

На правах рукописи

Яруллин Денис Владимирович

**ИНФОРМАЦИОННАЯ СИСТЕМА СБОРА И ОБРАБОТКИ
ТРЕБОВАНИЙ РАБОТОДАТЕЛЕЙ К КОМПЕТЕНЦИЯМ
ИТ-СПЕЦИАЛИСТОВ НА ОСНОВЕ МЕТОДОВ
ДЕНОТАТИВНОГО АНАЛИЗА**

2.3.4. Управление в организационных системах

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Пермь 2022

Диссертационная работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Пермский национальный исследовательский политехнический университет».

Научный руководитель: **Файзрахманов Рустам Абубакирович,**
доктор экономических наук, профессор

Официальные оппоненты: **Иващенко Антон Владимирович,**
доктор технических наук, профессор,
ФГБУ «Ордена Трудового Красного Знамени
Российский научно-исследовательский институт
радио имени М.И. Кривошеева».
Самарский филиал – «СОНИИР»,
ведущий научный сотрудник
Научно-технического центра ПР 048

Шиков Алексей Николаевич,
кандидат технических наук, доцент,
Северо-Западный институт управления — филиал
Российской академии народного хозяйства
и государственной службы при Президенте
Российской Федерации,
доцент кафедры бизнес-информатики

Ведущая организация: Федеральное государственное автономное
образовательное учреждение высшего
образования «Пермский государственный
национальный исследовательский университет»,
г. Пермь

Защита состоится «16» февраля 2023 года в 16:00 на заседании диссертационного совета Д ПНИПУ.05.21, созданного на базе ФГАОУ ВО «Пермский национальный исследовательский политехнический университет» по адресу: 614990, г. Пермь, Комсомольский проспект, дом 29, аудитория 345.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГАОУ ВО «Пермский национальный исследовательский политехнический университет» (<http://www.pstu.ru>).

Автореферат разослан « ____ » декабря 2022 года.

Ученый секретарь
диссертационного совета,
кандидат экономических наук,
доцент

Алексеев Александр Олегович

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. В соответствии с планом мероприятий по направлению «Кадры и образование» Национальной программы «Цифровая экономика Российской Федерации», развитие ИТ-отрасли и подготовка высококвалифицированных кадров являются одними из приоритетных направлений в экономике Российской Федерации. В настоящее время спрос на квалифицированных специалистов в значительной мере превышает предложение, так как потребность в ИТ-специалистах в различных областях, где происходят процессы автоматизации, цифровизации, аналитики, возрастает; одновременно с этим статистические исследования высшей школы демонстрируют тренд на снижение числа поступающих на специальности, связанные с информационными технологиями.

Это обусловлено тем, что компетенции, которые ожидают работодатели от специалистов, зачастую не соответствуют компетенциям, которые приобретают специалисты в процессе обучения в образовательных организациях. Как следствие, все большее число соискателей предпочитает приобретать необходимые компетенции бессистемно, путем самообразования, краткосрочных программ дополнительного образования, курсов повышения квалификации; укрепляется тренд на переход от очного образования к электронному на основе дистанционных технологий.

При этом значительное число работодателей размещают свои вакансии в открытом доступе на сайтах-агрегаторах; ИТ-компании используют подобные ресурсы в качестве основного метода поиска новых специалистов. Исходя из этого, становится целесообразным рассматривать размещенные на данных веб-ресурсах вакансии в качестве источника требований работодателей относительно компетенций специалистов на те или иные должности в сфере информационных технологий. Мониторинг, обработка и систематизация требований становятся необходимыми для составления и реализации конкурентноспособных образовательных программ в высших учебных заведениях, нацеленных на выпуск специалистов, востребованных на рынке труда, что особенно актуально именно в сфере ИТ.

Степень разработанности темы исследования. Подходы к построению модели специалиста с учетом требований работодателей предложены такими исследователями, как Калевко В. В., Лагерев Д. Г., Подвесовский А. Г., Булах К. В., Бурцева Е. Т., Харитоновна И. Ю. и Сергеев С. Ф. Предлагается анализировать рынок труда в области ИТ-сферы на основании вакансий с сайтов-агрегаторов, индивидуализировать образовательные программы и вовлекать работодателей в процесс обучения.

Исследованием подходов к построению модели специалиста с учетом требований образовательных организаций, Федерального государственного образовательного стандарта (ФГОС) и профессиональных стандартов занимались Галиев Р. М., Курзаева Л. В., Захарова А. А., Воронин А. А., Васильченко А. А., Зенович А. В., Лебедева Т. Ф., Муравьев С. А.

Концепции формирования подготовки специалистов с учетом как требований работодателей, так и образовательных институтов, описаны в работах Гузаирова М. Б., Герасимова И. Б., Васильевой Е. В., Мартынова В. В., Филосовой Е. И., Ширяева О. В., Бакановой А. П., Бюссемакер М., Марвана Х.

Логиновым К. В. описана модель управления процессом прохождения учебного курса с применением событийно-ориентированных игровых механик. Методы автоматизации сбора данных рассматриваются в работах Смида Д., Али А., Ильина Д. Ю.

Возможность учета различных требований рассматривается в большинстве работ, посвященных построению модели специалиста, однако исследователи не всегда подходят к решению задачи с точки зрения автоматизации, вследствие чего уменьшается число источников, которые возможно одновременно использовать при построении модели специалиста. Кроме того, не проводится унификация и анализ компетенций, которые требуются от молодых специалистов, что, учитывая слабоструктурированный характер исходных данных о требованиях работодателей, также указывает на необходимость

разработки моделей и методов обработки открытых данных о требованиях предприятий к соискателям.

Объектом исследования является процесс выявления обобщенных требований работодателей к компетенциям ИТ-специалистов.

Предмет исследования — модели, методы и алгоритмы анализа и обработки открытых слабоструктурированных данных для выявления обобщенных требований работодателей к компетенциям соискателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов.

Целью диссертационной работы является повышение эффективности выявления обобщенных требований работодателей к компетенциям соискателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов.

Для достижения указанной цели поставлены следующие **задачи**:

1. Анализ моделей и методов выявления обобщенных требований к ИТ-специалистам для формирования рекомендаций по подготовке ИТ-кадров с учетом выявленных требований;
2. Разработка концепции информационной системы для выявления динамически меняющихся обобщенных требований работодателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов;
3. Разработка методики и алгоритма мониторинга, сбора и предварительной обработки открытых слабоструктурированных данных о требованиях работодателей;
4. Разработка информационной модели ИТ-специалиста-соискателя на основе денотативного подхода и алгоритма ее автоматизированного построения;
5. Разработка информационной модели ИТ-специалиста-выпускника на основе денотативного подхода и методики обогащения данных и алгоритма ее автоматизированного построения;
6. Разработка алгоритма автоматизированного формирования рекомендаций по реализации образовательных программ с учетом обобщенных требований работодателей;
7. Разработка программного обеспечения информационной системы выявления обобщенных требований работодателей к компетенциям соискателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов на основе предложенных моделей и алгоритмов.

Положения, выносимые на защиту и обладающие **научной новизной**:

1. Предложена информационная модель специалиста-соискателя и разработан алгоритм ее автоматизированного построения, **отличающийся** использованием слабоструктурированных открытых данных на естественном языке для мониторинга и обновления представления специалиста с точки зрения работодателей и применением подхода, основанного на нечетких множествах, для идентификации направлений профессиональной деятельности специалиста, что **позволяет** структурировать и формализовать гетерогенные данные о требованиях работодателей для их использования в рекомендательной информационной системе (п. 5 — Разработка методов получения данных и идентификации моделей, прогнозирования и управления организационными системами на основе ретроспективной, текущей и экспертной информации; п. 4 — Разработка информационного и программного обеспечения систем управления и механизмов принятия решений в организационных системах);

2. Предложена информационная модель специалиста-выпускника и разработан алгоритм ее автоматизированного построения, учитывающий требования работодателей и **отличающийся** использованием денотативного подхода для обогащения данных при построении иерархии ключевых понятий предметной области, что **позволяет** обеспечить динамическое построение информационной модели предметной области для описания выпускника, востребованного на рынке труда (п. 5 — Разработка методов получения данных и идентификации моделей, прогнозирования и управления организационными системами на основе ретроспективной, текущей и экспертной информации; п. 4 — Разработка

информационного и программного обеспечения систем управления и механизмов принятия решений в организационных системах);

3. Разработан алгоритм автоматизированного формирования рекомендаций по реализации образовательных программ, **отличающийся** горизонтальной и вертикальной интеграцией изучаемых понятий в рамках предметной области и профессиональных задач заданного множества работодателей, что **позволяет** структурировать и систематизировать содержание образовательных программ в единую иерархическую информационную модель (п. 9 — Разработка методов и алгоритмов интеллектуальной поддержки принятия управленческих решений в организационных системах);

4. Разработано программное обеспечение информационной системы для выявления обобщенных требований работодателей к компетенциям соискателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов, реализующее разработанные методы и алгоритмы и **обеспечивающее** непрерывный сбор и анализ данных о вакансиях ИТ-специалистов и адаптируемость образовательных программ (п. 4 — Разработка информационного и программного обеспечения систем управления и механизмов принятия решений в организационных системах).

Теоретическая значимость диссертационного исследования заключается в уточнении теоретических основ автоматизированного описания предметной области информационных технологий за счет разработки алгоритмов сбора и обработки слабоструктурированных открытых данных на естественном языке и использовании методов теории графов для построения иерархического описания предметной области.

Практическая значимость работы заключается в программной реализации разработанных моделей и алгоритмов. Разработанная информационная система позволяет автоматически выявлять обобщенные требования работодателей к ИТ-специалистам на основе открытых слабоструктурированных веб-данных и формировать рекомендации по реализации образовательных программ подготовки специалистов в области информационных технологий для повышения конкурентоспособности выпускников на рынке труда за счет освоения ими компетенций, востребованных работодателями. В рамках обновления образовательных программ магистратуры по направлениям 09.04.01 «Информатика и вычислительная техника», 09.04.04 «Программная инженерия» на кафедре ИТАС ПНИПУ использование информационной системы позволило выявить 37 востребованных ключевых понятий со 140 связями между ними, что было учтено при подготовке рабочих программ дисциплин.

Методология и методы исследования. Методологическую основу исследований составили работы отечественных и зарубежных авторов в области информационного моделирования обучающихся с целью управления процессом подготовки высококвалифицированных специалистов. В ходе подготовки диссертации применялись методы системного анализа, алгоритмы обработки естественного языка и кластерного анализа, использовались положения теории нечетких множеств и теории графов.

Степень достоверности и апробация работы. Достоверность результатов исследования обеспечена системным подходом к решению задачи моделирования специалиста-выпускника с учетом требований работодателей; опорой на современные подходы к управлению, основанные на данных; использованием средств моделирования и проектирования сложных динамических систем.

Апробация результатов диссертационной работы проводилась на базе Пермского национального исследовательского политехнического университета и ИТ-компаний г. Перми: ООО «Датабриз», ООО «Иннфокус», ООО «Пармалогика». Акты о внедрении результатов диссертации включены в приложения к тексту работы.

Результаты исследования обсуждались на научно-методических семинарах кафедры «Информационные технологии и автоматизированные системы» ПНИПУ; докладывались на 7 международных научных конференциях: XXII, XXIII, XXIV Международных конференциях по мягким вычислениям и измерениям (г. Санкт-Петербург, 2019, 2020, 2021); XXXIII Международной научной конференции «Математические Методы в Технике и Технологиях

(ММТТ-33)» (г. Казань, 2020), Международной научной конференции «Кибер-физические системы: проектирование и моделирование CYBERPHY:2020» (г. Казань, 2020); XXXIV Международной научной конференции «Математические Методы в Технике и Технологиях (ММТТ-34)» (г. Санкт-Петербург, 2021), Международной научной конференции «Кибер-физические системы: проектирование и моделирование CYBERPHY:2021» (г. Санкт-Петербург, 2021).

Публикации. По материалам диссертации опубликовано 11 научных работ, из них 1 статья в рецензируемых научных изданиях, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, 4 статьи проиндексированы в международных реферативных базах и системах цитирования Scopus и Web of Science, получено 1 свидетельство о государственной регистрации программы для ЭВМ, 5 статей опубликованы в прочих изданиях.

Структура диссертации. Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложений. Работа изложена на 152 страницах, содержит 50 рисунков, 12 таблиц, 3 приложения. Список литературы включает 109 источников.

II. ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении описывается актуальность темы диссертации, сформулированы объект и предмет исследования, цель работы, основные задачи, перечислены методы исследования, описана теоретическая и практическая значимость работы, раскрыта научная новизна и сведения о полученных результатах, перечислены положения, выносимые на защиту.

Первая глава посвящена рассмотрению проблемы несоответствия компетенций выпускников образовательных организаций и требований работодателей и анализу существующих подходов к выявлению требований к подготовке специалистов с точки зрения работодателей и образовательных организаций.

Изучены подходы к моделированию ИТ-специалиста с использованием текстов вакансий с различных сайтов-агрегаторов. Указанный метод является эффективным для решения поставленных в работе задач, но при этом существующие подходы требуют доработки с точки зрения автоматизации сбора и обработки больших объемов данных. Также представленные разработки не учитывают специфику предметной области и не описывают методы обогащения данных, полученных с сайтов-агрегаторов вакансий, для выстраивания системы зависимостей ключевых понятий тех или иных компетенций.

Проанализированы методы, основанные на привлечении экспертов со стороны работодателей к формированию учебных планов. Экспертное знание используется для онтологического моделирования предметной области, далее в разработанную онтологию встраиваются изучаемые дисциплины. Подход обеспечивает высокую степень взаимной интеграции понятий, изучаемых в рамках дисциплин учебного плана, но не является универсальным и требует временных затрат.

Рассмотрены подходы к построению модели ИТ-специалиста с использованием официальных нормативных документов, таких, как ФГОС и профессиональные стандарты. Обобщенное описание трудовых функций специалиста не позволяет рассматривать данные документы как непосредственный источник требований к специалисту, но позволяет выявить контекст тех или иных ключевых понятий. Исходя из этого, нормативные документы могут служить одним из источников обогащения данных о требованиях предприятий и образовательных организаций.

Изучен подход к построению модели специалиста с точки зрения его компетенций. Компетентностный подход эффективен при подборе соискателей на конкретные должности и позволяет индивидуализировать процесс, но требует доработки с точки зрения автоматизации и обобщения при работе с множеством предприятий отрасли, региона, страны.

Вторая глава посвящена разработке метода повышения эффективности выявления требований работодателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов путем разработки информационной системы и

комплекса моделей и алгоритмов сбора и обработки открытых слабоструктурированных данных. Предложена общая концептуальная схема информационной системы, приведенная на рисунке 1. Концепция предполагает:

1. Мониторинг и сбор данных вакансий с сайтов-агрегаторов с заданной периодичностью для выявления требований работодателей;
2. Их последующую обработку для построения информационной модели ИТ-специалиста-соискателя, описывающей ключевые понятия, необходимые для профессиональной деятельности, и удовлетворяющей обобщенным требованиям работодателей;
3. Обогащение полученной модели открытыми данными о предметной области и информацией, содержащейся в нормативных документах;
4. Построение информационной модели специалиста-выпускника, учитывающей связи между ключевыми понятиями (денотатами), путем построения денотатного графа для восстановления иерархии обогащенного множества понятий;
5. Формирование на ее основе набора рекомендаций по реализации образовательных программ.
- 6.

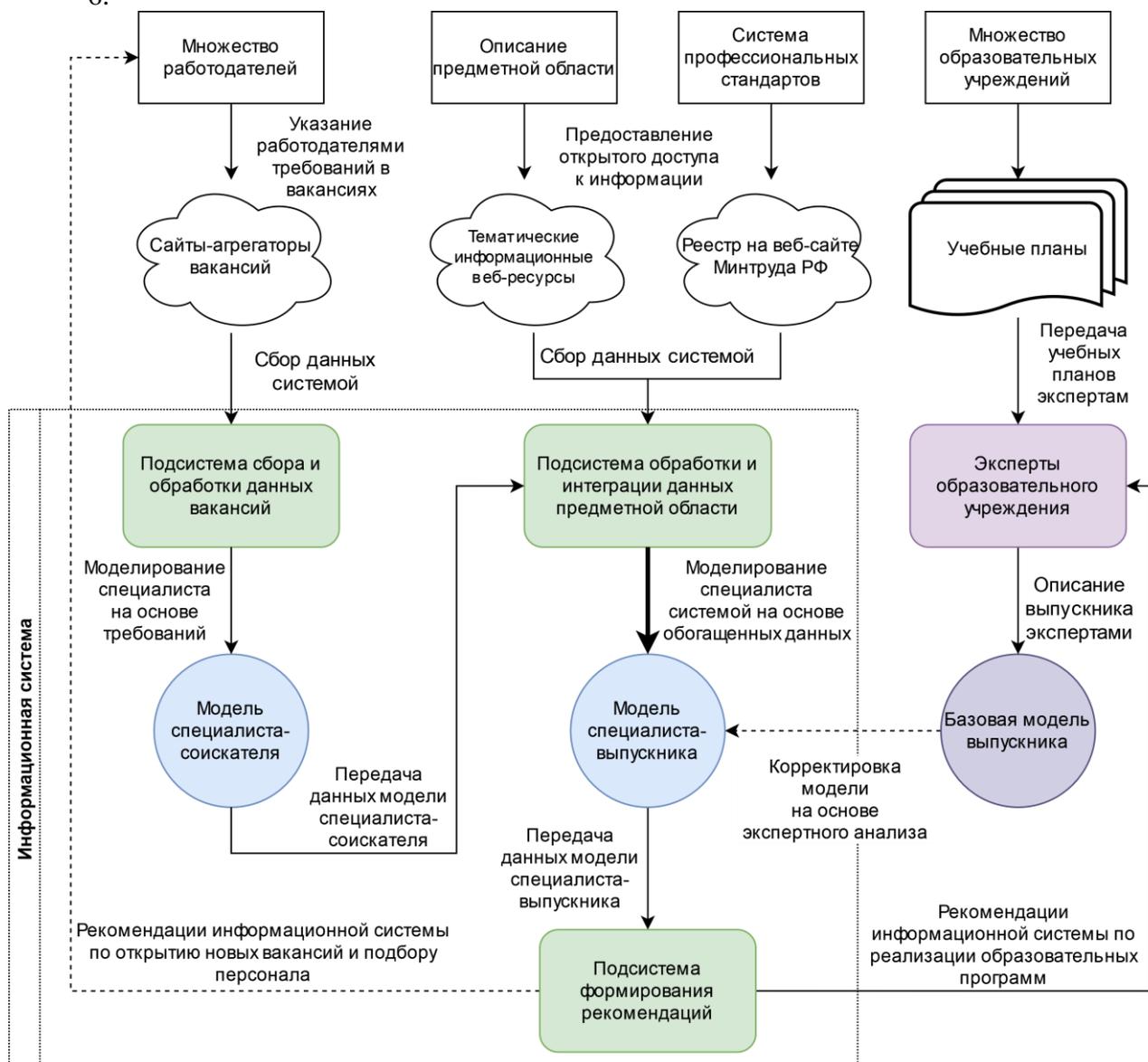


Рисунок 1 — Концептуальная схема выявления динамически меняющихся обобщенных требований работодателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов

При этом множество предприятий может быть ограничено с учетом целей, стоящих перед лицом, принимающим решения. В настоящей работе множество предприятий рассматривается на уровне региона Российской Федерации.

На общем уровне решается задача оптимизации, где необходимо максимизировать сходство между моделью специалиста-соискателя на вакантную должность, отвечающего требованиям множества предприятий (работодателей), и моделью специалиста-выпускника образовательного учреждения:

$$\text{sim}(M_{\text{соиск}}, M_{\text{вып}}) \rightarrow \max \quad (1)$$

Поскольку обе модели используют ключевые понятия в качестве базовых единиц, сходство между ними рассчитывается следующим образом:

$$\text{sim}(M_{\text{соиск}}, M_{\text{вып}}) = \frac{\sum_{i=1}^{|D|} x_i}{|D|}, |D| > 0, \quad (2)$$

$$x_i = \begin{cases} 1, & \text{если } d_i \in S \\ 0, & \text{в противном случае} \end{cases}, \quad (3)$$

где $M_{\text{соиск}}$ — модель специалиста-соискателя, $M_{\text{вып}}$ — модель специалиста-выпускника, D — множество ключевых понятий, входящих в модель специалиста-соискателя, S — множество ключевых понятий модели специалиста-выпускника, x_i — вхождение i -того ключевого понятия, описывающего требование работодателя, в модель специалиста-выпускника, d_i — i -тое ключевое понятие в модели специалиста-соискателя.

Исходные данные для моделей представляют собой тексты на естественном языке. Для их гомогенизации и структуризации разработан алгоритм, адаптирующий методы обработки естественного языка под специфику указанных данных. Алгоритм приведен на рисунке 2.

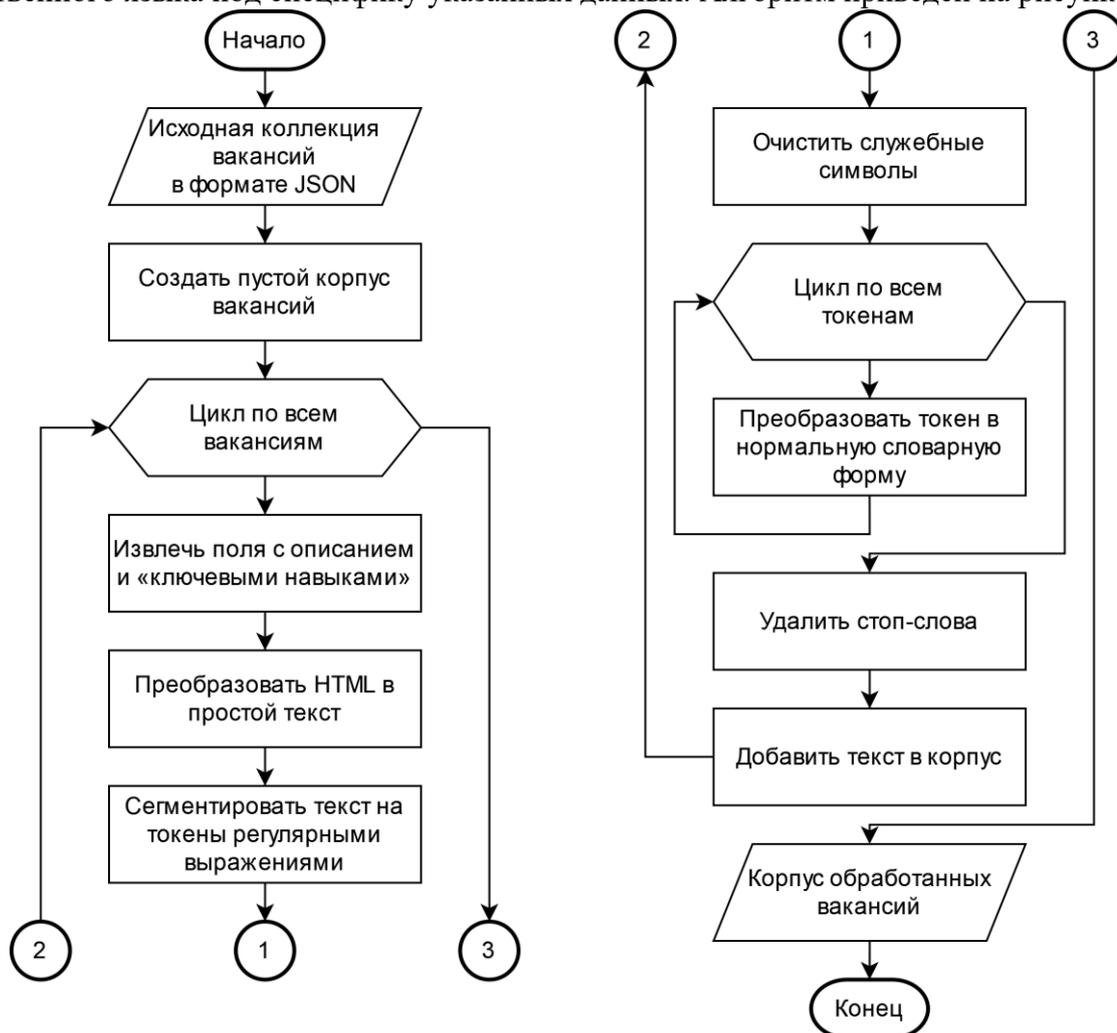


Рисунок 2 — Алгоритм предварительной обработки исходных текстов на естественном языке

Предложена модель специалиста-соискателя, учитывающая комплексную природу предметной области. Модель представляет собой семейство частных моделей, характеризующих направления профессиональной деятельности специалиста в рамках предметной области (например, «Веб-разработка», «Разработка для мобильных устройств», «Разработка баз данных») и объединенных множеством ключевых понятий D :

$$M_{\text{соиск}} = \{M_1^{\text{напр}}, M_2^{\text{напр}}, \dots, M_n^{\text{напр}}, D\} \quad (4)$$

Отмечено, что подмножества ключевых понятий $A_i^{\text{напр}} \subset D$ для каждого из соответствующих направлений профессиональной деятельности пересекаются. При этом ключевые понятия, входящие сразу в несколько подмножеств $A_i^{\text{напр}}$, встречаются с разной частотой в исходных данных для различных направлений профессиональной деятельности, что, в свою очередь, демонстрирует различия в требованиях работодателей относительно значимости освоения компетенций, связанных с указанными ключевыми понятиями.

В связи с этим предлагается подход, при котором каждая частная модель представляет собой нечеткое множество ключевых понятий, описывающее не только набор ключевых понятий направления профессиональной деятельности, но и величину принадлежности понятия данному направлению, что позволяет оценить значимость каждого ключевого понятия для специалиста данного направления:

$$M_i^{\text{напр}} = \{ \langle d, \mu_{M_i^{\text{напр}}}(d) \rangle \mid d \in D \}, \quad (5)$$

где D — множество ключевых понятий модели специалиста-соискателя; $\mu_{M_i^{\text{напр}}}(d)$ — функция принадлежности ключевого понятия d i -тому направлению профессиональной деятельности.

Направления профессиональной деятельности выявляются путем кластерного анализа корпуса вакансий, их число, в свою очередь, определяется предварительной кластеризацией множества ключевых понятий D . Кластеризация множества ключевых понятий осуществлялась методом распространения близости; для корпуса вакансий применялся алгоритм спектральной кластеризации. Качество кластеризации оценивалось при помощи интегральной оценки на основе коэффициента силуэта, индекса Калински-Харабаша и индекса Дэвиса-Болдина.

Для векторного представления ключевых понятий и вакансий на естественном языке используется сглаженная метрика TF-IDF:

$$TfIdf(d) = \left(\frac{f(d)}{k} \right) \times \ln \left(\left(\frac{1+n}{1+df(d)} \right) + 1 \right), \quad (6)$$

$$df(d) = \sum_{j=1}^n x_j, \quad (7)$$

$$x_j = \begin{cases} 1, & \text{если } d \in t_j \\ 0, & \text{в противном случае} \end{cases}, \quad (8)$$

где f — количество вхождений понятия d в тексте; k — общее количество понятий в тексте; n — общее количество текстов в корпусе, df — «document frequency», количество текстов в корпусе, в которых понятие d встречается минимум 1 раз; t_j — j -тый текст в корпусе. Полученные векторы нормализовались.

Размерность вектора в случае с вектором для ключевых понятий определяется мощностью корпуса вакансий; в случае с векторным представлением вакансий — мощностью множества ключевых понятий D .

Функция принадлежности ключевого понятия $d \in D$ i -тому профессиональному направлению использует частотную характеристику встречаемости понятия d в исходных данных i -того кластера вакансий, опираясь на используемый в (6) метод расчета переменной df , поскольку исходные данные также являются данными на естественном языке.

Функция рассчитывается следующим образом:

$$\mu_{M_i^{\text{напр}}}(d) = \frac{\sum_{j=1}^{|V_i|} x_{ij}}{|V_i|}, \quad (9)$$

$$x_{ij} = \begin{cases} 1, & \text{если } d \in v_{ij} \\ 0, & \text{в противном случае} \end{cases}, \quad (10)$$

где V_i — множество вакансий, входящих в i -тый кластер; v_{ij} — текст j -той вакансии в i -том множестве вакансий; x_{ij} — переменная хотя бы единственного вхождения ключевого понятия d в текст j -той вакансии в i -том множестве вакансий.

Автоматизированное построение модели обеспечивает разработанный алгоритм, приведенный на рисунке 3:



Рисунок 3 — Алгоритм автоматизированного моделирования специалиста-соискателя

Фрагмент модели приведен на рисунке 4:

(PHP, 0.884); (Git, 0.839); (MySQL, 0.670); (Объектно-ориентированное программирование, 0.589); (PostgreSQL, 0.536); (Laravel, 0.491); (JavaScript, 0.411); (API, 0.393); (SQL, 0.312); (REST, 0.312); (Проектирование, 0.304); (Docker, 0.295); (Symfony, 0.268); (Базы данных, 0.259); (CSS, 0.250); (Английский язык, 0.232); (Linux, 0.223); (HTML, 0.214); (Redis, 0.179); (Тестирование, 0.179); (REST API, 0.170); (AngularJS, 0.170); (MVC, 0.161); (Yii, 0.161); (React, 0.161); (MongoDB, 0.161)

(Мобильное приложение, 0.687); (Git, 0.626); (Android, 0.596); (iOS, 0.556); (Проектирование, 0.495); (Kotlin, 0.485); (Swift, 0.424); (Объектно-ориентированное программирование, 0.414); (Java, 0.394); (MVVM, 0.343); (Паттерны проектирования, 0.313); (API, 0.293); (REST, 0.283); (Android SDK, 0.283); (SOLID, 0.273); (UI, 0.242); (MVP, 0.242); (HTTP, 0.212); (Clean Architecture, 0.212); (CI, 0.212); (RXJAVA, 0.212); (Английский язык, 0.182); (Objective-C, 0.182); (Retrofit, 0.182); (GITHUB, 0.162); (REST API, 0.152); (FireBase, 0.152); (Алгоритмы, 0.152); (Room, 0.152)

Рисунок 4 — Фрагмент линейной модели ИТ-специалиста-соискателя для предприятий Пермского края

Полученная модель специалиста-соискателя отражает набор ключевых понятий, необходимых работодателям в тех или иных направлениях деятельности, но не учитывает связи между понятиями и не включает в себя «скрытые» понятия — понятия, описывающие те компетенции, освоение которых необходимо для изучения иных компетенций, но которые не были включены работодателями в качестве требований к вакансиям (например, ключевое понятие «git» в приведенном на рис. 4 фрагменте модели требуется во всех направлениях профессиональной деятельности, но ни одно из них не содержит требования «системы контроля версий», которое является родовым понятием для «git» и необходимо для его изучения).

Предложенная модель специалиста-выпускника расширяет и обогащает модель специалиста-соискателя, сохраняя при этом общую основу (множество ключевых понятий), что позволяет напрямую оценивать сходство моделей, в том числе после возможной экспертной корректировки модели специалиста-выпускника.

Модель специалиста-выпускника включает в себя множество частных моделей $G_i^{\text{напр}}$ и множество ключевых понятий предметной области S , при этом $D \subset S$, где D — множество ключевых понятий модели специалиста-соискателя, как указано в (4).

$$M_{\text{вып}} = \{G_1^{\text{напр}}, G_2^{\text{напр}}, \dots, G_n^{\text{напр}}, S\} \quad (11)$$

Каждая из частных моделей $G_i^{\text{напр}}$ является ориентированным графом:

$$G_i^{\text{напр}} = (V, E), \quad (12)$$

где V — множество вершин графа, E — множество дуг графа. Количество частных моделей $G_i^{\text{напр}}$ совпадает с количеством частных моделей, описанных в (5), для каждой модели $M_i^{\text{напр}}$ производится обогащение множества ключевых понятий и строится граф. Алгоритм построения модели приведен на рисунке 5:

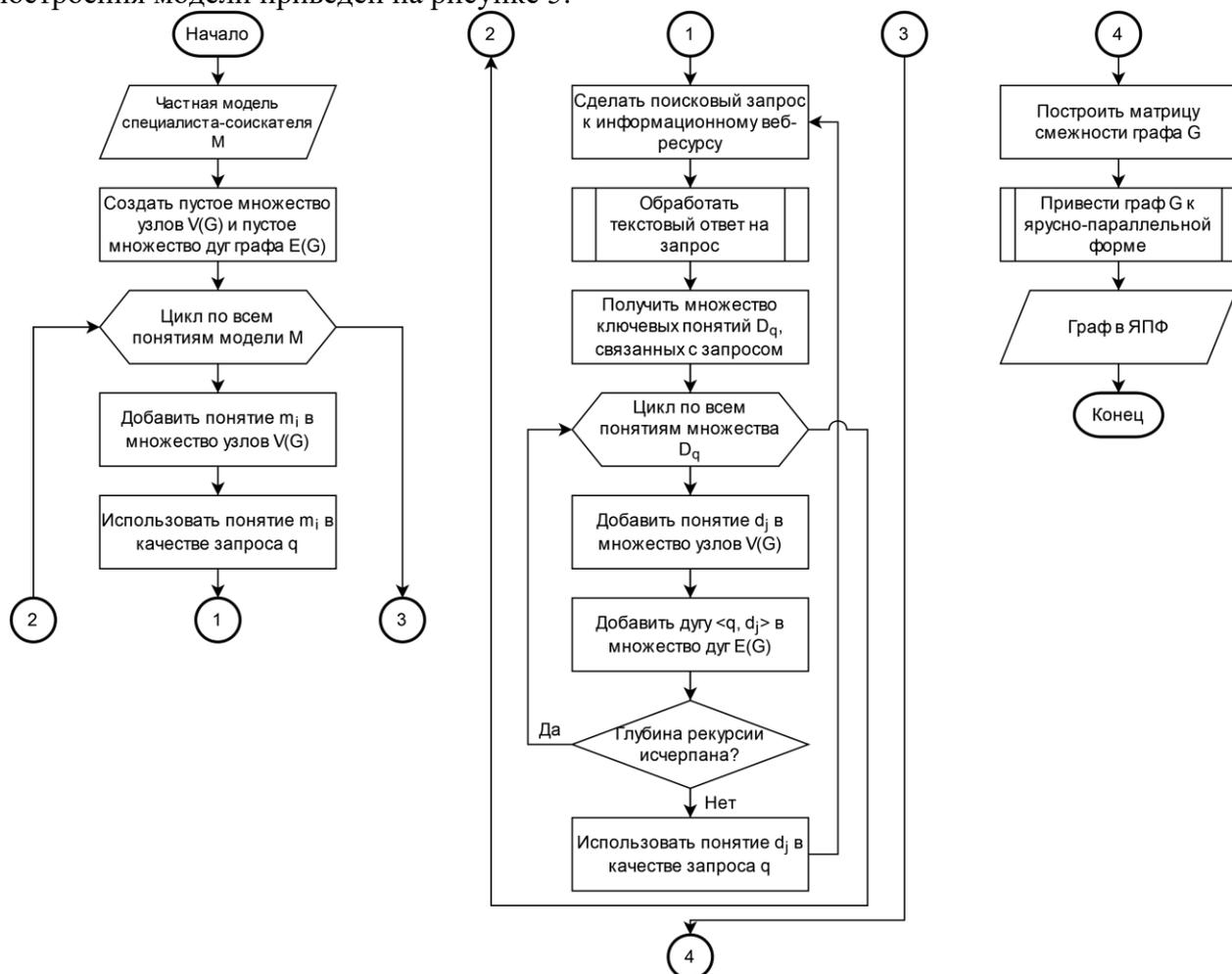


Рисунок 5 — Алгоритм построения частной модели специалиста-выпускника с обогащением данных

Итоговые графы модели специалиста-выпускника сохраняются в формате DOT, фрагмент графа приведен на рисунке 6.

```

"json" -> "javascript" [weight=97 penwidth=3.43 label=9.7]
"go" -> "sql" [weight=10 penwidth=1.00 label=1.0]
"net" -> "java" [weight=10 penwidth=1.00 label=1.0]
"svn" -> "sqlite" [weight=10 penwidth=1.00 label=1.0]
"css" -> "xml" [weight=79 penwidth=2.98 label=7.9]
    
```

Рисунок 6 — Фрагмент графа модели специалиста-выпускника, описанного на языке DOT

Для формирования рекомендаций разработан алгоритм, принимающий на вход целевое понятие $S_{\text{цел}}$ и модель специалиста-выпускника $M_{\text{вып}}$. Разработанный алгоритм приведен на рисунке 7.

На выходе алгоритма — линейно упорядоченное множество рекомендованных к изучению понятий $S_{\text{рек}}$ с расчетным количеством часов. Полученное множество передается экспертам образовательной организации и лицу, принимающему решения, как набор рекомендаций относительно составления и корректировки образовательных программ.

Указанные рекомендации как отражают требования работодателей, так и указывают на горизонтальную и вертикальную интеграцию ключевых понятий в рамках предметной области, что дает возможность эффективного внесения изменений не только в рамках отдельных дисциплин, но и учебных планов в целом.

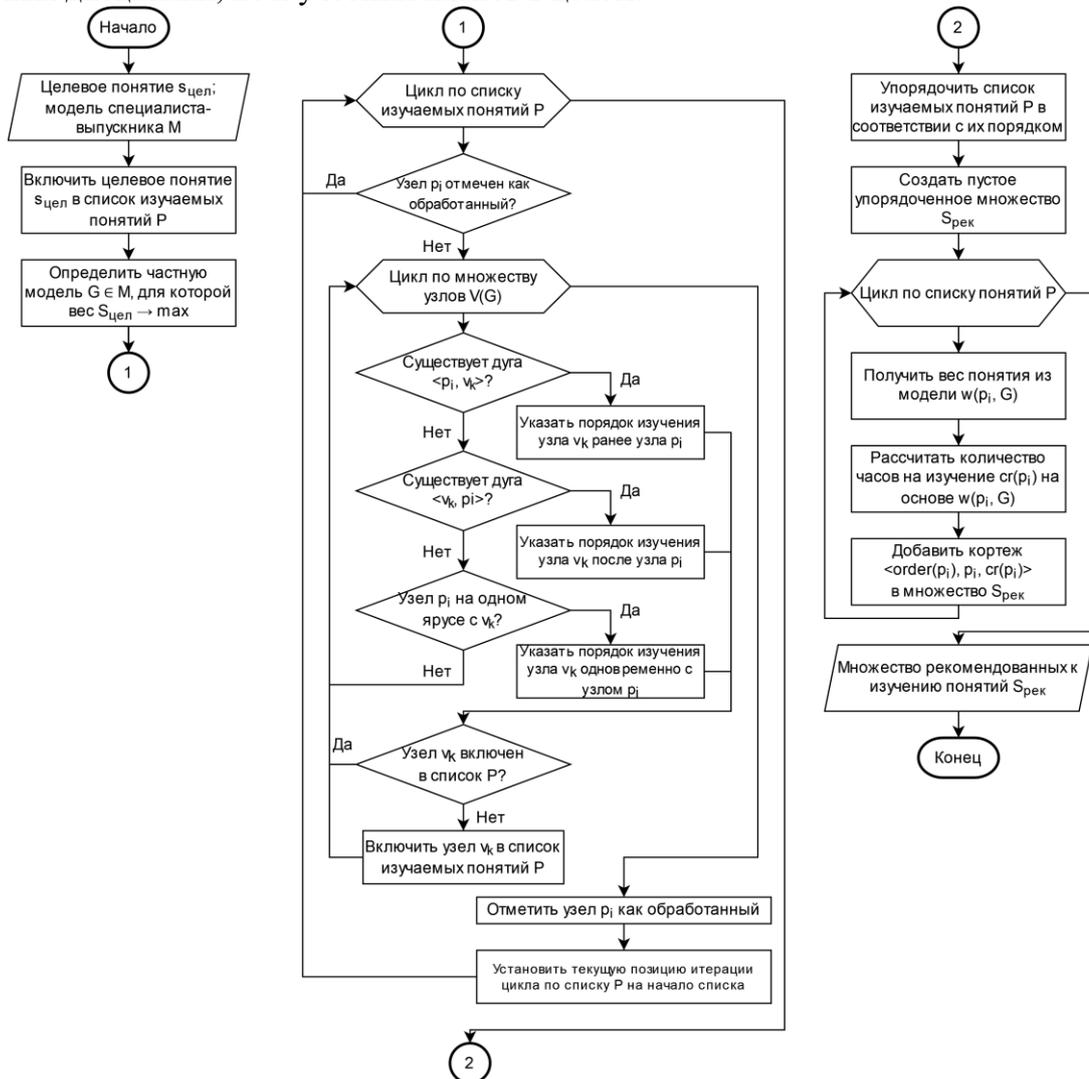


Рисунок 7 — Алгоритм формирования рекомендаций по включению в образовательные программы востребованных ключевых понятий

Третья глава посвящена программной реализации разработанных моделей и алгоритмов в информационной системе, обеспечивающей выявление обобщенных требований работодателей к компетенциям соискателей и формирование рекомендаций по реализации образовательных программ подготовки ИТ-специалистов с учетом требований.

Разработанная информационная система включает 5 подсистем, структурная схема приведена на рисунке 8.

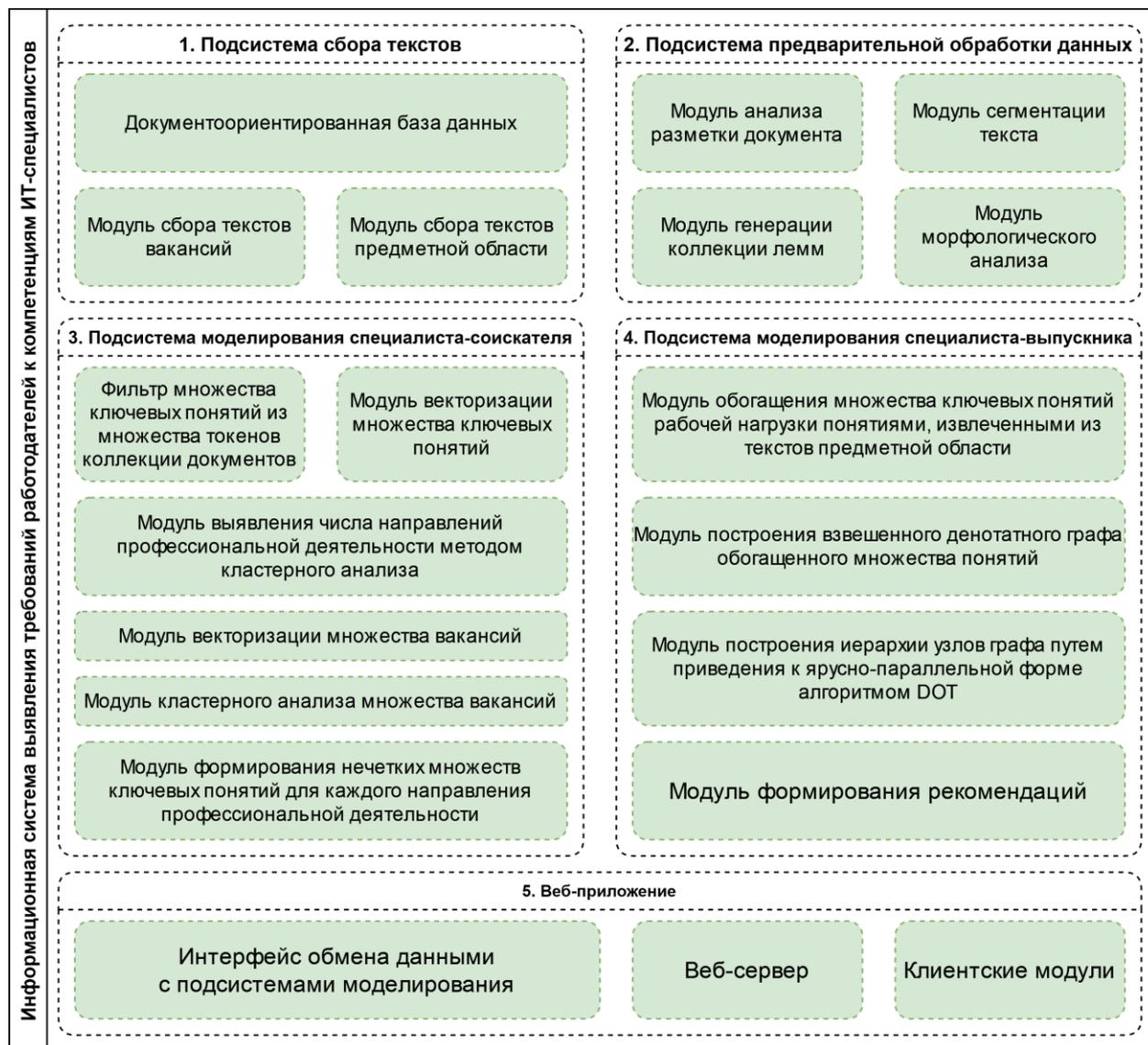


Рисунок 8 — Структурная схема информационной системы

Реализация серверной части осуществлена на языке программирования Python версии 3.9. Схема организации программных модулей приведена на рисунке 9.

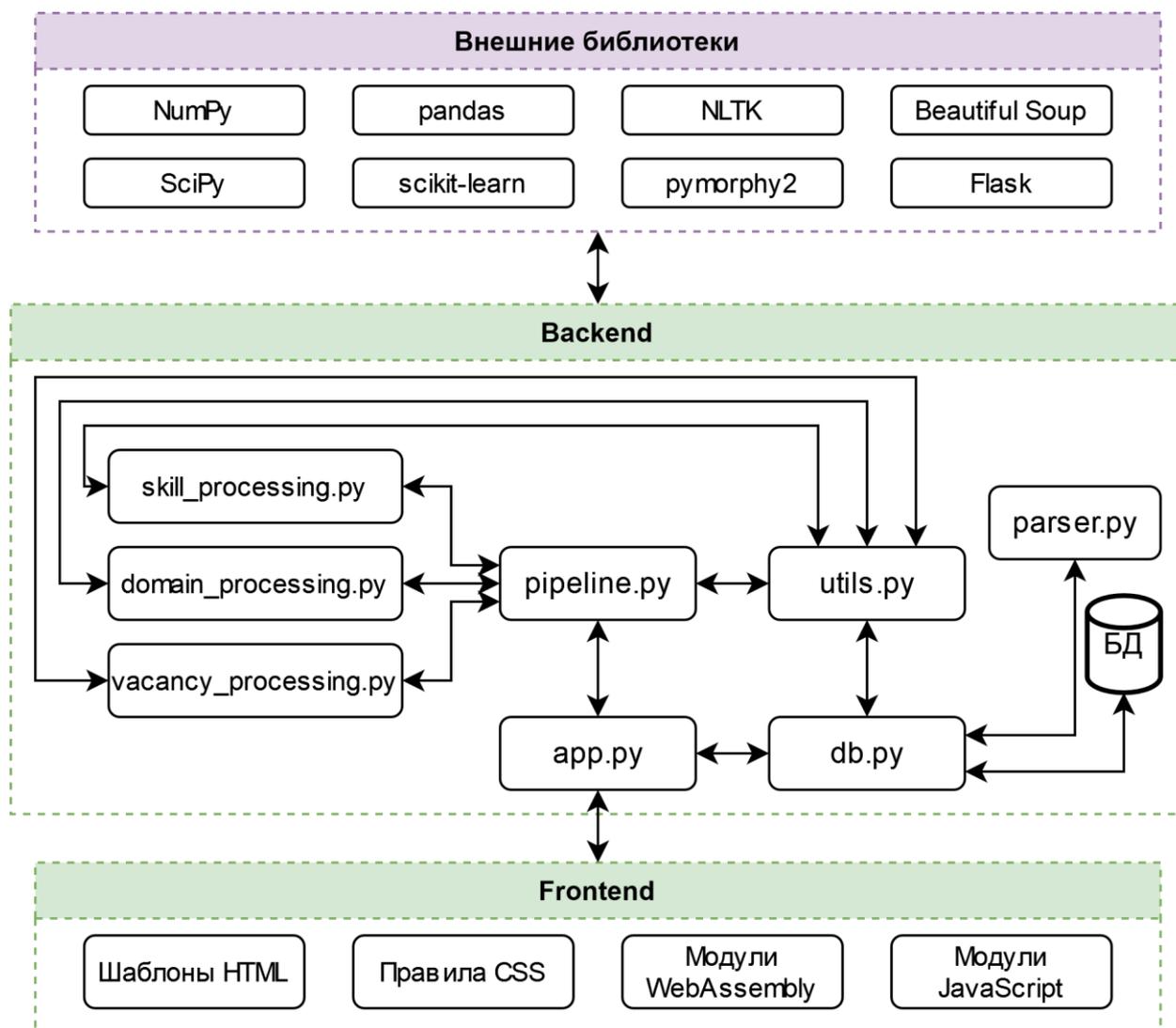


Рисунок 9 — Схема организации программных модулей информационной системы

Для ряда вспомогательных методов и алгоритмов используются готовые реализации из внешних модулей и библиотек (scikit-learn, SciPy, pymorphy2, NLTK). Веб-приложение реализовано при помощи фреймворка Flask и веб-сервера nginx. Описание функционала разработанных программных модулей приведено в таблице 1.

Таблица 1 — Описание функционала программных модулей

Модуль	Функционал
app.py	Запуск сервера веб-приложения; Обработка HTTP-запросов со стороны клиента (frontend) Формирование и вывод рекомендаций
pipeline.py	Предоставление интерфейса обмена данными с моделями для сервера веб-приложения; Запуск моделирования специалиста по заданным входным параметрам
db.py	Организация ввода/вывода для базы данных
utils.py	Анализ разметки документов; Сегментация (токенизация) документов; Морфологический анализ документов; Генерация коллекции лемм по заданным критериям; Фильтрация ключевых понятий из множества лемм коллекции

Модуль	Функционал
parser.py	Ежемесячный сбор данных о вакансиях с сайтов-агрегаторов; Организация и сохранение полученных данных в формате JSON
skill_processing.py	Векторизация ключевых понятий; Выявление числа стабильных групп ключевых понятий при помощи кластеризации алгоритмом распространения близости
vacancy_processing.py	Векторизация множества вакансий; Группировка вакансий алгоритмом спектральной кластеризации; Формирование нечетких множеств ключевых понятий
domain_processing.py	Сбор и кэширование текстов о предметной области с информационных веб-ресурсов; Обогащение заданного множества ключевых понятий понятиями, извлеченными из текстов предметной области; Построение взвешенного денотатного графа обогащенного множества понятий; Установление иерархии узлов графа через приведение к ярусно-параллельной форме алгоритмом DOT

Логика клиентской части реализована на языке программирования JavaScript. Инструменты визуализации графов используют предварительно скомпилированные C++-библиотеки в формате модулей WebAssembly. Пользовательский интерфейс реализован шаблонами Jinja2 и средствами веб-разметки — HTML и CSS.

Подсистема сбора текстов реализована и запущена автономно в феврале 2020 г. с целью ежемесячного мониторинга и сбора текстов вакансий по регионам Российской Федерации; собранный корпус насчитывает 520 615 вакансий.

В четвертой главе описан эксперимент по использованию разработанной концепции на примере краткосрочных интенсивных курсов по развитию компетенций студентов в интересах ИТ-компаний г. Перми.

Апробация проведена с фокус-группой студентов Пермского национального исследовательского политехнического университета (ПНИПУ) третьего курса бакалавриата направлений «Программная инженерия» и «Информатика и вычислительная техника». Для оценки результатов со стороны работодателей в апробации участвовали эксперты пермских ИТ-компаний «Датабриз» и «Иннфокус».

Для формирования рекомендаций в качестве целевого понятия определен язык программирования Java. Обобщенные требования работодателей относительно целевого понятия выявлены посредством моделирования специалиста-соискателя для Пермского края за период с ноября 2021 г. по апрель 2022 г, как указано на рисунке 10. Модель специалиста-соискателя в Пермском крае содержит 12 направлений профессиональной деятельности и 127 уникальных ключевых понятий. Частная модель направления «Java-разработчик» включает 34 ключевых понятия.

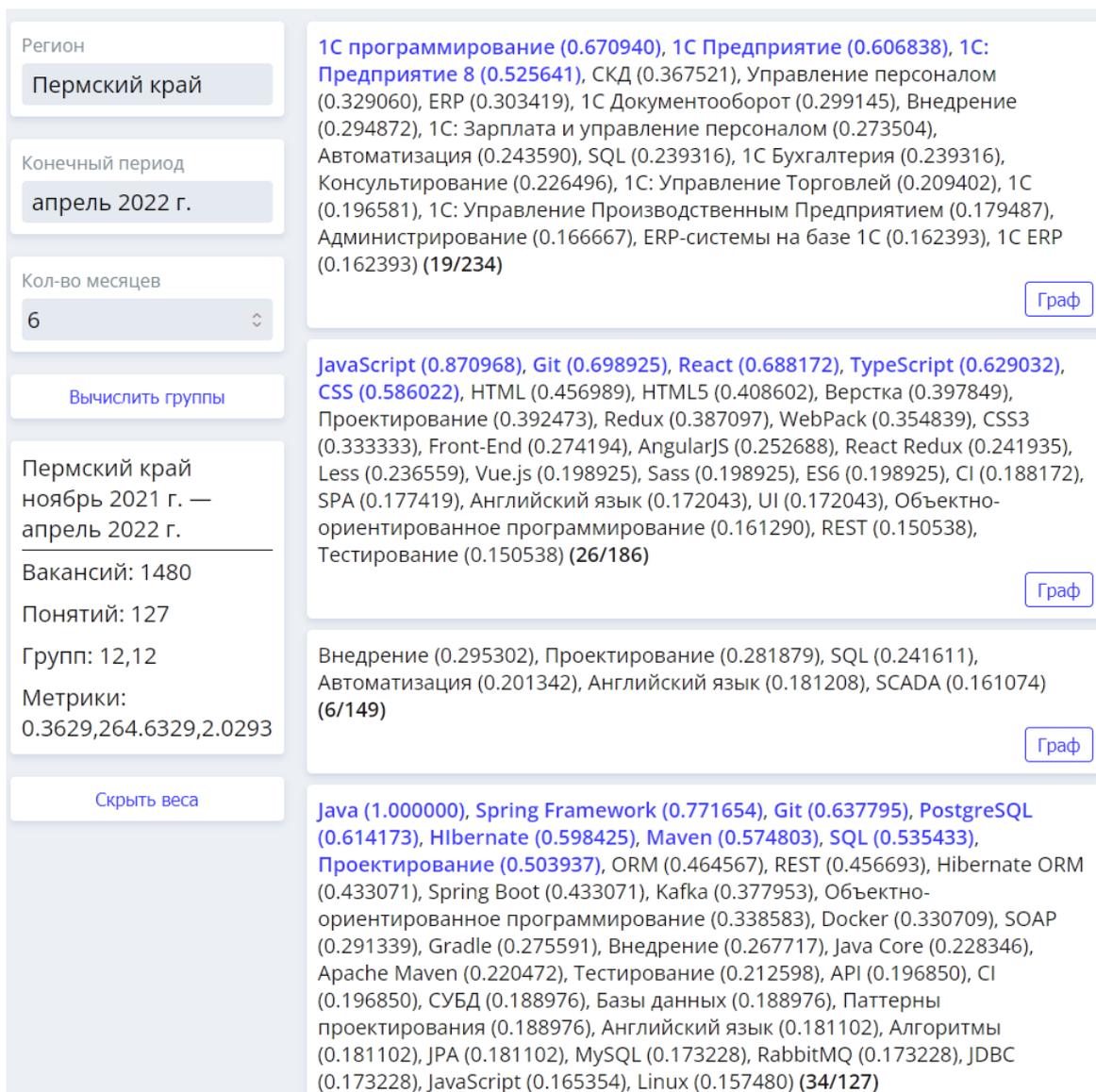


Рисунок 10 — Фрагмент модели специалиста-соискателя для Пермского края на основе данных в период с ноября 2021 г. по апрель 2022 г.

В интерфейсе веб-приложения выполнен подбор и ранжированы частные модели специалиста-соискателя для Пермского края за период с ноября 2021 г. по апрель 2022 г. в соответствии с наибольшим весом целевого понятия. Ранжирование продемонстрировано на рисунке 11, при этом целевое понятие выделено зеленым цветом; ключевые понятия, имеющие вес $w \geq 0,75$, выделены красным цветом; понятия, имеющие вес $w \geq 0,5$, выделены синим цветом.

Из общей модели специалиста-соискателя выделена частная модель с наибольшим весом целевого понятия. Для указанной частной модели соискателя построена модель специалиста-выпускника, произведено обогащение до 129 понятий (фрагмент графа модели приведен на рисунке 14; ключевые понятия, включенные в соответствующую модель специалиста-соискателя, обозначены зеленым цветом), построено 493 связи.

На основе полученной модели сформированы рекомендации по составлению программы двухнедельного интенсивного курса для студентов. Фрагмент содержания образовательной программы интенсива, учитывающей данные рекомендации, приведен на рисунке 13 в виде графа ключевых понятий. В рамках курса изучение «Spring Framework», предшествующее ему изучение «Hadoop»; также предложено горизонтально интегрировать изучение «PostgreSQL», «Oracle Database»; отмечена необходимость изучения «Oracle PL/SQL» в рамках подготовки к освоению по «Oracle Database».

Преподавателями кафедры ИТАС ПНИПУ разработана программа двухнедельного интенсива с учетом предложенных рекомендаций. Эксперты ИТ-компаний не были ознакомлены с программой и не влияли на ее содержание.

Регион
Пермский край

Конечный период
апрель 2022 г.

Кол-во месяцев
6

Целевое понятие
Java

[Подобрать рекомендации](#)

Найдено направлений: 5; проанализировано вакансий: 1480

Ключевых понятий: 34; вакансий: 127 Граф

Java (1.00), Spring Framework (0.77), Git (0.64), PostgreSQL (0.61), Hibernate (0.60), Maven (0.57), SQL (0.54), Проектирование (0.50), ORM (0.46), REST (0.46), Hibernate ORM (0.43), Spring Boot (0.43), Kafka (0.38), Объектно-ориентированное программирование (0.34), Docker (0.33), SOAP (0.29), Gradle (0.28), Внедрение (0.27), Java Core (0.23), Apache Maven (0.22), Тестирование (0.21), API (0.20), CI (0.20), СУБД (0.19), Базы данных (0.19), Паттерны проектирования (0.19), Английский язык (0.18), Алгоритмы (0.18), JPA (0.18), MySQL (0.17), RabbitMQ (0.17), JDBC (0.17), JavaScript (0.17), Linux (0.16)

Ключевых понятий: 28; вакансий: 51 Граф

Android (0.98), Kotlin (0.80), Git (0.63), Java (0.63), Android SDK (0.59), Мобильное приложение (0.55), Проектирование (0.43), RXJAVA (0.39), MVVM (0.37), Объектно-ориентированное программирование (0.35), Retrofit (0.35), MVP (0.35), Clean Architecture (0.35), SOLID (0.33), REST (0.27), API (0.27), Английский язык (0.25), iOS (0.25), Room (0.25), FireBase (0.24), Паттерны проектирования (0.24), CI (0.22), HTTP (0.18), UI (0.18), Алгоритмы (0.18), AGILE (0.18), JSON (0.16), C (0.16)

Ключевых понятий: 16; вакансий: 116 Граф

SQL (1.00), Oracle PL/SQL (0.71), Базы данных (0.61), Проектирование (0.59), СУБД (0.42), Автоматизация (0.33), ETL (0.31), PostgreSQL (0.26), Английский язык (0.24), Тестирование (0.23), Git (0.22), Оптимизация запросов (0.22), **Java (0.20),** Внедрение (0.19), BI (0.18), Работа с базами данных (0.16)

Рисунок 11 — Ранжирование частных моделей специалиста-соискателя для целевого понятия «Java», регион Пермский край, период с ноября 2021 г. по апрель 2022 г.

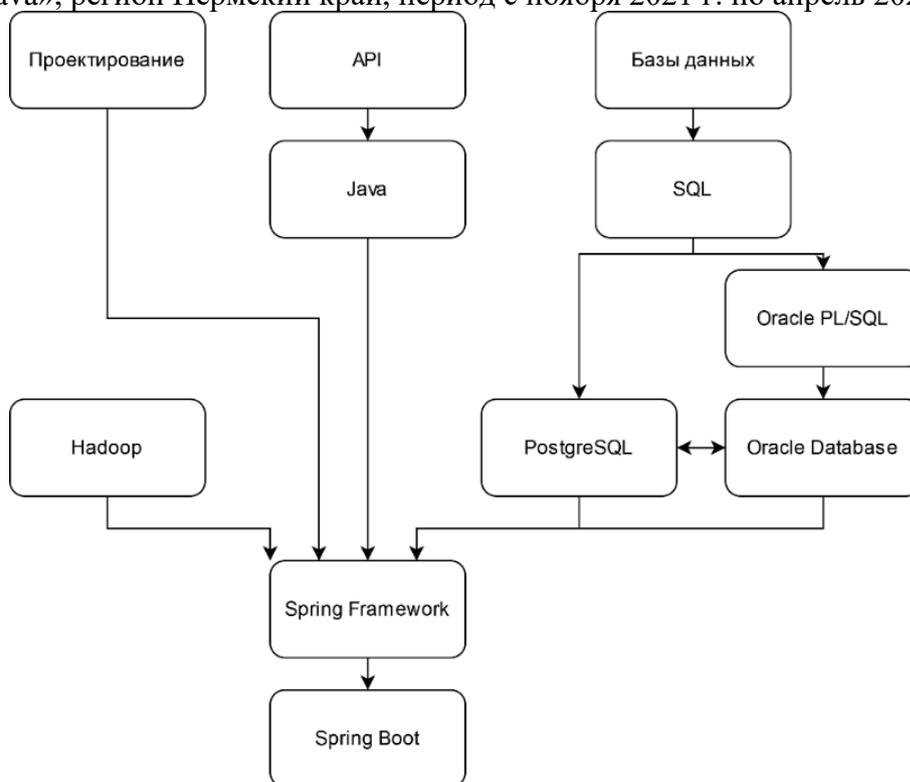


Рисунок 12 — Фрагмент графа ключевых понятий, раскрывающих содержание возможной программы интенсива для подготовки специалиста по направлению «Java-разработчик»

ИТ-экспертами проведены стандартизированные собеседования на квалификацию стажера-разработчика с каждым из студентов и заполнены опросники с указанием оценок студентов по шкале от 1 (низкий уровень) до 5 (наивысший уровень) по следующим критериям:

- уровень общей подготовки — наличие у студента компетенций по основам алгоритмизации и программирования, включая парадигмы программирования; теории алгоритмов и структурам данных;
- уровень владения стеком технологий — наличие у студента компетенций по требуемым компании технологиям, включая языки программирования, фреймворки, прикладное программное обеспечение;
- уровень технического кругозора — наличие у студента понимания взаимосвязей между различными технологиями; умение подбора альтернатив;
- вероятность рекомендации к трудоустройству — оценка экспертом вероятности рекомендации соискателя к трудоустройству на позицию стажера-разработчика в компании эксперта.

Результаты первичных собеседований показали, что, хотя студенты обладают достаточными компетенциями в области алгоритмизации и программирования (средний балл — 3,9375), отсутствие компетенций, связанных с требуемыми технологиями, и понимания их взаимосвязи (средняя оценка владения стеком технологий — 2,24; оценка технического кругозора — 2,375) сильно снижает вероятность рекомендации к трудоустройству (средний балл — 2,54).

После первичных собеседований студенты прошли двухнедельную краткосрочную подготовку по разработанной образовательной программе. По итогам студенты повторно прошли собеседование, при этом эксперты задавали дополнительные вопросы соискателям и вновь заполняли опросники. Список вопросов как для первичных собеседований, так и для повторных составлен и согласован всеми экспертами.

Повторные собеседования показали улучшения показателей по всем критериям (средний общий уровень подготовки вырос на 9%; владение стеком технологий — на 60%; технический кругозор — на 46%). Особенно важно отметить, что вероятность рекомендации трудоустройства выросла на 40%. Сравнение средних оценок по каждому из критериев до и после прохождения курса приведено на рисунке 13.

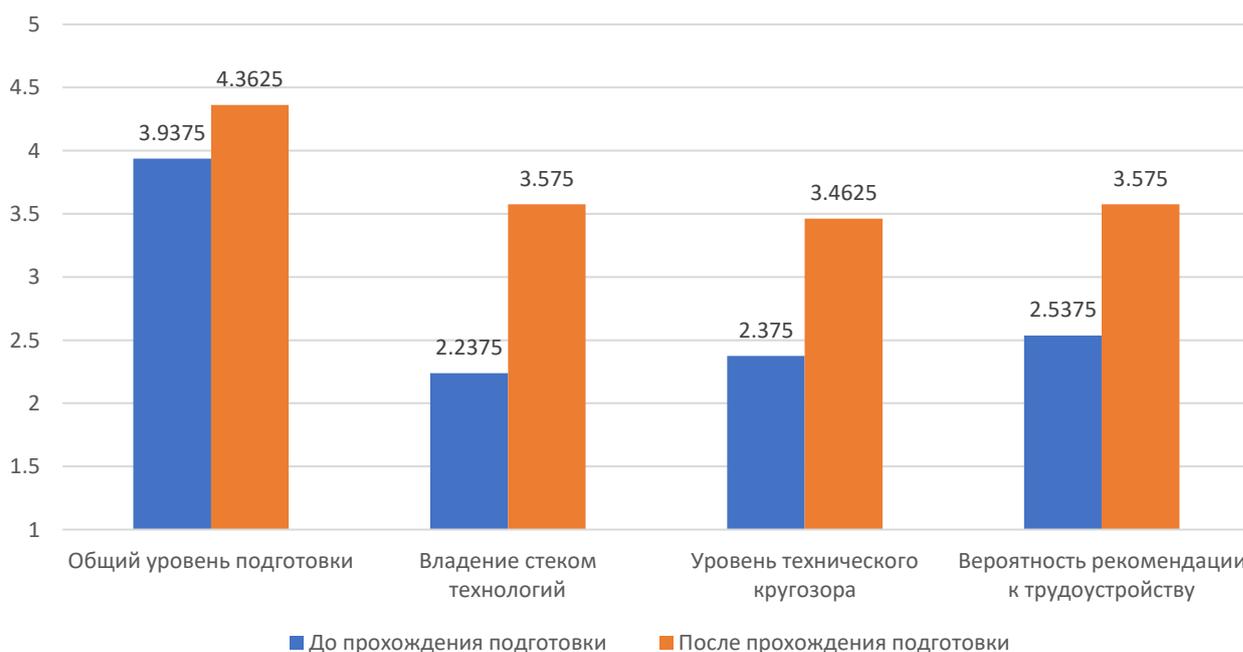


Рисунок 13 — Диаграмма средних оценок экспертов по каждому из критериев до и после прохождения интенсивного курса подготовки.

III. ЗАКЛЮЧЕНИЕ

Диссертационная работа посвящена повышению эффективности выявления обобщенных требований работодателей к компетенциям соискателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов.

В рамках проведенного исследования получены следующие основные результаты:

1. Проанализированы подходы к выявлению требований работодателей в сфере ИТ, изучен процесс подготовки ИТ-специалистов, выявлены факторы, приводящие к «кадровому голоду» в индустрии, показана важность выявления требований работодателей для подготовки кадров, обладающих компетенциями, необходимыми для экономики;

2. Предложена концепция информационной системы выявления обобщенных требований работодателей и формирования рекомендаций по реализации образовательных программ подготовки ИТ-специалистов на основе анализа открытых слабоструктурированных данных с использованием денотативного подхода и методик обогащения данных;

3. Предложена методика и разработан алгоритм автоматизированного мониторинга, сбора и предварительной обработки открытых слабоструктурированных данных о требованиях работодателей, что позволило сформировать базу данных из 520 615 вакансий ИТ-специалистов, охватывающих период в 33 месяца;

4. Предложена информационная модель специалиста-соискателя, основанная на открытых слабоструктурированных данных о ключевых понятиях профессиональной деятельности, и разработан алгоритм ее автоматизированного построения, что позволило выявлять обобщенные требования множества работодателей к ИТ-специалистам посредством одновременного анализа до 20 000 вакансий в пределах одного региона РФ;

5. Предложена модель специалиста-выпускника, использующая денотатный граф для выстраивания горизонтальных и вертикальных связей между ключевыми понятиями предметной области и обогащения исходных данных рынка труда, и разработан алгоритм ее автоматизированного построения, что позволило обеспечить обогащение исходной группы понятий в 3,8 раз и взаимную интеграцию 129 уникальных понятий в рамках одного направления профессиональной деятельности с построением более 490 связей между ними;

6. Разработан алгоритм автоматизированного формирования рекомендаций по реализации образовательных программ, учитывающий как обобщенные требования работодателей, так и иерархию понятий предметной области ИТ;

7. Разработано программное обеспечение информационной системы, обеспечивающее выявление обобщенных требований работодателей к компетенциям соискателей и формирование рекомендаций по реализации образовательных программ на основе предложенных моделей и алгоритмов. Программное обеспечение апробировано на базе ПНИПУ и ИТ-компаний г. Перми.

Результаты исследования использованы на краткосрочных курсах дополнительного образования с учетом интересов ИТ-компаний. Экспертами ИТ-компаний отмечено, что выпускники краткосрочных образовательных программ, реализованных с учетом рекомендаций информационной системы, на 25–40% более соответствуют требованиям ИТ-рынка г. Перми.

Разработанная информационная система использована в ходе обновления программ бакалавриата и магистратуры по направлениям «Информатика и вычислительная техника», «Программная инженерия» на кафедре ИТАС ПНИПУ. Экспертами кафедры отмечено, что предложенный подход позволил существенно сократить временные и трудовые затраты, связанные с внесением изменений в учебные планы и рабочие программы дисциплин с учетом потребностей работодателей. Информационная система обеспечила адаптируемость образовательных программ к заданной предметной области (направлению подготовки) за счет обогащения данных; ускорила и упростила процесс формирования рекомендаций по корректировке учебных планов.

Разработанное программное обеспечение позволило с заданной периодичностью осуществлять мониторинг обобщенных требований к специалисту. Модели и методы позволили также получить рекомендации для формирования программ в системе дополнительного образования и иных видов профессионального образования. По итогам внедрения получено положительное заключение кафедры ИТАС ПНИПУ.

IV. ОСНОВНЫЕ ПУБЛИКАЦИИ

Публикации в рецензируемых научных изданиях, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук

1. **Яруллин Д. В.** Интеллектуальная система управления подготовкой ИТ-специалистов на основе денотативной аналитики / Д. В. Яруллин. – DOI 10.15593/2499-9873/2022.3.08. – Текст : непосредственный // Прикладная математика и вопросы управления / Applied Mathematics and Control Sciences. – 2022. – № 3. – С. 141–164.

Публикации в изданиях, индексируемых в международных реферативных базах и системах цитирования

2. **Yarullin D. V.,** Faizrakhmanov R. A., Fominykh P. Y. A Vector Representation of the IT Specialists' Set of Workloads for HR Decision Support System / D. V. Yarullin, R. A. Faizrakhmanov, P. Y. Fominykh. – DOI 10.1007/978-3-030-95112-2_31. – Текст : непосредственный // Society 5.0: Human-Centered Society Challenges and Solutions. – Cham : Springer Verlag, 2022. – С. 383–394.

3. **Yarullin D. V.,** Faizrakhmanov R. A., Fominykh P. Y. Automation of Demand Planning for IT Specialists Based on Ontological Modelling / D. V. Yarullin, R. A. Faizrakhmanov, P. Y. Fominykh. – DOI 10.1007/978-3-030-63563-3_4. – Текст : непосредственный // Society 5.0: Cyberspace for Advanced Human-Centered Society. – Cham : Springer Verlag, 2021. – С. 35–45.

4. **Yarullin D. V.,** Fayzrakhmanov R. A., Fominykh P. Y. A Web-Data Driven HR Decision Support System Based on the IT Specialist Vector Model / D. V. Yarullin, R. A. Faizrakhmanov, P. Y. Fominykh. – DOI 10.1109/SCM52931.2021.9507144. – Текст : непосредственный // Proceedings of 2021 24th International Conference on Soft Computing and Measurements, SCM 2021 : 24, St. Petersburg, 26–28 мая 2021 года. – St. Petersburg, 2021. – С. 225–228.

5. **Yarullin D. V.,** Faizrakhmanov R. A. Web-data driven ontological approach to modelling IT specialists recruitment needs / D. V. Yarullin, R. A. Faizrakhmanov. – DOI 10.1109/SCM.2019.8903715. – Текст : непосредственный // Proceedings of 2019 22nd International Conference on Soft Computing and Measurements, SCM 2019 : 22, St. Petersburg, 23–25 мая 2019 года. – St. Petersburg, 2019. – С. 252–255.

Свидетельства о государственной регистрации программ для ЭВМ

6. Свидетельство о государственной регистрации программы для ЭВМ № 2022665489 Российская Федерация. Денограф. Интеллектуальная система денотативной аналитики для управления подготовкой ИТ-кадров / **Яруллин Д.В.**; заявка 2022663900; поступл. 19.07.2022, опублик. 17.08.2022, бюл. № 8. — 1 с.

Публикации в прочих изданиях

7. **Яруллин Д. В.**, Файзрахманов Р. А., Фоминых П. Ю. Векторное представление множества рабочих нагрузок ИТ-специалистов для системы поддержки принятия решений в сфере кадрового обеспечения / Д. В. Яруллин, Р. А. Файзрахманов, П. Ю. Фоминых. – DOI 10.52348/2712-8873_ММТТ_2021_7_41. – Текст : непосредственный // Математические методы в технологиях и технике. – 2021. – № 7. – С. 41–45.

8. **Яруллин Д.В.**, Файзрахманов Р.А., Фоминых П.Ю. Система поддержки принятия кадровых решений на основе векторной модели ИТ-специалиста / Д. В. Яруллин, Р. А. Файзрахманов, П. Ю. Фоминых. – Текст : непосредственный // Международная конференция по мягким вычислениям и измерениям. – 2021. – Т. 1. – С. 308–311.

9. **Яруллин Д.В.**, Файзрахманов Р.А., Фоминых П.Ю. Автоматизация планирования потребности в ИТ-специалистах на основе онтологического моделирования / Д. В. Яруллин, Р. А. Файзрахманов, П. Ю. Фоминых. – Текст : непосредственный // Математические методы в технике и технологиях - ММТТ. – 2020. – Т. 8. – С. 67–71.

10. **Яруллин Д.В.**, Файзрахманов Р.А. Построение модели ИТ-специалиста на основе нечеткой кластеризации для системы поддержки принятия решений в сфере кадрового обеспечения / Д. В. Яруллин, Р. А. Файзрахманов. – Текст : непосредственный // Международная конференция по мягким вычислениям и измерениям. – 2020. – Т. 1. – С. 250–254.

11. **Яруллин Д.В.**, Файзрахманов Р.А. Онтологическое моделирование потребностей в ИТ-специалистах на основе веб-данных / Д. В. Яруллин, Р. А. Файзрахманов. – Текст : непосредственный // Международная конференция по мягким вычислениям и измерениям. – 2019. – Т. 1. – С. 359–362.

Подписано в печать ____ .12.2022. Формат 60×84/16.
Усл. печ. л. 1,28. Тираж 100 экз. Заказ № ____/2022.

Отпечатано с готового оригинал-макета
в типографии издательства Пермского национального
исследовательского политехнического университета
614990, г. Пермь, Комсомольский пр., 29, к. 113.