

DATA MINING КАК ИНСТРУМЕНТ МУЛЬТИМОДАЛЬНОЙ БИЗНЕС-АНАЛИТИКИ: ТРАНСФОРМАЦИЯ ЛЕКСИКИ ГОДОВЫХ ОТЧЕТОВ ГОСКОРПОРАЦИИ «РОСТЕХ»

П. А. Михненко

Московский государственный технический университет имени Н. Э. Баумана,
Москва, Россия

Цель исследования состоит в применении инструментов Data Mining для выявления латентного характера трансформации лексики годовых отчетов государственной корпорации «Ростех» и обосновании перспективности расширения методологии бизнес-анализа инструментами мультимодальной бизнес-аналитики. Выявление трендов трансформации лексики годовых отчетов осуществлялось путем выделения индикаторных свойств уникальных ключевых лексем и их контекстов. Автором использовались методы мультимодальной бизнес-аналитики и инструменты Data Mining для анализа корпоративной отчетности с целью идентификации наиболее существенных признаков и трендов трансформации бизнес-лексики предприятий и корпораций, выявление которых остается за рамками традиционного бизнес-анализа. Исходными данными являлись русскоязычные тексты годовых отчетов госкорпорации «Ростех» за 2017–2021 гг. В качестве инструментов Data Mining использовалась Workflow-модель на онлайн-платформе анализа и визуализации данных Orange Data Mining 3.3.2. Исследование позволило выявить устойчивый тренд на обновление содержания годовой отчетности корпорации в анализируемый период. Максимальную динамику устойчивого роста частотности продемонстрировали уникальные ключевые лексемы «работник», «обучение», «платформа», «трансформация», «инициатива» и «академия». Контекстный анализ позволил сделать вывод, что в основе трансформации деловой лексики лежит прирост внимания корпорации к проблемам цифровой трансформации, развитию цифровых платформ, реализации технологических и социогуманитарных стратегических инициатив, повышению квалификации работников на основе обучения в корпоративной сетевой академии. Реализация этих мероприятий требует совершенствования стратегии управления финансовыми, коррупционными, рыночными и другими рисками. Мультимодальная бизнес-аналитика является перспективным направлением научных исследований и инструментом решения прикладных экономических и организационно-управленческих задач.

Ключевые слова: годовой отчет, лексема, динамика, контекст, трансформация.

DATA-MINING AS TOOL OF MULTIMODAL BUSINESS-ANALYSIS: LEXICAL TRANSFORMATION IN ANNUAL REPORTS OF ROSTECH CORPORATION

Pavel A. Mikhnenko

Bauman Moscow State Technical University,
Moscow, Russia

The goal of the research is to study tools of Data Mining to find latent nature of lexical transformation in annual reports of the state-owned corporation 'Rostech' and to estimate perspectives of extending methodology of business-analysis by tools of multimodal business-analysis. Identification of trends in vocabulary transformation in annual reports was done by finding indicator features of unique key lexical units and their contexts. The author used methods of multimodal business-analysis and tools of Data Mining to study corporate reports in order to identify

the most important signs and trends in business-vocabulary transformation at enterprises and corporations, which is usually ignored by conventional business-analysis. Russian texts of annual reports of 'Rostech' corporation for 2017–2021 act as initial data for the research. As Data Mining tools the author used Workflow-model on the on-line platform for analysis and data visualization Orange Data Mining 3.3.2. The research showed a clear trend to renewal of the annual report content within the period of analysis. Such units as 'работник' (worker), 'обучение' (training), 'платформа' (platform), 'трансформация' (transformation), 'инициатива' (initiative) and 'академия' (academy) demonstrated maximum dynamics in frequency growth. Context analysis allowed the author to draw a conclusion that transformation in business vocabulary is based on the growing attention of the corporation to problems of digital transformation, development of digital platforms, realization of technological and socio-humanitarian strategic initiatives, raising the quality of workers' qualification by training in corporate net academy. Implementation of these measures requires upgrading strategy of managing finance, corruption, market and other risks. Multimodal business-analysis is a promising trend in academic research and a tool for resolving applied economic and organizational-managerial problems.

Keywords: annual report, lexical unit, dynamics, context, transformation.

Введение

Мультимодальная бизнес-аналитика (Multimodal Business Analysis) – анализ нефинансовой информации и данных нечисловой природы – является сегодня развивающимся и перспективным направлением научных исследований и инструментом решения прикладных экономических и организационно-управленческих задач. Имеются основания полагать, что разработка и применение методов мультимодальной бизнес-аналитики позволяют идентифицировать латентные факторы и тенденции в экономической деятельности хозяйствующих субъектов, выявление которых традиционными методами бизнес-анализа оказывается затруднительным.

Одной из актуальных задач подобных исследований является выявление целей, факторов и тенденций стратегического развития отечественных предприятий в условиях макроэкономической нестабильности и рисков путем анализа нефинансовой информации (Non-Financial Information) и нечисловых данных (Non-numeric Data), извлекаемых из медиасреды с использованием технологий интеллектуального анализа данных (Data Mining), мультимодального слияния данных (Multimodal Data Fusion), анализа тональностей (Sentiment Analysis) и др.

При решении задачи анализа направлений стратегического развития компаний приходится учитывать изменение взглядов менеджмента и общественности на прин-

ципы корпоративного управления (Environmental, Social, and Corporate Governance – ESG), отражение в корпоративных документах положений Глобальной инициативы по отчетности и раскрытия нефинансовых показателей деятельности (Global Reporting Initiative – GRI), подходы к управлению инновациями на основе данных (Data-Driven Innovations) и др.

Основная цель данного исследования состоит в выявлении латентного характера трансформации лексики годовых отчетов Государственной корпорации по содействию разработке, производству и экспорту высокотехнологичной промышленной продукции «Ростех» (ГК «Ростех») путем выделения индикаторных свойств уникальных ключевых лексем и их контекстов для идентификации факторов и тенденций развития корпорации в период 2017–2021 гг. с применением инструментов Data Mining.

ГК «Ростех» – российская государственная корпорация, созданная в 2007 г. для содействия в разработке, производстве и экспорте высокотехнологичной промышленной продукции гражданского и военного назначения. Сегодня в ее состав входит более 800 научных и производственных организаций в 60 регионах России. Ключевыми направлениями деятельности корпорации являются автомобилестроение, авиастроение, двигателестроение, металлургия, оптика, композиционные и другие современные конструкционные материалы, медицинская техника, фарма-

цевтика, промышленные биотехнологии, радиоэлектроника, производство вооружений и военной техники и др. В соответствии с актуализированной стратегией миссией корпорации является повышение качества жизни людей через создание высокотехнологичных умных продуктов.

Под трансформацией лексики в статье понимается количественное и качественное изменение словарного состава и основных контекстов наиболее частотных и динамично изменяющихся ключевых лексем русскоязычных текстов годовых отчетов корпорации.

Наиболее полный обзор проблем, связанных с мультимодальным слиянием данных (MMDF), приводится в работе Д. Лахата, Т. Адали и С. Ютгена [10]. Актуальным методом обработки экономических данных, извлекаемых из разнородных источников путем их объединения, признается машинное обучение [11]. Проблема анализа тональностей текстов экономической и финансовой тематики посвящена совместная работа К. Ахмад, Дж. Дж. Хан, Э. Хатсона, К. Кирни и С. Лю [7]. В работе М. Бейкера и Дж. Вурглера [8] описан характер влияния настроений инвесторов на доходность акций. Ученые Т. В. Теплова, Т. В. Соколова, А. Ф. Томтосов, Д. В. Бучко и Д. Д. Никулин [3] предлагают авторский индикатор настроений, влияющих на биржевые характеристики, с широким спектром прокси-переменных. В исследовании «Новости и социальные сети российских компаний: степень влияния на рынок ценных бумаг» [5] доказано, что доходность фондового рынка России объясняется не только экономическими показателями, но и тональностью официальной и неофициальной текстовой информации в медиасреде. Е. Федорова, С. Дручок и П. Дрогозов [9] выявили характер влияния нарративов, тем общественных дискуссий и новостных настроений на занижение цены IPO. В работе «О влиянии зарубежных СМИ на российский фондовый рынок: текстовый анализ» [1] авторами сделан вывод о наличии асимметрии в реакции ком-

позитного фондового индекса Мосбиржи на позитивные и негативные высказывания в СМИ.

Отдельным направлением исследований можно назвать анализ влияния характера раскрытия информации о деятельности компаний на их инвестиционную привлекательность с использованием методов и инструментов Data Mining. В работе Е. А. Федоровой, Л. Е. Хрустовой и И. С. Демина [6] эмпирически доказана значимость полноты раскрытия нефинансовой информации для формирования показателей инвестиционной привлекательности российских компаний. Анализируя ежегодные отчеты, отчеты устойчивого развития, социальные и экологические отчеты компаний, авторы исследования «Влияние нефинансовой информации на основные показатели российских компаний» [4] доказали влияние некоторых тематических групп нефинансовой отчетности на основные показатели российских компаний. В работе С. А. Кузубова и М. С. Евдокимовой [2] на основе анализа корпуса компаний в 2004–2014 гг. показано, что выпуск отчетов в соответствии с принципами GRI оказывает положительное влияние на отношение рыночной стоимости капитала к балансовой в течение года и через год после выпуска.

Обзор основных источников подтверждает тезис о перспективности исследования взаимосвязи информации нечисловой природы с экономической деятельностью хозяйствующих субъектов.

Методология исследования

Методология данного исследования состоит в применении инструментов Data Mining для анализа русскоязычных отчетов ГК «Ростех» за 2017–2021 гг. с целью идентификации наиболее существенных признаков трансформации лексики отчетов, выявление которых, как правило, остается за рамками целей и методов традиционного бизнес-анализа.

Под лексикой в статье понимается словарный состав текстов русскоязычных от-

четов – совокупность лексем. Лексемой называется слово во всех словоформах, встречающихся в отчетах и допускающее выявление контекстов его использования. Ключевой называется лексема (КЛ), частотность использования которой в отчете превышает установленный уровень. Под частотностью понимается процентная мера количества использований данной лексемы по отношению к общему количеству лексем в отчете. В данном исследовании использовался пороговый уровень – 0,02%.

Data Mining – обобщенный термин, используемый сегодня для описания сово-

купности методов обнаружения в большом объеме данных практически значимой и доступной для последующей интерпретации информации, необходимой для принятия решений в различных сферах человеческой деятельности [12]. В качестве одного из инструментов Data Mining в статье использовалась онлайн-платформа анализа и визуализации данных Orange Data Mining 3.3.2. На рис. 1 показана Workflow-модель анализа данных, составленная автором на платформе Orange.

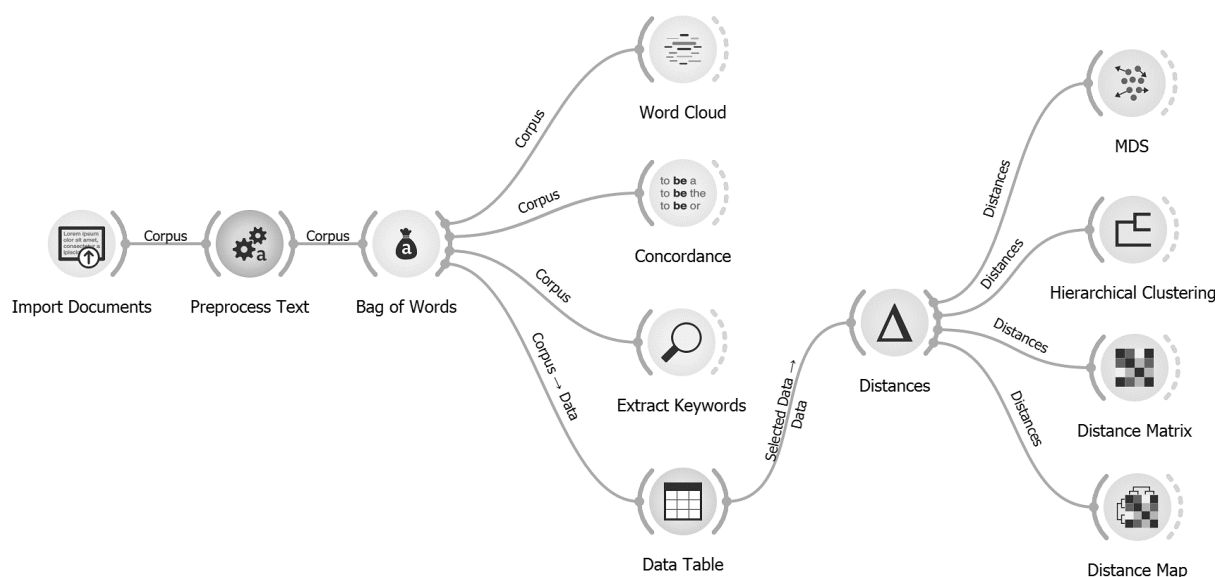


Рис. 1. Workflow-модель анализа данных Data Mining

Анализ трансформации лексики отчетов осуществлялся в объеме корпуса уникальных ключевых лексем (УКЛ). Уникальной в работе названа КЛ, зафиксированная как минимум в одном из пяти отчетов исследуемого периода.

Для количественного описания состава и динамики трансформации лексики отчетов использовались два основных показателя:

1) средняя частотность УКЛ – среднее арифметическое (по числу отчетов) величин частотности конкретной УКЛ в каждом из отчетов:

$$F = \frac{1}{N} \sum_{i=1}^N f_i,$$

где $N = 5$ – количество отчетов;

i – порядковый номер отчета (года);

f_i – частотность УКЛ в i -м отчете;

2) показатель средней динамики изменения (роста или спада) частотности УКЛ – среднее арифметическое (по числу интервалов изменения) величин абсолютного прироста частотности УКЛ за один год:

$$D = \frac{1}{N-1} \sum_{i=1}^{N-1} (f_{i+1} - f_i).$$

На рис. 2 показано статистическое распределение величин частотности УКЛ (F) с

выходом на верхнюю границу интервала 0,2%. Как видно, около 64% УКЛ принадлежат диапазону крайне низкой частотности [0,00...0,03%]. Условное ядро максимально частотных УКЛ составляют 22 лек-

семы с показателем $F > 0,2\%$. К ним, например, относятся «развитие» ($F = 0,47\%$), «организация» ($F = 0,61\%$) и «государственный» ($F = 0,81\%$).

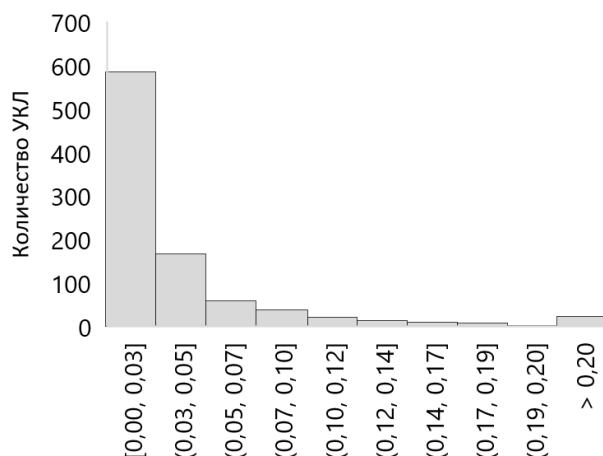


Рис. 2. Статистическое распределение величин средней частотности УКЛ

На рис. 3 показано статистическое распределение величин динамики УКЛ (D) с

выходом на верхнюю и нижнюю границы интервала 0,05% и -0,05% соответственно.

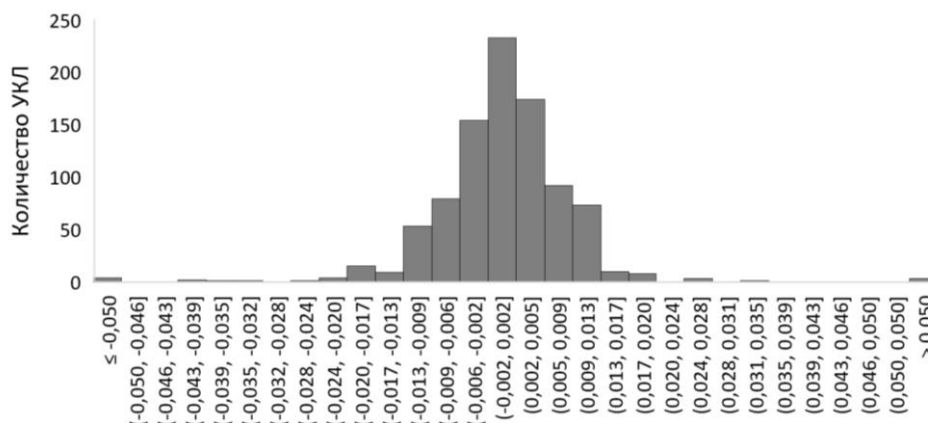


Рис. 3. Статистическое распределение показателей динамики УКЛ

Более 91% УКЛ находятся в диапазоне $D = (-0,010...0,014\%)$, соответствующем средней динамике спада/роста частотности. Для сравнения, максимальную динамику спада продемонстрировала УКЛ «государственный» с показателем $D = -0,12\%$, максимальный рост с показателем $D = 0,07\%$ характерен для УКЛ «организация».

В соответствии с целью и методологией данного исследования наибольший интерес представляют максимально динамичные и высокочастотные УКЛ как наиболее релевантные индикаторы содержательной трансформации лексики отчетов.

Для оценки степени устойчивости тренда динамики УКЛ в статье использо-

вался показатель достоверности линейной аппроксимации (R^2) ряда их частотностей по годам. Устойчивыми в работе названы УКЛ, для которых $R^2 \geq 0,6$.

На рис. 4 приведены примеры устойчивой динамики убывающей и растущей УКЛ с высокими значениями R^2 .

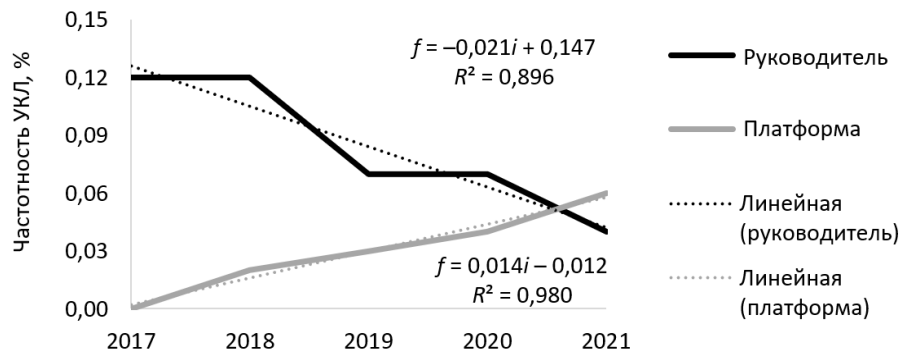


Рис. 4. Примеры устойчивой динамики убывающей и растущей УКЛ

Результаты исследования

Предварительные результаты исследования представлены в виде карты «Динамика / Средняя частотность» наиболее устойчивых УКЛ с показателем динамики $D \geq 0,02\%$ (рис. 5). На этой диаграмме используется логарифмическая шкала средней частотности УКЛ (F). Позиция каждой

лексемы определяется по середине подписи (наименованию). Чем выше расположена лексема, тем выше значение ее среднегодовой частотности. Расположение УКЛ левее вертикальной оси указывает на динамику спада частотности, справа – роста.

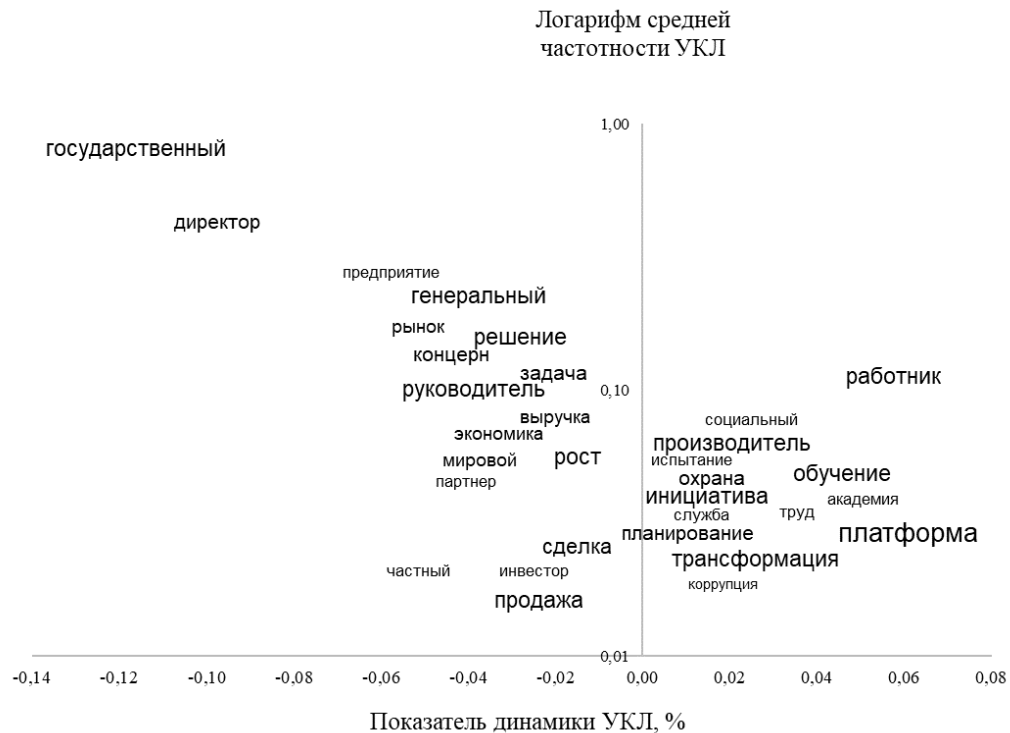


Рис. 5. Карта «Динамика / Средняя частотность» устойчивых УКЛ

Размер шрифта приближенно пропорционален величине достоверности линейной аппроксимации пятилетнего тренда ($R^2_{\text{макс}} = 0,866$, $R^2_{\text{мин}} = 0,605$) и отражает степень устойчивости тренда УКЛ.

Наиболее заметным признаком трансформации лексики отчетов ГК «Ростех» за 2017–2021 гг. стал спад частотности формальных терминов корпоративного управления (генеральный, директор, руководитель и т. п.) и коммерческой деятельности (выручка, продажа, сделка и т. п.). Этот процесс сопровождался устойчивым ростом ряда лексем (платформа, работник, обучение и др.), анализ контекстов которых представляет интерес. Средняя частотность ядра растущих лексем не превышает 0,04% с близким к этой величине темпом роста.

Как видно, максимально устойчивую ($R^2 = 0,92$) линейную динамику спада ($D = -0,12\%$) демонстрирует УКЛ «государственный», оставаясь при этом самой высокочастотной лексемой ($F = 0,81\%$) в общем корпусе УКЛ за пять лет. Среди растущих наибольшую устойчивость демонстрирует УКЛ «платформа», максимальная динамика характерна для лексем «работник», «обучение», «академия».

Имеются основания полагать, что ротация наиболее частотных лексем в отчетах отражает содержание и факторы протекающих в корпорации организационных изменений. Однако для более полной идентификации их характера требуется проведение контекстного анализа лексем, релевантных проблеме стратегических трансформаций.

В данной статье широкие лексические контексты релевантных УКЛ определялись с использованием виджета «Concordance» Workflow-модели Data Mining на основе группировки по частотности их лексического окружения в отчетах. Под широкими лексическими контекстами понимаются контексты лексемы в масштабе одного-двух предложений, в отличие от узких контекстов, выявляемых в масштабах словосочетания.

Результаты контекстного анализа представлены диаграммами типа «облако основных контекстов». Размер шрифта на диаграмме приближенно соответствует частотности контекста в общем объеме всех выявленных контекстов. Наиболее частотные из них расположены ближе к центру диаграммы и выделены жирным шрифтом.

На рис. 6 представлено облако основных контекстов УКЛ «государственный».

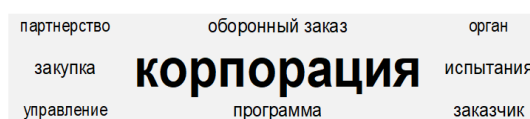


Рис. 6. Облако основных контекстов УКЛ «государственный»

Очевидно, что данная УКЛ преимущественно используется в словосочетаниях «государственная корпорация». Снижение ее частотности имеет относительный характер и обусловлено ростом частотности других лексем.

Далее рассмотрим результаты контекстного анализа выборки лексем, динамика устойчивого роста которых показана на рис. 7.

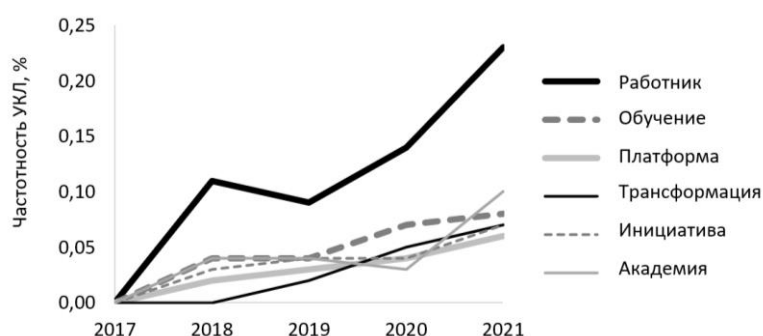


Рис. 7. Динамика выборки растущих УКЛ

Анализ облака основных контекстов УКЛ «работник» (рис. 8) показывает, что основным контекстом этой лексики на протяжении всего исследуемого периода была поддержка работников корпорации.

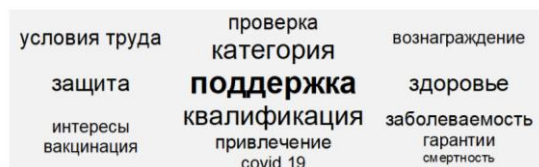


Рис. 8. Облако основных контекстов УКЛ «работник»

Этот тезис раскрывается такими контекстами, как повышение квалификации различных категорий работников, защита их интересов, улучшение условий труда и забота о сохранении здоровья. Проблема заболеваемости, смертности и вакцинации активно обсуждалась в отчетах за 2020 и 2021 гг. в связи с пандемией COVID-19.

Среди основных контекстов лексики «обучение» выделяется программа обучения (рис. 9).

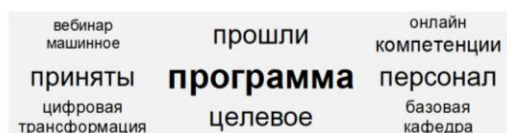


Рис. 9. Облако основных контекстов УКЛ «обучение»

Цели, задачи и результаты обучения персонала корпорации обсуждаются в контексте приема и прохождения обучения, развития компетенций, в том числе в рамках цифровой трансформации. В отчетах 2019–2021 гг. обсуждались результаты целевой подготовки специалистов на базовых кафедрах вузов.

УКЛ «платформа» представлена широким спектром технических контекстов с заметным доминированием словосочетания «цифровая платформа», отражающим внимание корпорации к проблеме цифровой трансформации (рис. 10), что подтверждается одноименным доминирующим

контекстом лексики «трансформация» (рис. 11).

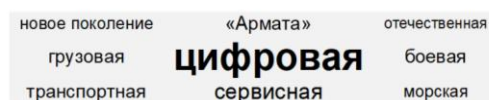


Рис. 10. Облако основных контекстов УКЛ «платформа»

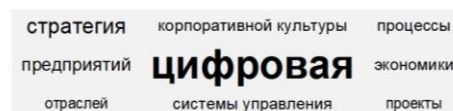


Рис. 11. Облако основных контекстов УКЛ «трансформация»

Кроме того, в отчетах корпорации обсуждаются основные положения стратегии трансформации, а также цели и результаты трансформации корпоративной культуры, системы управления и др.

Термин «стратегический» доминирует в контексте реализации стратегических инициатив, в том числе технологической, социокультурной, образовательной и других направленностей (рис. 12).

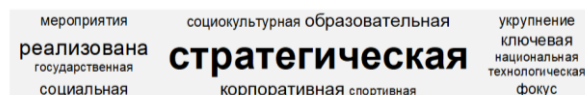


Рис. 12. Облако основных контекстов УКЛ «инициатива»

УКЛ «академия» может быть отнесена к мультиконтекстным, однако ее доминирующим контекстом является обсуждение целей и результатов работы корпоративной сетевой академии ГК «Ростех» – центра экспертизы госкорпорации в сфере развития человеческого капитала, главная цель которой – помощь сотрудникам в получении знаний и навыков, необходимых для реализации стратегических целей корпорации (рис. 13).

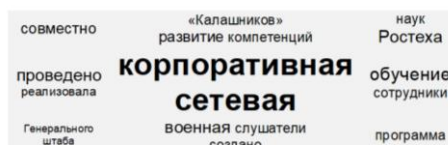


Рис. 13. Облако основных контекстов УКЛ «академия»

Задачи и результаты обучения сотрудников корпорации обсуждаются также в контексте совместных программ с российскими вузами.

Отдельного внимания в данном исследовании заслуживает УКЛ «риск», не во-

шедшая в выборку (см. рис. 4) по причине отсутствия устойчивого линейного тренда. Однако скачкообразный двукратный рост частотности этой лексемы в отчете 2021 г. требует отдельного контекстного анализа (рис. 14).

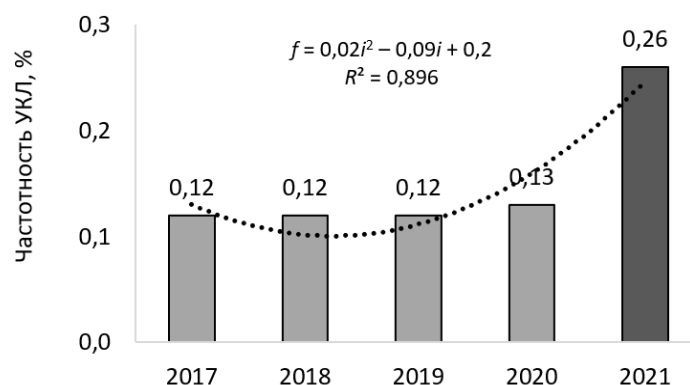


Рис. 14. Динамика УКЛ «риск»

На рис. 15 представлено облако контекстов УКЛ «риск», основным из которых является «управление рисками», заключающийся в разработке стратегии управления финансовыми, коррупционными, рыночными и другими рисками, их своевременного выявления, оценки параметров и минимизации влияния на деятельность корпорации.

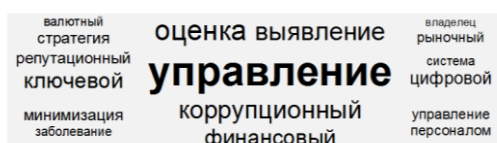


Рис. 15. Облако основных контекстов УКЛ «риск»

Заключение

В статье представлены методология и основные результаты исследования латентных тенденций трансформации лексики годовых отчетов ГК «Ростех» с использованием инструментов Data Mining. Результаты исследования позволили выявить устойчивый тренд на обновление содержания отчетов корпорации в 2017–

2021 гг., в основе которого лежит прирост внимания к проблемам цифровой трансформации, обуславливающей необходимость развития цифровых платформ; реализации стратегических инициатив технологической и социогуманитарной направленности; повышения квалификации работников на основе программ обучения в корпоративной сетевой академии. Реализация этих мероприятий требует от корпорации совершенствования стратегии управления финансовыми, коррупционными, рыночными и другими рисками.

Идентификация выявленных трендов, как правило, лежит за рамками целей и методологии традиционного бизнес-анализа, причем его базовый инструментарий не позволяет в полной мере решить эту задачу. Расширение методологии организационно-экономического и финансового анализа методами и инструментами мультимодальной бизнес-аналитики приведет к повышению качества выводов о результатах их хозяйственной деятельности и прогнозов развития предприятий и корпораций.

Список литературы

1. Афанасьев Д. О., Федорова Е. А., Демин И. С., Рогов О. Ю. О влиянии зарубежных СМИ на российский фондовый рынок: текстовый анализ // Экономика и математические методы. – 2020. – Т. 56. – № 1. – С. 153–165.
2. Кузубов С. А., Евдокимова М. С. Повышает ли стоимость компании публикация нефинансовых отчетов по стандартам GRI (на примере стран БРИКС)? // Учет. Анализ. Аудит. – 2017. – № 2. – С. 28–36.
3. Теплова Т. В., Соколова Т. В., Томтосов А. Ф., Бучко Д. В., Никулин Д. Д. Сентимент частных инвесторов в объяснении различий в биржевых характеристиках акций российского рынка // Журнал Новой экономической ассоциации. – 2022. – № 1 (53). – С. 53–84.
4. Федорова Е. А., Афанасьев Д. О., Нерсисян Р. Г., Ледяева С. В. Влияние нефинансовой информации на основные показатели российских компаний // Журнал Новой экономической ассоциации. – 2020. – № 2 (46). – С. 73–96.
5. Федорова Е. А., Пыльцин И. В., Ковальчук Ю. А., Дрогвоз П. А. Новости и социальные сети российских компаний: степень влияния на рынок ценных бумаг // Журнал Новой экономической ассоциации. – 2022. – № 1 (53). – С. 32–52.
6. Федорова Е. А., Хрустова Л. Е., Демин И. С. Влияние качества раскрытия нефинансовой информации российскими компаниями на их инвестиционную привлекательность // Российский журнал менеджмента. – 2020. – Т. 18. – № 1. – С. 51–72.
7. Ahmad K., Han J. G., Hutson E., Kearney C., Liu S. Media-Expressed Negative Tone and Firm-Level Stock Returns // Journal of Corporate Finance. – 2015. – N 37. – P. 152–172.
8. Baker M., Wurgler J. Investor Sentiment and the Cross-Section of Stock Returns // The Journal of Finance. – 2006. – Vol. LXI. – N 4. – P. 1645–1680.
9. Fedorova E., Druchok S., Drogovoz P. Impact of News Sentiment and Topics on IPO Underpricing: US Evidence // International Journal of Accounting & Information Management. – 2022. – Vol. 30. – N 1. – P. 73–94.
10. Lahat D., Adali T., Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects // Proceedings of the IEEE. – 2015. – N 103 (9). – P. 1449–1477.
11. Li M., Wang F., Jia X. et al. Multi-Source Data Fusion for Economic Data Analysis // Neural Computing and Applications. – 2021. – N 33. – P. 4729–4739.
12. Piatetsky-Shapiro G. Data Mining and Knowledge Discovery 1996 to 2005: Overcoming the Hype and Moving from “University” to “Business” and “Analytics” // Data Mining and Knowledge Discovery. – 2007. – N 15. – P. 99–105.

References

1. Afanasev D. O., Fedorova E. A., Demin I. S., Rogov O. Yu. O vliyaniy zarubezhnykh SMI na rossiyskiy fondovyy rynek: tekstovyy analiz [On the Influence of Foreign Media on the Russian Stock Market: Text Analysis]. *Ekonomika i matematicheskie metody* [Economics and Mathematical Methods], 2020, Vol. 56, No. 1, pp. 153–165. (In Russ.).
2. Kuzubov S. A., Evdokimova M. S. Povyshaet li stoimost kompanii publikatsiya nefinansovykh otchetov po standartam GRI (na primere stran BRIKS)? [Does the Company Value Increase through the Publication of Non-Financial Reports under GRI guidelines? (On the Example of BRICS Countries)?]. *Uchet. Analiz. Audit* [Accounting. Analysis. Auditing], 2017, No. 2, pp. 28–36. (In Russ.).
3. Teplova T. V., Sokolova T. V., Tomtosov A. F., Buchko D. V., Nikulin D. D. Sentiment chastnykh investorov v obyasnении razlichiy v birzhevyykh kharakteristikakh aktsiy rossiyskogo rynka [The Sentiment of Private Investors in Explaining the Differences in the Trade

Characteristics of the Russian Market Stocks]. *Zhurnal Novoy ekonomicheskoy assotsiatsii* [Journal of the New Economic Association], 2022, No. 1 (53), pp. 53–84. (In Russ.).

4. Fedorova E. A., Afanasev D. O., Nersesyan R. G., Ledyeva S. V. Vliyanie nefinansovoy informatsii na osnovnye pokazateli rossiyskikh kompaniy [Impact of Non-Financial Information on Key Financial Indicators of Russian Companies]. *Zhurnal Novoy ekonomicheskoy assotsiatsii* [Journal of the New Economic Association], 2020, No. 2 (46), pp. 73–96. (In Russ.).

5. Fedorova E. A., Pylytsin I. V., Kovalchuk Yu. A., Drogovoz P. A. Novosti i sotsialnye seti rossiyskikh kompaniy: stepen vliyaniya na rynek tsennykh bumag [News and Social Networks of Russian Companies: Degree of Influence on the Securities Market]. *Zhurnal Novoy ekonomicheskoy assotsiatsii* [Journal of the New Economic Association], 2022, No. 1 (53), pp. 32–52. (In Russ.).

6. Fedorova E. A., Khrustova L. E., Demin I. S. Vliyanie kachestva raskrytiya nefinansovoy informatsii rossiyskimi kompaniyami na ikh investitsionnyuyu privlekatel'nost' [Completeness of Non-Financial Disclosure by Russian Companies: The Influence on Investment Attractiveness]. *Rossiyskiy zhurnal menedzhmenta* [Russian Management Journal], 2020, Vol. 18, No. 1, pp. 51–72. (In Russ.).

7. Ahmad K., Han J. G., Hutson E., Kearney C., Liu S. Media-Expressed Negative Tone and Firm-Level Stock Returns. *Journal of Corporate Finance*, 2015, No. 37, pp. 152–172.

8. Baker M., Wurgler J. Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 2006, Vol. LXI, No. 4, pp. 1645–1680.

9. Fedorova E., Druchok S., Drogovoz P. Impact of News Sentiment and Topics on IPO Underpricing: US Evidence. *International Journal of Accounting & Information Management*, 2022, Vol. 30, No. 1, pp. 73–94.

10. Lahat D., Adali T., Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE*, 2015, No. 103 (9), pp. 1449–1477.

11. Li M., Wang F., Jia X. et al. Multi-Source Data Fusion for Economic Data Analysis. *Neural Computing and Applications*, 2021, No. 33, pp. 4729–4739.

12. Piatetsky-Shapiro G. Data Mining and Knowledge Discovery 1996 to 2005: Overcoming the Hype and Moving from “University” to “Business” and “Analytics”. *Data Mining and Knowledge Discovery*, 2007, No. 15, pp. 99–105.

Сведения об авторе

Павел Александрович Михненко

доктор экономических наук, доцент,
профессор кафедры предпринимательства
и внешнеэкономической деятельности
МГТУ им. Н. Э. Баумана.
Адрес: ФГБОУ ВО «Московский
государственный технический университет
имени Н. Э. Баумана (национальный
исследовательский университет)»,
105005, Москва, 2-я Бауманская ул., д. 5, стр. 1.
E-mail: pmihnenko@bmstu.ru
ORCID: 0000-0002-1766-8029

Information about the author

Pavel A. Mikhnenko

Doctor of Economics,
Assistant Professor,
Professor of the Department
for Entrepreneurship and Foreign Economic
Activity of the BMSTU.
Address: Bauman Moscow State Technical
University, 1 building,
5 2nd Baumanskaya Str., Moscow,
105005, Russian Federation.
E-mail: pmihnenko@bmstu.ru
ORCID: 0000-0002-1766-8029