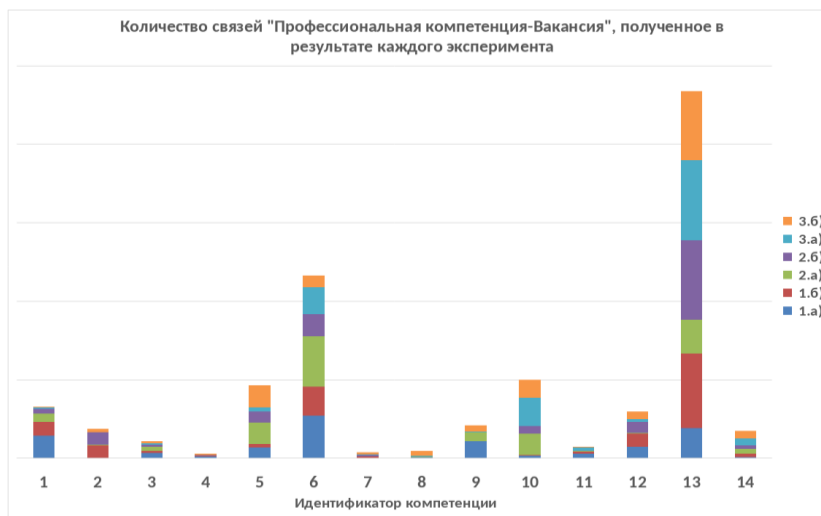


Рис. 4. Столбчатая диаграмма с накоплением, отображающая количество связей «Профессиональная компетенция - Вакансия» для каждой профессиональной компетенции в процентах от общего количества связей

Табл. 1. Процент связей формулировок компетенций разного уровня детализации с формулировками поля «Обязанности» объявлений о вакансиях с требованиями к опыту работы от общего количества всех связей «Профессиональная компетенция - Вакансия» в графе

Опыт работы	Процент от общего количества связей					
	1.а)	1.б)	2.а)	2.б)	3.а)	3.б)
От 1 года до 3 лет	53,80	53,78	53,11	53,85	54,16	54,18
От 3 до 6 лет	34,05	34,01	33,92	32,12	33,18	33,14
Более 6 лет	2,81	2,83	2,77	2,54	2,73	2,73
Нет опыта	9,34	9,38	10,2	11,5	9,92	9,95

Заключение

Разработанный в [1–3] с использованием семантической модели *Word2Vec* подход к определению востребованности на рынке труда профессиональных компетенций высшего образования показал близкие результаты и при использовании других нейросетевых языковых моделей (*FastText* и *BERT*), а также относительную устойчивость к изменению исходных данных в рамках компетентностной образовательной модели. Это свойство указывает на возможность практического применения разработанного подхода для исследования востребованности образовательных компетенций на рынке труда. Сравнение проводилось на примере данных образовательных программ ИТ-направлений Государственного университета «Дубна». В результате проведенной работы представлен ряд выводов о закономерностях, наблюдаемых во взаимосвязи системы высшего образования и рынка труда.

Список источников

1. Мониторинг потребностей рынка труда в выпускниках вузов на основе аналитики с интенсивным использованием данных / П. В. Зрелов, В. В. Кореньков, Н. А. Кутковский, А. Ш. Петросян [и др.] // Труды XVIII Межд. конф. DAMDID/RCDL'2016 «Аналитика и управление данными в областях с интенсивным использованием данных», Ершово, 11-14 октября 2016. – Москва: Федеральный исследовательский центр "Информатика и управление" Российской академии наук, 2016. – С. 124-131.

2. Мониторинг соответствия профессионального образования потребностям рынка труда / С. Д. Валентей, П. В. Зрелов, В. В. Кореньков, С. Д. Белов [и др.] // *Общественные науки и современность*. – 2018. – №3. – С. 5-16.
3. Methods and algorithms of the analytical platform for analyzing the labor market and the compliance of the higher education system with market needs / S. Belov [et al.] // *Proceedings of Science, 2022. – Conf. DLCP2022*. – Pp. 028. – DOI: <https://doi.org/10.22323/1.429.0028>.
4. MEET: A Method for Embeddings Evaluation for Taxonomic Data / L. Malandri, F. Mercori, M. Mezzanatica, N. Nobani // *2020 International Conference on Data Mining Workshops (ICDMW), 2020*. – Pp. 31-38. – DOI: [10.1109/ICDMW51313.2020.00014](https://doi.org/10.1109/ICDMW51313.2020.00014).
5. An AI-based open recommender system for personalized labor market driven education / M. Tavakoli, A. Faraji, J. Vrolijk, M. Molavi [et al.] // *Advanced Engineering Informatics, 2022*. – Vol. 52. – Pp. 101508. – DOI: <https://doi.org/10.1016/j.aei.2021.101508>.
6. OntoJob: Automated Ontology Learning from Labor Market Data / J. Vrolijk, S. T. Mol, C. Weber, M. Tavakoli [et al.] // *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*. – 2022 Pp. 195-200. – DOI: [10.1109/ICSC52841.2022.00040](https://doi.org/10.1109/ICSC52841.2022.00040).
7. Wowczko I. A. Skills and Vacancy Analysis with Data Mining Techniques. – *Informatics*. – 2015. – Vol. 2 (4). – Pp. 31-49. – DOI: [10.3390/informatics2040031](https://doi.org/10.3390/informatics2040031).
8. Classifying online Job Advertisements through Machine Learning / R. Boselli, M. Cesarini, F. Mercurio, M. Mezzanatica // *Future Generation Computer Systems*. – Vol. 86, Issue C. – 2018. – Pp. 319-328.
9. Sparreboom T., Labour market information and analysis systems // *Perspectives on labour economics for development*, Geneva: ILO, 2013. – Pp. 255-282.
10. Efficient Estimation of Word Representations in Vector Space / T. Mikolov et al., [arXiv:1301.3781v3, 2013].
11. Bag of Tricks for Efficient Text Classification / A. Joulin, E. Grave, P. Bojanowski, T. Mikolov. // arXiv:1607.01759v3 [cs:CL] 9 Aug 2016.
12. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // J. Devlin, M.-W. Chang, Kenton Lee, K. Toutanova. // arXiv:1810.04805v2 [cs.CL] 24 May 2019. DOI: <https://doi.org/10.48550/arXiv.1810.04805>
13. Kutuzov A., Kuzmenko E. RusVectōrēs: models. – *RusVectores*, 2023. – URL: <https://rusvectors.org/ru/models/> (дата обращения: 09.09.2023)
14. Блог компании SberDevices. Обучение модели естественного языка с BERT и Tensorflow // Хабр : [сайт]. – Habr, 2006–2023. – Дата публикации: 12.11.2020. – URL: <https://habr.com/ru/companies/sberdevices/articles/527576/>.
15. Cherven K., Magdy M. *Mastering Gephi Network Visualization* – Birmingham, UK: Packt Publishing Ltd., 2015. – 357 с.
16. OpenOrd · gephi/gephi Wiki // GitHub : [web platform]. – GitHub, Inc., 2023. – URL: <https://github.com/gephi/gephi/wiki/OpenOrd> (дата обращения: 18.10.2023).
17. OpenOrd: an open-source toolbox for large graph layout / S. Martin, W. M. Brown, R. Klavans, K. W. Boyack // *Proc. SPIE 7868, Visualization and Data Analysis 2011*. – Pp. 786806 (24 January 2011). – DOI: <https://doi.org/10.1117/12.871402>.

Приложение

Табл. 1. Наименования и индикаторы достижения используемых в работе профессиональных компетенций, извлеченных из образовательных программ Государственного университета «Дубна». Каждой профессиональной компетенции проставлен числовой индекс для однозначной идентификации компетенции

Индекс профессиональной компетенции	Наименование профессиональной компетенции	Индикаторы достижения профессиональной компетенции
1	Способен проектировать и внедрять компоненты корпоративных ИС и ИТ-инфраструктуры, соответствующие оперативным и стратегическим целям деятельности предприятия	Проектирует компоненты корпоративных информационных систем и ИТ-инфраструктуры; Выполняет внедрение корпоративных информационных систем и элементов ИТ-инфраструктуры; Анализирует и реализует оперативные и стратегические цели предприятия средствами информационных технологий
2	Способен проектировать архитектуру ИС, отвечающую требованиям проекта (ТЗ заказчика)	Выполняет анализ и управление требованиями заказчика в проекте автоматизации; Проектирует архитектуру информационных систем
3	Способен осуществлять планирование и организацию выполнения задач организационного управления с учетом утвержденных параметров	Выполняет планирование задач организационного управления с учетом утвержденных параметров; Организует выполнение задач организационного управления в соответствии с существующим планом и критериями оценки
4	Способен проектировать и разрабатывать компоненты корпоративных информационных систем и информационных систем электронного бизнеса	Проектирует компоненты информационных систем электронного бизнеса; Разрабатывает компоненты информационных систем
5	Способен разрабатывать ИТ-сервисы и Интернет-ресурсы, формировать контент и интерфейс, соответствующие требованиям проекта	Разрабатывает ИТ-сервисы и интерфейсы и функционал интернет-ресурсов в соответствии с требованиями проекта; Формирует контент интернет-ресурсов в соответствии с требованиями заказчика
6	Способен выполнять проектную деятельность по разработке и созданию (модификации) ИС, разработке технико-экономического обоснования проектов для улучшения бизнес-процессов и ИТ-инфраструктуры предприятия	Реализует проекты по разработке и модификации информационных систем; Разрабатывает технико-экономическое обоснование проектов по улучшению бизнес-процессов и ИТ-инфраструктуры предприятия

7	Способен разрабатывать требования и проектировать программное обеспечение	<p>Анализирует проблемную ситуацию заинтересованных лиц и выявляет требования к программному обеспечению, продукту, средству, программно-аппаратному комплексу, ИС;</p> <p>Разрабатывает требования к программному обеспечению, продукту, средству, программно-аппаратному комплексу, ИС;</p> <p>Проектирует программное обеспечение согласно требованиям</p>
8	Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности	<p>Выбирает, обосновывает, согласовывает и описывает концептуальную архитектуру программного обеспечения, продукта, средства, программно-аппаратного комплекса, ИС программного обеспечения, продукта, средства, программно-аппаратного комплекса, ИС;</p> <p>Применяет методы и средства проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов;</p> <p>Использует существующие типовые решения и шаблоны проектирования программного обеспечения</p>
9	Способен разрабатывать программные модули и компоненты, осуществлять их интеграцию и проверку работоспособности выпусков программного продукта	<p>Осуществляет интеграцию программных модулей и компонент программного продукта;</p> <p>Осуществляет проверку работоспособности выпусков программного продукта;</p> <p>Разрабатывает программные модули и компоненты</p>
10	Способен выполнять работы и управлять работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы	<p>Способен выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы;</p> <p>Способен управлять работами по созданию и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы</p>
11	Владение навыками использования различных технологий разработки программного обеспечения	Использует различные технологии разработки программного обеспечения
12	Способен анализировать требования к разрабатываемой информационной системе (ИС), включая требования к информационной безопасности; способен участвовать в работах по созданию ИС, ее тестированию и	<p>Анализирует и формирует требования к информационной системе, включая требования к информационной безопасности;</p> <p>Применяет современные технологии и инструментальные средства при решении задач на всех этапах жизненного цикла информационной системы, ее тестированию и интеграции с другими системами</p>

	интеграции с другими системами	
13	Способен проектировать и создавать программное обеспечение, соответствующее требованиям заказчика, включая разработку программного интерфейса, проектирование структур данных, модулей и компонентов программного обеспечения, выполнять процедуры их сборки, и проводить оценку работоспособности программного продукта	<p>Осуществляет проектирование и разработку программных модулей и компонентов, соответствующих требованиям заказчика, включая разработку программного интерфейса и проектирование структур данных с использованием современных технологий и языков программирования;</p> <p>Планирует и выполняет работы по сборке компонентов и модулей программного обеспечения и подтверждению работоспособности и корректности работы программной системы путем организации тестирования и представления результатов тестов</p>
14	Способен участвовать в работах по предотвращению потерь и повреждений данных и обеспечению информационной безопасности при проектировании и эксплуатации информационных технологий и систем	<p>Определяет проектные решения, обеспечивающие предотвращение потерь и повреждений данных и обеспечение информационной безопасности информационных технологий и систем;</p> <p>Планирует и выполняет работы по предотвращению потерь и повреждений данных и обеспечению информационной безопасности при эксплуатации информационных технологий и систем</p>