

СИСТЕМНЫЙ АНАЛИЗ И УПРАВЛЕНИЕ SYSTEM ANALYSIS AND MANAGEMENT

УДК 504.53.052

DOI 10.52575/2687-0932-2025-52-4-928-945

EDN TMEYRY

Методика извлечения типовых требований из описания вакансий методами графо-семантического анализа

¹Логинов И.В., ²Щербин А.А.

¹ Академия ФСО России,

Россия, 302020, г. Орел, ул. Приборостроительная, д. 35

² Среднерусский институт управления – филиал РАНХиГС,

Россия, 302028, г. Орел, бульвар Победы, д. 5а

loginov_iv@bk.ru, alex.sherbin1@mail.ru

Аннотация. В работе рассмотрена проблема извлечения и обобщения информации о требованиях к компетенциям специалистов на основе анализа наборов текстовых описаний вакансий. Цель исследования – повышение обоснованности обобщения и извлечения данных с учетом дополнительной информации об уровне требований. Достижение цели исследования обеспечивается за счет применения подхода к анализу текстовых данных на основе объединения графо-семантического анализа с нечеткой лингвистической моделью уровня «желательности». На основе указанного подхода предложена формальная модель описания вакансий и методика извлечения обобщенных требований к профессиональным навыкам из текстовых описаний вакансий специалистов. Результаты применения предложенного подхода к корпусу вакансий для профессий бизнес-аналитика, разработчика и DevOps-инженера с использованием разработанного программного средства «ВекторКогнитив» показали возможность комбинации нечеткой лингвистической модели и графо-семантического анализа для формирования обобщенных требований к специалистам.

Ключевые слова: семантический анализ, графы, вакансии, обобщение, обработка естественного языка, лингвистические выражения, нечеткие множества

Для цитирования: Логинов И.В., Щербин А.А. 2025. Методика извлечения типовых требований из описания вакансий методами графо-семантического анализа. *Экономика. Информатика*, 52(4): 928–945. DOI 10.52575/2687-0932-2025-52-4-928-945; EDN TMEYRY

The Method of Extracting Typical Requirements from Job Descriptions by Grapho-Semantic Analysis

¹Ilia V. Loginov, ²Alex A. Sherbin

¹ The Academy of the Federal Guard Service of the Russian Federation,
35 Priborostroiteльnaya St., Orel 302020, Russia

² Mid-Russia Institute of Management, RANEPA branch in Orel,
5a Pobedy Blvd., Orel 302028, Russia
loginov_iv@bk.ru, alex.sherbin1@mail.ru

Abstract. The paper addresses the problem of extracting and summarizing information about the requirements for the competencies of specialists based on the analysis of sets of textual job descriptions. The purpose of the

© Логинов И.В., Щербин А.А., 2025

study is to increase the validity of the generalization and extraction of data, taking into account additional information about the level of requirements. The achievement of the research goal is ensured by applying an approach to the analysis of textual data based on combining grapho-semantic analysis with a fuzzy linguistic model of the "desirability" level. Based on this approach, a formal model of job descriptions and a methodology for extracting generalized professional skills requirements from textual job descriptions of specialists are proposed. The results of applying the proposed approach to the vacancy corpus for the professions of business analyst, developer and DevOps engineer using the developed VectorCognitive software tool showed the possibility of combining a fuzzy linguistic model and grapho-semantic analysis to form generalized requirements for specialists.

Keywords: semantic analysis, graphs, vacancies, generalization, natural language processing, linguistic expressions, fuzzy sets

For citation: Loginov I.V., Shcherbin A.A. 2025. The Method of Extracting Typical Requirements from Job Descriptions by Grapho-Semantic Analysis. *Economics. Information technologies*, 52(4): 928–945 (in Russian). DOI 10.52575/2687-0932-2025-52-4-928-945; EDN TMEYRY

Введение

Развитие технологий и изменения рынка труда приводят к изменению в требованиях к профессиональным навыкам и знаниям специалистов. При этом изменения происходят довольно быстро, что требует непрерывного уточнения профессиональных стандартов и соответственно адаптации учебных программ. Возникает необходимость разработки инструментария выявления изменений и соответственно формирования новых требований. Исходное множество описаний вакансий представляет собой текстовые документы, в которых изложены основные требования к профессиональным знаниям и навыкам, а также содержатся общие положения об условиях рабочей деятельности. Исходя из указанного формата исходных данных задача извлечения типовых требований относится к классу задач семантического анализа. Высокая интенсивность изменений требований к ИТ-специалистам и развитие методов графо-семантического анализа показывает актуальность разработки механизмов обобщения типовых требований на основе анализа описаний множества вакансий от работодателей.

Обзор подходов к обобщению информации из описаний вакансий в ИТ-сфере

Проблема извлечения и обобщения информации из наборов текстовых данных является частью задач семантического анализа текстов. Цель настоящей работы заключается в разработке механизма обобщения навыков, знаний и умений, заданных в интернет-вакансиях ИТ-специалистов на естественном языке, и формирования наборов формальных требований:

$$\{V\} \rightarrow \{\text{Tr}\},$$

где V – структурированное текстовое описание вакансии ИТ-специалиста (в работе рассматриваются вакансии ИТ-специалистов на русском языке); $\{\text{Tr}\}$ – множество устойчиво повторяющихся требований к профессиональным навыкам, знаниям и умениям специалистов, указанных в вакансиях.

Анализ подходов к извлечению информации из текстовых описаний вакансий

Извлечение информации из вакансий нашло широкое применение при решении прикладных задач образования, подготовки и подбора кадров. Исходными данными в общем случае является первоисточник – объявление о вакансии, которое размещают потенциальные работодатели.

Задача определения потребностей в навыках (знаниях, а также дополнительные требования) широко рассмотрены в литературе. Так, в работе [Деев, 2024] выполнено исследование для сравнительного анализа компетентностных требований к образовательному

процессу. В работе [Wowczko, 2015] решена задача разработки эффективного метода мониторинга потребностей в профессиональных навыках. В работе [Минаев, 2022] рассматривалась задача выделения характерных укрупненных трудовых функций на материале текстовых требований. Полученные результаты могут быть использованы как специалистами – для адаптации портфеля компетенций в соответствии с требованиями, предъявляемыми к реальным работам, для получения более высокой заработной платы и большей мобильности в условиях неопределенности, так и образовательными организациями (в первую очередь курсами переподготовки) – для оперативной корректировки образовательных программ.

Математические методы обработки текстов вакансий заключаются в использовании двух подходов: на основе поисковых запросов и извлечения информации из заранее подготовленного корпуса вакансий, либо на основе непосредственной обработки корпуса текстовых вакансий.

В рамках первого подхода широко используются методы на основе построения модели предметной области: например, в работе [Яруллин, 2019] для получения упрощенной модели предметной области использован метод автоматического построения денотатного графа гипертекста [Курушин и др. 2018], либо на основе извлечения данных с использованием поисковых запросов: с использованием инструментария RapidMiner и R, модуля KNIME, Topic Extractor (Parallel LDA). Вариант алгоритма [Минаев, 2022] разработан на основе библиотеки моделирования Mallet: A Machine Learning for Language Toolkit.

В рамках второго подхода использовались семантические подходы обработки текстов. Так в статье [Vanetik, Kogan, 2023] выполнен сравнительный анализ оригинального алгоритма ОКАРІ BM25, модифицированного BERT метода – BERT-rank, на основе комбинации 4 классов текстовых описаний и 5 вариантов представлений текста (п-граммы слов и символов, BERT вложения в предложения, векторы tf-idf и объединения всех этих векторов). Для выделения семантических элементов стандартно использовались лемматизация, токенизация, удаление стоп-слов [Перстенева, Кучко, 2020], создание векторно-семантических моделей Word2vec. Для дальнейшего анализа выполнялся компартивный анализ на основе выделения словарей из 150 наиболее часто встречающихся слов с последующей оценкой похожести бинарными мерами сходства: коэффициент Жаккара и косинусное сходство, широко использованы методы кластеризации, в частности: кластеризации k-NN, метод кластеризации с использованием аппарата нечеткой логики и нейронной сети Fuzzy C-means, а также интеллектуальный анализ веб-контента [Litecky et al, 2010].

Область информационных технологий как сфера подготовки и применения кадров широко рассмотрена в литературе. Авторами [Диков, Широбокова, 2022] предложен механизм сопоставления навыков выпускников и требований рынка труда для решения задачи определения актуальности различных направлений подготовки «Прикладная информатика» для трудоустройства. В исследованиях [Деев, 2024] выполнено сравнительное обобщение компетентностных требований к образовательному процессу. В работе [Перстенева, Кучко, 2020] решалась задача анализа степени соответствия содержания рабочих программ СГЭУ – потребностям предприятий на основе текстовых вакансий. В работе [Минаев, 2022] рассматривалась задача выделения характерных укрупненных трудовых функций на материале требований, которые выставляют работодатели. В работе [Litecky et al, 2010] на основе метода интеллектуального анализа веб-контента проведен эксперимент с выделением информации о двадцати категориях ИТ-должностей. Для каждой из должностей выявлены сходные наборы навыков, которые соответствуют конкретным профессиям. При этом вопросы обобщения требований на основе углубленного семантического анализа пока остались без должного разрешения.

В качестве основного источника данных используются тексты сайтов – агрегаторов вакансий: для российской аудитории это «НН.ru», «Мой круг», «Яндекс.Работа», «Авито работа», а также специализированные [Яруллин, 2019; Перстенева, Кучко, 2020]. В работе [Litecky et al, 2010], используя приложение интеллектуального анализа веб-контента, проанализирована четверть миллиона описаний требований к ИТ-вакансиям из различных

поисковых систем. В статье [Vanetik, Kogan, 2023] использован набор данных, состоящий из резюме разработчиков программного обеспечения, извлеченных из общедоступного Telegram-канала. В качестве примера архитектуры системы интеллектуального анализа вакансий можно рассматривать работы [Яруллин, 2019; Деев, 2024], где решение задачи обеспечивается кластеризацией векторных моделей компетенций.

Обзор подходов к графо-семантическому анализу текстов вакансий

В рамках решения задачи извлечения информации из текстовых описаний вакансий широко используются механизмы обработки естественного языка NLP (Natural Language Processing) [Иванченко, Барауля, 2022]. Данный метод представляет собой механизмы обработки естественного языка, и соответственно выполнение операций по извлечению информации из текстовых сообщений является областью искусственного интеллекта по направлению обработки текста. В рамках метода обработки естественного языка (NLP) предложено множество векторных моделей семантического представления текстов, задаваемых в векторной форме. В соответствии с [Baroni et al, 2014], эти вектора могут быть подвергнуты дальнейшей обработке с помощью статистического моделирования: LSA [Landauer et al, 1997] и LDA [Blei et al, 2003], либо с использованием семантических нейронных сетей, таких как CW [Collobert, Weston, 2008], Word2Vec 3 [Mikolov et al, 2013] и GloVe [Pennington et al, 2014].

Явные модели пространства концептов, такие как ESA [Gabrilovich, Markovitch, 2007], SSA [Hassan, Mihalcea, 2011], а также их обобщения в виде NASARI [Camacho-Collados et al, 2015], для представления текстовых структур конструируют вектора типа «мешка слов» Bag of Concepts (BoC), что в общем случае позволяет описать основные темы текстовых данных и таким образом смоделировать его семантику.

Другим подходом является применение моделей, основанных на знаниях (так называемых Knowledge-based models). Указанные семантические модели используют словари Wordnet [Fellbaum, 1998] и Wiktionary, что позволяет задавать явные отношения между концептами для определения степени их семантической связи:

1) явный семантический анализ – Explicit Semantic Analysis (ESA) [Gabrilovich, Markovitch, 2007]. ESA конструирует концептуальное пространство термина путем поиска по инвертированному индексу совпадений термина и понятия, что позволяет извлекать концепты, соответствующие целевым поисковым запросам;

2) существенный семантический анализ – Salient Semantic Analysis (SSA) [Hassan, Mihalcea, 2011]. Метод определяет значение слова по его непосредственному контексту и использует прямые ассоциации между концептами;

3) латентный семантический анализ Latent Semantic Analysis (LSA) [Deerwester et al, 1990]. Метод предназначен для построения матрицы совпадений концептов и документов из текстового корпуса, в которой концепты и документы, имеющие схожее значение, располагаются близко;

4) ориентированный (поисковый) семантический анализ Mined semantic analysis (MSA) [Shalaby, Zadrozny, 2017] – модель концептуального пространства, которая использует обучение без учителя для генерации семантических представлений текста. Метод MSA представляет текстовые структуры на основе механизма мешка слов, что позволяет выявлять неявные связи между понятиями путем анализа их ассоциаций:

$$S = S(w_i, w_j) = \begin{cases} 1, & S_{cos}(w_i, w_j) \geq \lambda \\ \frac{S_{cos}(w_i, w_j)}{\lambda}, & S_{cos}(w_i, w_j) < \lambda \end{cases}.$$

где w_i, w_j – концепты, $S(w_i, w_j)$ – оценка семантической схожести концептов, S_{cos} – косинусная мера сходства, λ – заданное расстояние между концептами.

5) модель случайных блужданий [Hughes, Ramage, 2007] предполагает построение концептуального пространства векторов – набора концептов на основе случайных блужданий в больших ассоциативных базах.

С позиций семантики текстовое (гипертекстовое) описание вакансии V представляет собой некоторый вид текста, который в отдельных случаях может иметь мультимедийную форму. В литературе рассматриваются различные модели графо-семантического анализа применительно к заданному виду графовых семантических структур [Zhang et al, 2013; Бермудес, 2017], [Логинова, 2024]: на основе анализа структур графов; на основе расчета мер информации; векторные модели семантического анализа; частотный анализ; методы на основе синтеза глоссариев, а также гибридные методы, объединяющие достоинства указанных. Наличие иерархической упорядоченности в текстовом (мультимедийном) описании вакансии позволяет применить методы графо-семантического анализа класса интеллект-карт [Логинов, 2024].

С точки зрения теории графов, текстовое описание вакансии ИТ-специалиста представляет собой ориентированное дерево с семантическими узлами и ребрами. Для решения рассматриваемой в работе задачи обобщения семантических графов широко используются методы определения похожести на основе механизмов обучения. В соответствии с [Guixiang et al, 2021] выделены следующие механизмы, позволяющие выполнять задачи оценки семантической схожести иерархически упорядоченных описаний вакансий:

- на основе расширенных графов:
 - уровень узлов (node2vec-PCA, Bag-of-vectors);
 - уровень структур графов (graph2vec, Neural networks with Structure2vec, Simple permutation-invariant GCN, SEED, DGCNN, N-Gram graph embedding);
- на основе методов графовых нейросетей:
 - модели GNN-CNN (GSimCNN, SimGNN);
 - модели двойных (сиамских) нейросетей (Siamese GCN, Higher-order Siamese GCN, Community-preserving Siamese GCN, Hierarchical Siamese GNN, Siamese GCN for image retrieval);
 - модели поисковых сетей (GMN: graph matching network, NeuralMCS: neural maximum common subgraph GMN, Hierarchical graph matching network, NCMN: neural graph matching network, GMNs for image matching);
- на основе глубоко обученных графовых ядер (Deep graph kernels, Deep divergence graph kernels, Graph neural tangent kernel).

Результаты анализа методов графо-семантического анализа показывают, что объединение подходов семантического анализа текста на основе мешка слов ВоС с дополнительным учетом структуры вакансии как документа в виде графа позволяет повысить точность обобщения информации. Это позволяет сделать вывод о потенциальной возможности применения данного метода при решении задачи анализа множеств текстовых описаний вакансий.

Методика обобщения информации из множества текстовых описаний вакансий

В рамках исследования сформулирована практическая задача, которая заключается в изучении вакансий ИТ-специалистов на предмет формирования требований к необходимым профессиональным компетенциям, динамически изменяемым с течением времени. Для решения данного вопроса возникает необходимость в разработке механизма обработки текстовых описаний вакансий специалистов в интересах обобщения требуемых навыков специалистов (знаний, умений, компетенций). Цель методики заключается в преобразовании множества текстовых описаний вакансий, относящихся к некоторой выборке, сформированной по профессиональному принципу, в множество типовых профессиональных требований к специалистам в соответствующей группе вакансий:

$$A: \{V\} \xrightarrow{A} \{\text{Tp}\},$$

где V – структурированное текстовое представление вакансии ИТ-специалиста (в работе рассматриваются вакансии ИТ-специалистов на русском языке); $\{\text{Тр}\}$ – множество стабильно повторяющихся требований к профессиональным навыкам, знаниям и умениям, зафиксированных в описаниях вакансий.

Формальная модель описания вакансий

Исходной информацией о вакансиях специалистов является его текстовое (гипертекстовое) представление V . С формальной точки зрения в составе объявления о вакансии специалиста выделяются до нескольких десятков параметров [Herlambang, Nur, 2021], часть из которых является обязательным параметром (например, «название должности», «место работы»), а часть – опциональными (например, «размер оплаты»).

Смысловое содержание вакансий, используемое в работе, опирается на модель вакансии, рассмотренной в работе [Диков, Широбокова, 2022], и описывается формальной моделью:

$$V = \langle V^t, V^\Phi, V^c \rangle,$$

где V^t – текстовое (гипертекстовое или мультимедийное) описание вакансии; V^Φ – формальная модель вакансии; V^c – семантическое представление вакансии;

$$V^\Phi = \langle URL, Nm, Pay, SC, Ctr, EMP, ASK, Cond, PT, Reg \rangle,$$

где URL – ссылка вакансии на онлайн-сервисе вакансий, Nm – наименование вакансии (профессии), Pay – диапазон заработной платы, SC – график работы для данной вакансии (вахтовый метод, гибкий график, сменный график и т. д.), Ctr – информация о компании, которая предоставляет вакансию ИТ-специалиста, EMP – занятость для данной вакансии (полная, удаленная и т. д.), ASK – множество требований к основным профессиональным навыкам (знаниям, умениям, компетенциям) для соответствующей вакансии, $Cond$ – условия выполнения профессиональных обязанностей, PT – профессиональные обязанности, Reg – регион вакансии, указанный в объявлении.

В работе основное исследование связано с анализом требований к профессиональным навыкам, исходя из этого рассмотрим семантическую модель множества требований $ASK^c \in V^c$ в форме семантического дерева:

$$ASK^c: G = (V, E),$$

где узлы семантического описания требований к профессиональным навыкам из вакансий $v \in V$ – представляют собой термины в кортежи следующего вида:

$$v = \langle ask, req \rangle,$$

где ask – представляет собой лексическое выражение LE , которое весьма компактно (количество слов во фразе лежит в диапазоне от 1 до 15), req – лексическая оценка уровня требований {обязательные, рекомендуемые, сулящие преимущество}, описываются лексическими переменными: {«рекомендуется», «будет плюсом» ...}.

Примеры лексических выражений, задающих требования к профессиональным компетенциям:

$$LE_1 = \text{"Знание нотаций UML и IDEF"},$$

$$LE_2 = \text{"Знание английского языка на уровне Intermediate"},$$

$$LE_3 = \text{"Знание нотации BPMN 2.0, навыки моделирования в Business Studio (желательно)"}.$$

В отличие от выражений 1 и 2, выражение 3 имеет сложную структуру и после предобработки может быть представлено в форме следующего подграфа семантического дерева описания требований:

$$G_3 = (V_3, E_3), G_3 \in G,$$

$$V_3 = \{v_3^1, v_{31}^2, v_{32}^2\},$$

$$E_3 = \{e_{31}^2, e_{32}^2\},$$

где:

$$v_3^1 = \langle \text{"Знание нотации BPMN 2.0, навыки моделирования в Business Studio", "желательно"} \rangle$$

$$v_{31}^2 = \langle \text{"Знание нотации BPMN 2.0", "желательно"} \rangle$$

$$v_{32}^2 = \langle \text{"Навыки моделирования в Business Studio", "желательно"} \rangle$$

e_{31}^2, e_{32}^2 – связи узлов семантического дерева типа «Часть-целое», показывающие путь от корня указанного дерева v_3^1 до листьев v_{31}^2, v_{32}^2 .

Описание методики извлечения обобщенных требований к профессиональным навыкам из текстовых описаний

Основные положения методики извлечения обобщенных требований к профессиональным навыкам представлены на рис. 1. Выполнение задачи реализуется посредством следующих шагов:

1. Формирование исходного множества текстовых описаний вакансий: $\{V\}$ – реализуется путем загрузки данных вакансий из нескольких источников и их преобразования к текстовому виду.

2. Преобразование текстовых описаний вакансий в семантическую модель требований к вакансиям:

$$\forall V_i \in \{V\}: V_i^T \rightarrow ASK_i^C,$$

$$\{V\} \rightarrow \{ASK^C\}.$$

Преобразование выполняется в три этапа:

– выделение отдельных строк требований с составными или простыми лексическими выражениями;

– разбиение составных лексических выражений на простые;

– представление требований в форме кортежа путем выделения лексической модели уровня требований (при отсутствии соответствующего выражения проставляется оценка «не задано»).

3. Преобразование текстовых данных описания вакансий в нормализованную форму, предназначенную для машинной обработки:

$$\{v_{ij}\} \rightarrow \{v_{ij}^H\},$$

где v_{ij} – j -е требование к i -й вакансии в текстовом виде, приведенное к нижнему регистру.

Преобразование выполняется в два этапа и заключается в двухступенчатом процессе фильтрации вспомогательных слов, не несущих значимой семантической нагрузки.

4. Векторизация текстовых данных. Для проведения векторизации текста применяется библиотека машинного обучения «scikit-learn» и её модуль «CountVectorizer», который преобразует входные текстовые данные в матрицу значений числовых ключей вида:

$$\{v_{ij}^H\} \rightarrow UID, \{v_k^{uid}\},$$

где UID – вектор базовых лексических единиц в нормализованной форме, v_k^{uid} – k -й блок требований ($k = (ij)$), $v_k^{uid} = \langle v_{uid}^{ask}, v_{uid}^{req} \rangle_k$, v_{uid}^{ask} – вектор ключей соответствия лексического выражения требований вектору UID , v_{uid}^{req} – лексическая оценка уровня требований.

5. Предобработка матрицы $\{v_{uid}^{ask}\}$ для снижения ее размера методом сингулярного разложения

$$\{v_{uid}^{ask}\} \rightarrow \{v_{uid}^{ask}\}^*.$$



Рис. 1. Методика извлечения обобщенных требований к профессиональным навыкам из текстовых описаний вакансий специалистов

Fig. 1. Method for extracting generalized professional skills requirements from text descriptions of specialist vacancies

6. Расчет попарной схожести требований v_i^{ask} и v_j^{ask} на основе расчета метрики взвешенного косинусного сходства:

$$s = s(v_i^{ask}, v_j^{ask}) = \frac{v_i^{ask} \odot v_j^{ask}}{\|v_i^{ask}\| \cdot \|v_j^{ask}\| \cdot \sum_m \alpha_m},$$

где \odot – операция взвешенного упорядоченного произведения координат векторов требований к ИТ-специалистам v_i^{ask} и v_j^{ask} в заданном семантическом пространстве:

$$le_{im} \odot le_{jm} = \alpha_m \cdot le_{im} \cdot le_{jm},$$

где требования $v_i^{ask}, v_j^{ask} \in \{v_{uid}^{ask}\}^*$, α_m – весовой коэффициент значимости m -го по похожести концепта из вектора слов, координаты которого упорядочены по похожести.

Матрица попарной схожести требований S в форме лингвистических концептов имеет следующий вид:

$$S = \begin{matrix} & v_{11} & \dots & v_{ij} & \dots & v_{IJ} \\ v_{11} & 1,00 & \dots & s = s(v_{11}^{ask}, v_{ij}^{ask}) & \dots & s = s(v_{11}^{ask}, v_{IJ}^{ask}) \\ \dots & s = s(v_{ij}^{ask}, v_{11}^{ask}) & \dots & \dots & \dots & s = s(v_{ij}^{ask}, v_{IJ}^{ask}) \\ v_{ij} & \dots & \dots & 1,00 & \dots & \dots \\ \dots & s = s(v_{IJ}^{ask}, v_{11}^{ask}) & \dots & s = s(v_{IJ}^{ask}, v_{ij}^{ask}) & \dots & 1,00 \\ v_{IJ} & \dots & \dots & \dots & \dots & \dots \end{matrix}$$

7. Иерархическая кластеризация требований v_{ij} по уровню сходства. В рамках исследования рассматривался иерархический кластерный анализ. Метрикой (расстоянием между точками в пространстве требований) выступала величина обратная величине схожести объектов:

$$d(v_i^{ask}, v_j^{ask}) = 1 - s = 1 - s(v_i^{ask}, v_j^{ask}).$$

В качестве метода агломерации использовался метод средней связи:

$$d(M_k, M_l) = \frac{1}{n_k n_l} \sum_{v_i^{ask} \in M_k} \sum_{v_i^{ask} \in M_l} d(v_i^{ask}, v_j^{ask}).$$

Поскольку объединение элементов в группы однотипных требований целесообразно выполнять только для случая высокой степени схожести, то кластеризация выполняется до тех пор, пока расстояние между элементами (кластерами) будет меньше заданного: $d_k \leq d^{kp}$. Исходя из условий проведения исследований, критическое значение величины расстояния лежит в диапазоне: $d^{kp} \in [0,3; 0,4]$. Результатом кластеризации является множество кластеров, включающих от 1 до n требований в каждом:

$$\{v_{ij}^h\} \rightarrow \{v_k^{обоб}\},$$

где $v_k^{обоб} = \langle \{v_{ik}^{ask}\}, \{v_{ik}^{req}\} \rangle$ – вектор, содержащий однотипные (схожие) требования, $\{v_{ik}^{ask}\} = \{v_{1k}^{ask}, v_{2k}^{ask} \dots v_{n_k k}^{ask}\}$, $\{v_{ik}^{req}\} = \{v_{1k}^{req}, v_{2k}^{req} \dots v_{n_k k}^{req}\}$.

8. Восстановление лингвистического значения требования:

$$v_k^{обоб,ask} \rightarrow v_k^{обоб,le}.$$

Задача решается нахождением i -го требования в k -м кластере v_{ik}^{ask} , наиболее близкого к среднему значению кластера. Таким образом для каждого кластера формируется кортеж, состоящий из лексического выражения требований из вакансии и количества повторений их в выборке: $v_k^{обоб,le} = \langle le, n \rangle$.

9. Оценка уровня необходимости требования выполняется путем обобщения нечетких лингвистических оценок на предварительно сформированной шкале. Все новые лингвистические оценки, отсутствующие в предварительно сформированном множестве, обрабатываются экспертом путем определения их функций принадлежности $\mu(x)$:

$$v_{ik}^{req} = \mu_{ik}(x),$$

$$v_k^{обоб,req} = \text{среднее}(\mu_{1k}(x), \mu_{2k}(x), \dots, \mu_{n_k k}(x),),$$

где среднее – оператор нахождения среднего значения нескольких нечетких чисел.

Далее полученная функция принадлежности переводится в лингвистическую оценку.

10. Формирование набора требований осуществляется путем использования предварительно выполненных настроек, либо путем указания оператором соответствующих настроек в автоматизированном режиме. Используется два параметра: минимальный уровень необходимости требования req^{min} и минимальная частота встречаемости требования ϑ^{min}

$$req^{min}, \vartheta^{min}: \{V\} \rightarrow \{v_k^{обоб}\} \xrightarrow{req^{min}, \vartheta^{min}} \{\text{Тр}\}.$$

Минимальная частота встречаемости требования ϑ^{min} – показывает долю вакансий, в которых встречалось требование: $\vartheta^{min} \in [0,4; 0,5]$.

Минимальный уровень необходимости требования req^{min} = "желательно".

Выявление требований к специалистам из множества текстовых описаний вакансий

В рамках исследования выполнена работа по разработке макета информационной системы, позволяющей выделять актуальные обобщенные квалификационные требования специалистов из набора описаний вакансий.

Пример выполнения иерархической кластеризации требований к вакансиям

Исходные данные включают в свой состав описание трех вакансий, из которых было выделено 13 простых требований (лингвистических выражений). Матрица парной похожести лингвистических выражений и оценки необходимости требований (на основе частотного анализа) приведены в табл. 1.

Таблица 1
Table 1

Оценки парной похожести и необходимости требований
Assessment of paired similarity and necessity of requirements

v_{ij}	№	1	2	3	4	5	6	7	8	9	10	11	12	13	req
v_{11}	1	1,00	0,41	0,41	0,41	0,20	0,00	0,33	0,00	0,71	0,00	0,00	1,00	0,32	«не задано»
v_{12}	2	0,41	1,00	0,33	0,00	0,20	0,00	0,29	0,00	0,00	0,00	0,41	0,50	0,32	«требуется»
v_{13}	3	0,41	0,33	1,00	0,00	0,20	0,00	0,86	0,00	0,00	0,00	0,00	0,41	0,52	«требуется»
v_{14}	4	0,41	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,58	0,00	0,00	0,41	0,00	«желательно»
v_{15}	5	0,20	0,20	0,20	0,00	1,00	0,00	0,17	0,35	0,00	0,00	0,00	0,00	0,00	«требуется»
v_{16}	6	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,58	0,00	0,00	0,00	0,00	0,00	«требуется»
v_{21}	7	0,33	0,29	0,86	0,00	0,17	0,00	1,00	0,00	0,00	0,00	0,00	0,35	0,89	«требуется»
v_{22}	8	0,00	0,00	0,00	0,00	0,35	0,58	0,00	1,00	0,00	0,00	0,00	0,00	0,00	«требуется»
v_{23}	9	0,71	0,00	0,00	0,58	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,71	0,10	«желательно»
v_{31}	10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	«не задано»
v_{32}	11	0,00	0,41	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,17	0,11	«не задано»
v_{33}	12	1,00	0,50	0,41	0,41	0,00	0,00	0,35	0,00	0,71	0,00	0,17	1,00	0,29	«не задано»
v_{34}	13	0,32	0,32	0,52	0,00	0,00	0,00	0,89	0,00	0,10	0,00	0,11	0,29	1,00	«не задано»

Результат кластеризации с $d^{kp} = 0,5$ позволил сформировать 8 кластеров (иерархическая кластеризация приведена на рис. 2), обобщенные оценки требований к которым приведены ниже по тексту:

$$V^{обоб} = \{v_k^{обоб}\} = \{\{v_{ik}^{ask}\}, \{v_{ik}^{req}\}\}$$

$$V^{\text{обоб}} = \left\{ \begin{array}{l} v_1^{\text{обоб}} \\ v_2^{\text{обоб}} \\ v_3^{\text{обоб}} \\ v_4^{\text{обоб}} \\ v_5^{\text{обоб}} \\ v_6^{\text{обоб}} \\ v_7^{\text{обоб}} \\ v_8^{\text{обоб}} \end{array} \right\} = \left\{ \begin{array}{l} \langle \{v_{11}^{\text{ask}}, v_{23}^{\text{ask}}, v_{33}^{\text{ask}}\}, \{v_{11}^{\text{req}}, v_{23}^{\text{req}}, v_{33}^{\text{req}}\} \rangle \\ \langle \{v_{12}^{\text{ask}}\}, \{v_{12}^{\text{req}}\} \rangle \\ \langle \{v_{13}^{\text{ask}}, v_{21}^{\text{ask}}, v_{34}^{\text{ask}}\}, \{v_{13}^{\text{req}}, v_{21}^{\text{req}}, v_{34}^{\text{req}}\} \rangle \\ \langle \{v_{14}^{\text{ask}}\}, \{v_{14}^{\text{req}}\} \rangle \\ \langle \{v_{15}^{\text{ask}}\}, \{v_{15}^{\text{req}}\} \rangle \\ \langle \{v_{16}^{\text{ask}}, v_{22}^{\text{ask}}\}, \{v_{16}^{\text{req}}, v_{22}^{\text{req}}\} \rangle \\ \langle \{v_{31}^{\text{ask}}\}, \{v_{31}^{\text{req}}\} \rangle \\ \langle \{v_{32}^{\text{ask}}\}, \{v_{32}^{\text{req}}\} \rangle \end{array} \right\}$$

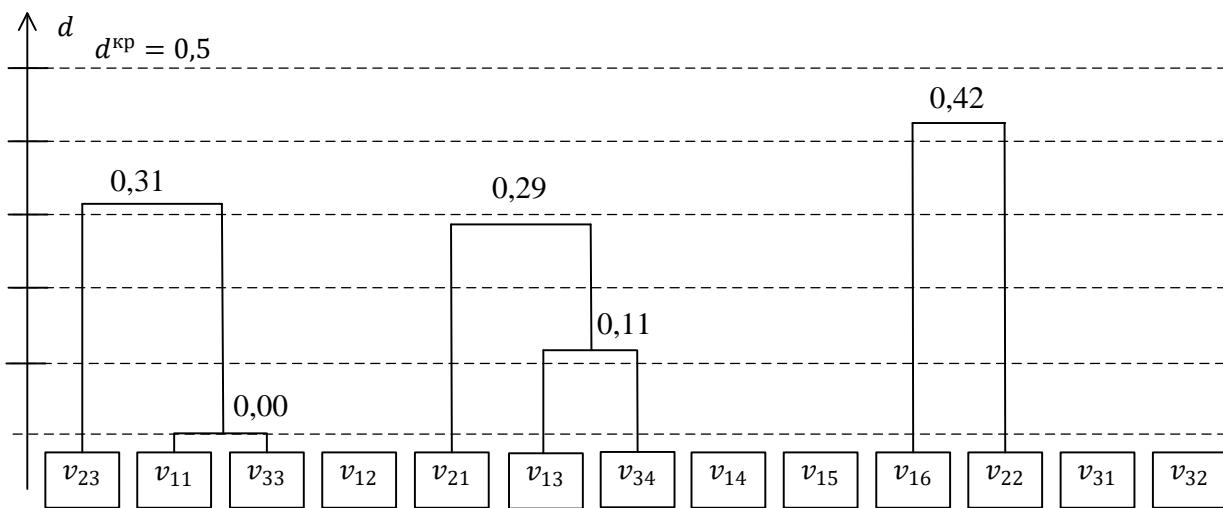


Рис. 2. Результаты иерархической кластеризации требований, выраженных в лингвистической форме
 Fig. 2. Results of hierarchical clustering of requirements expressed in linguistic form

$$V^{\text{обоб, рек}} = \left\{ \begin{array}{l} \text{"между требуется и желательно"} \\ \text{"требуется"} \\ \text{"требуется"} \\ \text{"желательно"} \\ \text{"требуется"} \\ \text{"требуется"} \\ \text{"требуется"} \\ \text{"требуется"} \end{array} \right\}$$

Результаты формирования обобщенных квалификационных требований:

– при $req^{\min} = \text{"желательно"}, \vartheta^{\min} = 1$:

$$\{\text{Tp}\} = \{v_{11}^{\text{ask}}, v_{12}^{\text{ask}}, v_{34}^{\text{ask}}, v_{14}^{\text{ask}}, v_{15}^{\text{ask}}, v_{16}^{\text{ask}}, v_{31}^{\text{ask}}, v_{32}^{\text{ask}}\};$$

– при $req^{\min} = \text{"требуется"}, \vartheta^{\min} = 1$:

$$\{\text{Tp}\} = \{v_{12}^{\text{ask}}, v_{34}^{\text{ask}}, v_{15}^{\text{ask}}, v_{16}^{\text{ask}}, v_{31}^{\text{ask}}, v_{32}^{\text{ask}}\};$$

– при $req^{\min} = \text{"желательно"}, \vartheta^{\min} = 2$:

$$\{\text{Tp}\} = \{v_{11}^{\text{ask}}, v_{34}^{\text{ask}}, v_{16}^{\text{ask}}\};$$

– при $req^{\min} = \text{"требуется"}, \vartheta^{\min} = 2$:

$$\{\text{Tp}\} = \{v_{34}^{\text{ask}}, v_{16}^{\text{ask}}\}.$$

Разработка программного макета информационной системы по выявлению и анализу квалификационных требований к ИТ-специалистам

Для решения задачи исследования разработан прототип системы «ВекторКогнитив» [Щербин, 2024], предназначенный для анализа вакансий, их обобщения и возможности сопоставления требований с заявляемым уровнем компетенций. Программный макет включает в себя компоненты (разработан на базе веб-фреймворка «Django»):

- загрузки и предобработки данных с рекрутинговых ресурсов («HeadHunter», «Авито», а также поддерживается ручная загрузка данных);
- формирования подмножества вакансий из набора загруженных в базу данных (на основе фильтрации по дате размещения и тематике) и их обработке с целью формирования множества требований в соответствии с предложенной методикой;
- отображения и визуализации результатов обработки данных с использованием сводной аналитической панели (дашборда). Дашиборд представляет собой визуальную панель, на которой отображаются данные, необходимые для анализа. С помощью дашборда можно объединить все необходимые данные на одной странице. На рис. 3 представлен интерфейс разработанного автором программного продукта, который помогает визуально представить результат выполнения латентно-семантического анализа данных;
- хранения наборов загруженных данных и результатов их обработки в форме множеств требований (разработана структура базы данных).

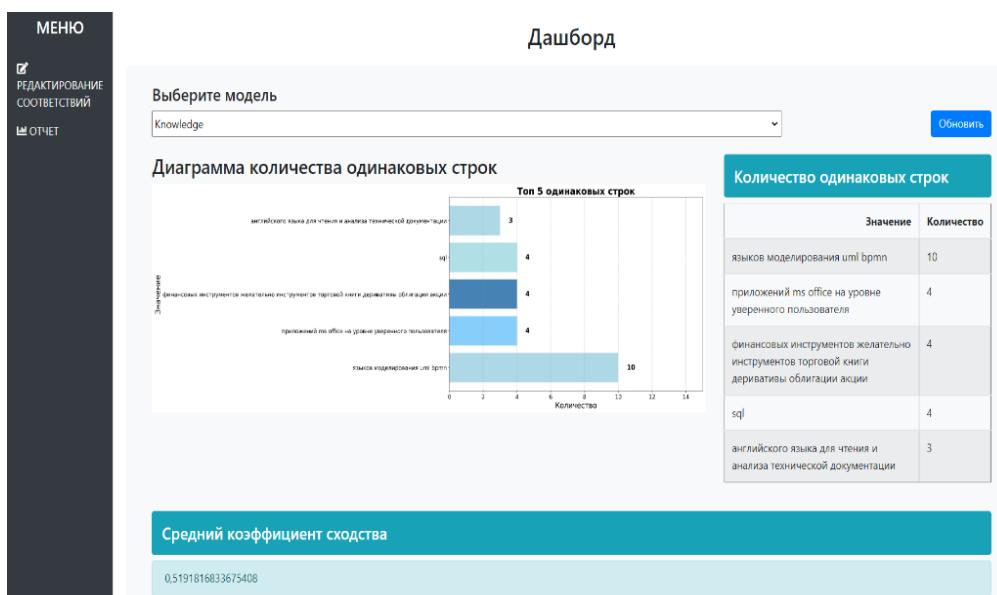


Рис. 3. Пример аналитического дашборда (сводного отчета) с анализом квалификационных требований в программном средстве «ВекторКогнитив»
Fig. 3. Example of VectorCognitive software analytical dashboard with the analysis of qualification requirements

Тестирование разработанной методики и программного средства выполнено на выборке данных вакансий «HeadHunter». Формирование обобщенных требований к квалификации выполнено для следующих категорий: бизнес-аналитики, DevOps-инженеры и разработчики (инженеры-программисты). В рамках эксперимента загружено следующее количество вакансий для каждой категории (табл. 2).

Для каждой группы специалистов было выделено множество первичных требований к навыкам и знаниям (табл. 3):

$$\forall V_i \in \{V\}: V_i^T \rightarrow ASK_i^c = \langle ASK_i^{c, \text{знан}}, ASK_i^{c, \text{нав}} \rangle, \\ \{V\} \rightarrow \{ASK^c\} = \{ASK_i^{c, \text{знан}}, ASK_i^{c, \text{нав}}\}.$$

Таблица 2
 Table 2

Количество вакансий по профессиям
 Number of vacancies by profession

Профессия	Количество вакансий
Бизнес-аналитики	450
DevOps-инженеры	160
Разработчики	460

Таблица 3
 Table 3

Первичные требования к специалистам по категориям
 Primary requirements for specialists by category

Профессия	Категории	
	навыки	знания
Бизнес-аналитики	1860	220
DevOps-инженеры	880	80
Разработчики	2300	200

В результате проведения анализа с использованием разработанного приложения «ВекторКогнитив» для рассмотренных профессий обобщены требования к навыкам и знаниям. Результаты эксперимента – сформированные перечни требований приведены в табл. 4. Пример частоты встречаемости указанных требований приведен на рис. 4.

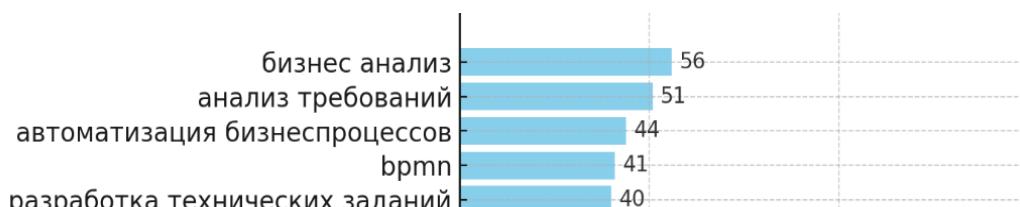


Рис. 4. Результаты эксперимента по формированию требований к навыкам
 Fig. 4. Results of the experiment to form skill requirements

Таблица 4
 Table 4

Результаты эксперимента по выявлению требований
 к профессиональным знаниям специалистов

Results of the experiment to identify the requirements for professional knowledge of specialists

Профессия	Бизнес-аналитик	DevOps-инженер	Разработчик
1	2	3	4
Требования к навыкам (пять наиболее встречающихся)	анализ бизнес-требований; бизнес-анализ; автоматизация бизнес-процессов; моделирование бизнес-процессов; навык работы с bpmn;	навык работы в kubernetes; навык работы docker; навык работы ansible; навык работы с postgresql; навык работы gitlab;	навык работы на javascript; навык работы с git; навык работы sql; навык работы spring framework; навык работы с python;

Окончание табл. 4
End of Table 4

1	2	3	4
Требования к знаниям (пять наиболее встречающихся)	знание языков моделирования uml и bpmn; знание приложений ms office на уровне уверенного пользователя; знание финансовых инструментов; знание sql; знание английского языка для чтения и анализа технической документации;	знание инфраструктуры кода; знание работы облачных технологий; знание языков программирования; знание основ кибербезопасности; знание основ работы с базами данных;	знание языков программирования, фреймворков и библиотек; знание работы web-сервисов; знание архитектуры ПО; знание механизмов тестирования и отладки; знание облачных платформ.

Результат работы программного средства «ВекторКогнитив» на реальных исходных данных текстовых описаний вакансий подтверждает возможность решения задачи извлечения типовых требований из описания множества вакансий методами графо-семантического анализа с использованием предложенной методики.

Заключение

В работе рассмотрена проблема анализа наборов текстовых описаний вакансий специалистов с целью извлечения и обобщения информации о требованиях с учетом уровня «желательности». Результаты проведенного анализа показали, что решение указанной задачи рассмотрено в литературе путем непосредственного семантического анализа корпуса вакансий, либо методом извлечения данных через поисковые запросы, однако они не учитывают уровень неопределенности необходимости требования. Для снятия этого ограничения в работе предложен подход к анализу текстовых данных на основе объединения графо-семантического анализа с нечеткой лингвистической моделью уровня «желательности». На основе указанного подхода предложена формальная модель описания вакансий и методика извлечения обобщенных требований к профессиональным навыкам из текстовых описаний вакансий специалистов. В результате проведения экспериментального анализа на основе корпуса вакансий с использованием разработанного приложения «ВекторКогнитив» для профессий бизнес-аналитика, разработчика и DevOps-инженер сформированы требования к часто встречающимся навыкам и знаниям с учетом уровня их необходимости. Результаты показали возможность комбинации нечеткой лингвистической модели и графо-семантического анализа для формирования обобщенных требований к специалистам.

Список литературы

- Бермудес С.Х.Г. 2017. Метод измерения семантического сходства текстовых документов. *Известия ЮФУ. Технические науки*, 3(188): 17–29. DOI 10.23683/2311-3103-2017-3-17-29.
- Деев М.В. 2024 Архитектура системы интеллектуального анализа компетенций для актуализации образовательных программ вуза. *Информатика и образование*, 39(3): 29–43. DOI 10.32517/0234-0453-2024-39-3-29-43.
- Диков М.Е., Широбокова С.Н. 2022. О проектных решениях цифрового инструментария профориентации по определению востребованности направлений подготовки на основе анализа описаний вакансий. *Инженерный вестник Дона*, 12(96): 65–74.
- Иванченко О.В., Барауля Е.В. 2022. Влияние обработки естественного языка (NLP) на цифровой маркетинг. Развитие логистики в условиях санкционных ограничений и международной

- экономической интолерантности: материалы международной научно-практической конференции: XVIII Южно-Российский логистический форум, Ростов-на-Дону, 07–08 октября 2022 года. Ростовский государственный экономический университет «РИНХ»; Южно-Российская ассоциация логистики. Ростов-на-Дону: Ростовский государственный экономический университет «РИНХ». 133–137.
- Курушин Д.С., Леонов Е.Р., Соболева О.В. 2018. О возможном подходе к автоматическому построению денотатного графа гипертекста. Информационная структура текста. Сб.статьй. РАН. ИНИОН. М., 113–118.
- Логинов И.В., Логинова Ю.В. 2024. Сравнение схожести интеллект-карт в задачах маркетингового анализа. *Вестник Кемеровского государственного университета. Серия: Политические, социологические и экономические науки*, Т. 9, № 3(33): 410–423. DOI 10.21603/2500-3372-2024-9-3-410-423.
- Логинова Ю.В., Логинов И.В. 2024. Применение семантического анализа в стратегическом маркетинге при использовании инструмента интеллект-карт. *Вестник Тюменского государственного университета. Социально-экономические и правовые исследования*, Т. 10, № 1(37): 103–123. DOI 10.21684/2411-7897-2024-10-1-103-123.
- Минаев Д.В. 2022. Исследование компетентностной модели образовательной программы на основе интеллектуального анализа профессиональных требований рынка труда. *Управленческое консультирование*, 10: 65–83.
- Перстенева Н.П., Кучко А.Ю. 2020. Компаративный анализ описаний вакансий и рабочих программ. *Наука XXI века: актуальные направления развития*. 1-2: 37–42.
- Щербин А.А. 2024. Применение методов обработки естественного языка для решения задач анализа вакансий бизнес-аналитиков на предмет требований к знаниям. *Образование и наука без границ: социально-гуманитарные науки*, 23: 134–138.
- Яруллин Д.В. 2019. Автоматическое построение профессионального портфеля ИТ-специалиста на основе текстов вакансий. *Филология в XXI веке*, 1(3): 17–21.
- Baroni M., Dinu G., Kruszewski G. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 1: 238–247.
- Blei D.M., Ng A.Y., Jordan M.I. 2003. Latent dirichlet allocation. *The Journal of machine Learning research*, 3: 993–1022.
- Camacho-Collados J., Pilehvar M.T., Navigli R. 2015. Nasari: a novel approach to a semantically-aware representation of items. In Proceedings of NAACL, 567–577.
- Collobert R., Weston J. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, 160–167. ACM.
- Deerwester S.C., Dumais S.T., Landauer Th.K., Furnas G.W., Harshman R.A. 1990. Indexing by latent semantic analysis. *JAsIs*, 41(6): 391–407.
- Fellbaum Ch. 1998. WordNet: An Electronic Lexical Database. Bradford Books.
- Gabrilovich E., Markovitch Sh. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In IJCAI, 7: 1606–1611.
- Guixiang M., Nesreen A., Theodore W., Philip Y. 2021. Deep graph similarity learning: a survey. *Data Mining and Knowledge Discovery*, 35. 10.1007/s10618-020-00733-5.
- Hassan S., Mihalcea R. 2011. Semantic relatedness using salient semantic analysis. In AAAI.
- Herlambang P., Nur R. 2021. Job Standard Parameters from Online Job Vacancy. *IPTEK Journal of Proceedings Series*, 46. 10.12962/j23546026.y2020i6.8905.
- Hughes Th., Ramage D. 2007. Lexical Semantic Relatedness with Random Graph Walks. Conference: EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, June 28–30, 2007, Prague, Czech Republic. 581–589.
- Landauer Th.K., Laham D., Rehder B., Schreiner M.E. 1997. How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In Proceedings of the 19th annual meeting of the Cognitive Science Society, 412–417. Citeseer.
- Litecky Ch., Aken A., Ahmad A., Nelson, H. 2010. Mining for Computing Jobs. *Software, IEEE*, 27: 78–85. 10.1109/MS.2009.150.

- Mikolov T., Chen K., Corrado G., Dean J. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Pennington J., Socher R., Manning Ch. 2014. Glove: Global vectors for word representation. Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014), 12: 1532–1543.
- Shalaby W., Zadrozny W. 2017. Mined semantic analysis: A new concept space model for semantic representation of textual data. 2122–2131. 10.1109/BigData.2017.8258160.
- Vanetik N., Kogan G. 2023. Job Vacancy Ranking with Sentence Embeddings, Keywords, and Named Entities. *Information*, 14: 468. 10.3390/info14080468.
- Zhang Z., Gentile A., Ciravegna F. 2013. Recent advances in methods of lexical semantic relatedness – A survey. *Natural Language Engineering*, 19. 10.1017/S1351324912000125.
- Wowczko I. 2015. Skills and Vacancy Analysis with Data Mining Techniques. *Informatics*, 2: 31–49. 10.3390/informatics2040031.

References

- Bermudes S.H.G. 2017. Metod izmereniya semanticeskogo skhodstva tekstovyh doku-mentov [A method for measuring the semantic similarity of text documents]. *Izvestiya YuFU. Tekhnicheskie nauki*, 3(188): 17–29. DOI 10.23683/2311-3103-2017-3-17-29.
- Deev M.V. 2024. Arhitektura sistemy intellektual'nogo analiza kompetencij dlya aktualizacii obrazovatel'nyh programm vuza [Architecture of the intellectual competence analysis system for updating university educational programs]. *Informatika i obrazovanie*, 39(3): 29–43. DOI 10.32517/0234-0453-2024-39-3-29-43.
- Dikov M.E., Shirobokova S.N. 2022. O proektnyh resheniyah cifrovogo instrumen-tariya proforientacii po opredeleniyu vostrebovannosti napravlenij podgotovki na osnove analiza opisanij vakansij [On design solutions for a digital career guidance tool to determine the relevance of training areas based on the analysis of job descriptions]. *Inzhenernyj vestnik Doma*, 12(96): 65–74.
- Ivanchenko O.V., Baraulya E.V. 2022. Vliyanie obrabotki estestvennogo yazyka (NLP) na cifrovoj marketing [The impact of natural language processing (NLP) on digital marketing]. Razvitie logistiki v usloviyah sankcionnyh ogranicenij i mezhdunarod-noj ekonomiceskoy intolerantnosti: materialy mezhdunarodnoj nauchno-prakticheskoy konfe-rencii: XVIII Yuzhno-Rossijskij logisticheskij forum, Rostov-na-Donu, 07–08 oktyabrya 2022 goda. Rostovskij gosudarstvennyj ekonomiceskij universitet \"RINH\"; Yuzhno-Rossijskaya as-sociaciya logistiki. Rostov-na-Donu: Rostovskij gosudarstvennyj ekonomiceskij universi-tet \"RINH\", 133–137.
- Kurushin D.S., Leonov E.R., Soboleva O.V. 2018. O vozmozhnom podhode k avtomatičeskoj postroeniyu denotatnogo grafa giperteksta [On a possible approach to the automatic construction of a denotation graph of hypertext]. *Informacionnaya struktura teksta*. Sb. statej. RAN. INION. M., 113–118.
- Loginov I.V., Loginova Yu.V. 2024. Sravnenie skhozhesti intellekt-kart v zadachah marketingovogo analiza [Comparison of the similarity of intelligence maps in marketing analysis tasks]. *Vestnik Kemerovskogo gosudarstvennogo universiteta. Seriya: Poli-ticheskie, sociologicheskie i ekonomicheskie nauki*, T. 9, № 3(33): 410–423. DOI 10.21603/2500-3372-2024-9-3-410-423.
- Loginova Yu. V., Loginov I.V. 2024. Primenenie semanticeskogo analiza v strategi-cheskom marketinge pri ispol'zovanii instrumenta intellekt-kart [Application of semantic analysis in strategic marketing when using the intelligence maps tool]. *Vestnik Tyumenskogo gos-udarstvennogo universiteta. Social'no-ekonomicheskie i pravovye issledovaniya*, T. 10, № 1(37): 103–123. DOI 10.21684/2411-7897-2024-10-1-103-123.
- Minaev D.V. 2022. Issledovanie kompetentnostnoj modeli obrazovatel'noj programmy na osnove intellektual'nogo analiza professional'nyh trebovaniij rynka truda [The study of the competence model of the educational program based on the intellectual analysis of professional requirements of the labor market]. *Upravlencheskoe konsul'tirovaniye*, 10: 65–83.
- Persteneva N.P., Kuchko A.Yu. 2020. Komparativnyj analiz opisanij vakansij i rabochih programm [Comparative analysis of job descriptions and work programs]. Nauka XXI veka: aktual'nye napravleniya razvitiya. 1-2: 37–42.
- Shcherbin A.A. 2024. Primenenie metodov obrabotki estestvennogo yazyka dlya resheniya zadach analiza vakansij biznes-analitikov na predmet trebovaniij k znaniyam [Application of natural language processing methods to solve the problems of analyzing business analyst vacancies for knowledge requirements]. *Obrazovanie i nauka bez granic: social'no-gumanitarnye nauki*, 23: 134–138.

- Yarullin D.V. 2019. Avtomaticheskoe postroenie professional'nogo portreta IT-specialista na osnove tekstov vakansij [Automatic construction of a professional portrait of an IT specialist based on vacancy texts]. *Filologiya v XXI veke*, 1(3): 17–21.
- Baroni M., Dinu G., Kruszewski G. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 1: 238–247.
- Blei D.M., Ng A.Y., Jordan M.I. 2003. Latent dirichlet allocation. *The Journal of machine Learning research*, 3: 993–1022.
- Camacho-Collados J., Pilehvar M.T., Navigli R. 2015. Nasari: a novel approach to a semantically-aware representation of items. In Proceedings of NAACL, 567–577.
- Collobert R., Weston J. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, 160–167. ACM.
- Deerwester S.C., Dumais S.T., Landauer Th.K., Furnas G.W., Harshman R.A. 1990. Indexing by latent semantic analysis. *JAsIs*, 41(6): 391–407.
- Fellbaum Ch. 1998. WordNet: An Electronic Lexical Database. Bradford Books.
- Gabrilovich E., Markovitch Sh. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In IJCAI, 7: 1606–1611.
- Guixiang M., Nesreen A., Theodore W., Philip Y. 2021. Deep graph similarity learning: a survey. *Data Mining and Knowledge Discovery*, 35. 10.1007/s10618-020-00733-5.
- Hassan S., Mihalcea R. 2011. Semantic relatedness using salient semantic analysis. In AAAI.
- Herlambang P., Nur R. 2021. Job Standard Parameters from Online Job Vacancy. *IPTEK Journal of Proceedings Series*, 46. 10.12962/j23546026.y2020i6.8905.
- Hughes Th., Ramage D. 2007. Lexical Semantic Relatedness with Random Graph Walks. Conference: EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, June 28-30, 2007, Prague, Czech Republic. 581–589.
- Landauer Th.K., Laham D., Rehder B., Schreiner M.E. 1997. How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In Proceedings of the 19th annual meeting of the Cognitive Science Society, 412–417. Citeseer.
- Litecky Ch., Aken A., Ahmad A., Nelson, H. 2010. Mining for Computing Jobs. *Software, IEEE*, 27. 78-85. 10.1109/MS.2009.150.
- Mikolov T., Chen K., Corrado G., Dean J. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Pennington J., Socher R., Manning Ch. 2014. Glove: Global vectors for word representation. Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014), 12: 1532–1543.
- Shalaby W., Zadrozny W. 2017. Mined semantic analysis: A new concept space model for semantic representation of textual data. 2122–2131. 10.1109/BigData.2017.8258160.
- Vanetik N., Kogan G. 2023. Job Vacancy Ranking with Sentence Embeddings, Keywords, and Named Entities. *Information*, 14: 468. 10.3390/info14080468.
- Zhang Z., Gentile A., Ciravegna F. 2013. Recent advances in methods of lexical semantic relatedness – A survey. *Natural Language Engineering*, 19. 10.1017/S1351324912000125.
- Wowczko I. 2015. Skills and Vacancy Analysis with Data Mining Techniques. *Informatics*, 2: 31–49. 10.3390/informatics2040031.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: no potential conflict of interest related to this article was reported.

Поступила в редакцию 11.06.2025

Received June 11, 2025

Поступила после рецензирования 07.09.2025

Revised September 07, 2025

Принята к публикации 10.09.2025

Accepted September 10, 2025

ИНФОРМАЦИЯ ОБ АВТОРАХ

Логинов Илья Валентинович, доктор технических наук, сотрудник Академии ФСО России, г. Орел, Россия

Щербин Алексей Алексеевич, студент, Среднерусский институт управления – филиал РАНХиГС, г. Орел, Россия

INFORMATION ABOUT THE AUTHORS

Ilia V. Loginov, Doctor of Technical Sciences, employee at the Academy of the Federal Guard Service of the Russian Federation, Orel, Russia

Aleksey A. Sherbin, Student of Mid-Russia Institute of Management, RANEPA branch in Orel, Orel, Russia