

DOI: 10.29141/2218-5003-2022-13-5-2

EDN: DLNXWM

JEL Classification: C51, D91, O2

## Трансформация деловой лексики годовых отчетов крупнейших российских компаний: Data Mining

П.А. Михненко<sup>1</sup><sup>1</sup> МГТУ им. Н.Э. Баумана, г. Москва, РФ

**Аннотация.** Одним из перспективных направлений бизнес-анализа является разработка новых методов и инструментов учета нефинансовой и нечисловой информации. В этой области имеются научный задел и практические решения, представленные в публикациях по данной тематике. Однако недостаточно глубоко проработаны вопросы, связанные с анализом динамики трансформации деловой лексики компаний. Исследование направлено на выявление и интерпретацию латентной информации, отражающей стратегические ориентиры и условия экономического развития российских компаний. Методологическим фундаментом работы послужили основные положения нарративной экономики и концепции мультимодальной бизнес-аналитики, включающей систему научно-практических методов анализа деятельности хозяйствующих субъектов за счет использования данных из разнородных источников. Применялись методы и инструменты анализа и систематизации больших объемов текстовой информации – Data Mining. Информационную базу сформировали годовые отчеты крупнейших российских компаний за 2018–2020 гг. В качестве основных показателей трансформации деловой лексики рассматривались частотность уникальных ключевых лексем (УКЛ) и динамика ее изменения, а также основные контексты УКЛ, релевантные проблеме развития. Выявлены заметные изменения рассмотренной лексики: спад активности в освещении формальных аспектов хозяйственной деятельности и рост интенсивности обсуждения направлений развития в условиях риска. Установлено, что наиболее явно эти тенденции проявились в отчетности металлургических и энергетических компаний. Результаты исследования могут служить основой для повышения аналитической и прогнозной результативности современного бизнес-анализа.

**Ключевые слова:** Data Mining; текстовый анализ; российские компании; годовой отчет; лексема; деловая лексика.

**Информация о статье:** поступила 8 июня 2022 г.; доработана 11 августа 2022 г.; одобрена 24 августа 2022 г.

**Ссылка для цитирования:** Михненко П.А. (2022). Трансформация деловой лексики годовых отчетов крупнейших российских компаний: Data Mining // Управленец. Т. 13, № 5. С. 17–33. DOI: 10.29141/2218-5003-2022-13-5-2. EDN: DLNXWM.

## Transformation of the largest Russian companies' business vocabulary in annual reports: Data Mining

Pavel A. Mikhnenko<sup>1</sup><sup>1</sup> Bauman University, Moscow, Russia

**Abstract.** One of the promising areas of business analysis is the development of new methods and tools for accounting of non-financial and non-numeric information. There is a significant number of theoretical and practical solutions in this field; however, the issues of the transformation dynamics of companies' business vocabulary need to be studied more extensively. The article aims to identify and interpret latent information reflecting strategic guidelines and conditions for the economic development of Russian enterprises. The methodology of the study is based on the concepts of narrative economics and multimodal business analytics, which is a system of scientific-practical methods for analyzing the activities of economic entities through the use of data from heterogeneous sources. The Data Mining methods and tools for analyzing and systematizing large volumes of textual information were used. The data for research were retrieved from the annual reports of the largest Russian companies for 2018–2020. Among the main indicators of the business vocabulary transformation considered in the paper are the occurrence of unique key tokens (UKTs) and the dynamics of its change, as well as the main contexts of UKTs relevant to the problem of development. The findings indicate noticeable changes in the vocabulary of Russian companies' annual reports, such as a decline in covering formal aspects of economic activity and a growing debate on the development in the presence of risk. It is shown that these trends were most clearly manifested in the reports of metallurgical and energy enterprises. The research results can serve as a basis for enhancing the analytical and predictive effectiveness of modern business analysis.

**Keywords:** Data Mining; text analysis; Russian companies; annual report; token; business vocabulary.

**Article info:** received June 8, 2022; received in revised form August 11, 2022; accepted August 24, 2022

**For citation:** Mikhnenko P.A. (2022). Transformation of the largest Russian companies' business vocabulary in annual reports: Data Mining. *Upravlenets/The Manager*, vol. 13, no. 5, pp. 17–33. DOI: 10.29141/2218-5003-2022-13-5-2. EDN: DLNXWM.

## ВВЕДЕНИЕ

Научные исследования, связанные с анализом нефинансовой и нечисловой информации при решении прикладных экономических задач, становятся все более популярны среди отечественных и зарубежных экономистов [Пятов и др., 2018]. Одним из актуальных направлений является разработка методов и инструментов мультимодальной бизнес-аналитики (Multimodal Business Analysis), предназначенных для идентификации латентных факторов и тенденций экономической деятельности хозяйствующих субъектов [Yager, 2004; Lahat, Adali, Jutten, 2015].

Использование этих методов в экономических исследованиях предусматривает расширение состава и контекстное структурирование факторов социально-экономического развития посредством учета нефинансовой информации (Non-financial Information) и нечисловых данных (Non-numeric Data), извлекаемых из релевантных текстовых и иных источников с использованием технологий мультимодального слияния данных (Multimodal Data Fusion, MMDF) и анализа тональностей (Sentiment Analysis) и др.

В процессе идентификации факторов и направлений стратегического развития компаний имеет смысл учитывать такие тенденции, как управление инновациями на основе данных (Data-Driven Innovations), отражение в корпоративных документах принципов экологического, социального и корпоративного управления (Environmental, Social and Corporate Governance – ESG) [Батаева, Кокурина, Карпов, 2021], использование в годовых отчетах положений Глобальной инициативы по отчетности (Global Reporting Initiative, GRI) [Кузубов, Евдокимова, 2017] и др. Теоретическую базу этой области исследований составляют институциональная экономическая теория, сигнальная теория (Signaling Theory), концепция поведенческой экономики (Behavioral Economics), нарративная экономика (Narrative Economics) [Ширяев, Курышева, Вольчик, 2021], методы моделирования производственно-финансовых процессов и др.

В зависимости от структуры и характера источников и объемов анализируемых данных могут использоваться технологии интеллектуального анализа данных (Data Mining) и мультимодального глубокого обучения (Multimodal Deep Learning), инструменты анализа больших данных (Big Data), кластерный и регрессионный анализ, краудсорсинг (Crowdsourcing) – категоризация и пополнение данных с помощью сторонних лиц и др. [Yager, 2004; Siltaoja, 2006; Adams, Frost, 2008 и др.]

Data Mining – обобщенный термин, используемый сегодня для описания совокупности методов обнаружения в большом объеме данных практически значимой и доступной для интерпретации информации, необходимой для принятия решений в различных сферах человеческой деятельности [Berry, Linoff, 2004; Piatetsky-Shapiro, 2007].

Цель исследования – выявление динамики и содержания трансформации лексики годовых отчетов публичных российских компаний (далее – отчеты) путем идентификации индикаторных свойств ключевых лексем и их контекстов для выделения наиболее общих признаков и тенденций развития. Под трансформацией лексики в работе понимается изменение в отчетах словарного состава и контекстов максимально частотных лексем.

Научная новизна и практическая значимость исследования состоят в развитии методологии бизнес-анализа предприятий за счет использования новых методов и инструментов анализа данных нечисловой природы. Его оригинальность заключается в идентификации динамики частотности и основных контекстов лексем, отражающих проблемы организационных изменений и стратегического развития российских компаний.

## ОБЗОР ЛИТЕРАТУРЫ И ГИПОТЕЗА ИССЛЕДОВАНИЯ

**Обзор основных публикаций по тематике исследования.** Проблемам анализа больших объемов данных нечисловой природы посвящено значительное количество научных трудов как в России, так и за рубежом. Это направление исследований можно признать одним из наиболее перспективных подходов к выявлению информации о деятельности компаний за пределами корпуса данных финансовой отчетности. Для решения данной задачи применяются различные методы и инструменты анализа [Davis, Searcy, 2010; Skouloudis, Evangelinos, Kourmousis, 2010; Brennan, Merkl-Davies, 2013; Lahat, Adali, Jutten, 2015].

Наиболее полный обзор проблем, связанных с мультимодальным слиянием данных в различных дисциплинах, приводится в работе [Lahat, Adali, Jutten, 2015]. Одним из актуальных методов обработки экономических данных, извлекаемых из разнородных источников путем их слияния, признается машинное обучение [Li, Wang, Jia et al., 2021].

В последние годы все большую популярность набирают методы анализа экономических нарративов. Исследователи продемонстрировали связь нарративного анализа с программой институциональной экономической теории: на основе рассмотрения разнородных информационных источников периода 2010–2021 гг. были выявлены распространенные нарративы об особенностях институционального регулирования российской инновационной системы, затрудняющих коммерциализацию инноваций [Ширяев, Курышева, Вольчик, 2021].

Проблемам анализа тональностей текстов экономической и финансовой тематики посвящен обширный корпус работ (см., например: [Ahmad et al., 2015; Tschopp, Huefner, 2015; Renault, 2017]). Описан характер влияния настроений инвесторов на доходность раз-

ных видов акций [Baker, Wurgler, 2006]. Доказано, что доходность фондового рынка России объясняется как экономическими показателями, так и тональностью официальной (новости РБК) и неофициальной (твиты компаний) текстовой информации [Федорова и др., 2022].

Публичная нефинансовая отчетность играет существенную роль в повышении прозрачности деятельности компаний, которая является показателем высоких стандартов управления и инвестиционной привлекательности [Корнилова, Никонов, 2017]. Основными причинами роста популярности данной отчетности являются использование ее как инструмента самопрезентации, поиск конкурентных преимуществ и стремление следовать концепции корпоративной социальной ответственности [Пятов и др., 2018]. Указанная концепция предполагает, что мотив прибыли следует рассматривать в контексте социальной ответственности бизнеса. При этом наблюдается рост интереса к факторам устойчивости со стороны инвесторов, которые изучают публикуемые организациями отчеты о развитии наряду с финансовой информацией [Калабихина, Крикунов, 2018]. Компании, уделяющие внимание корпоративной устойчивости, реагирующие на глобальные тенденции и национальные приоритеты, представляют более полные данные для участия в ESG-рейтингах и индексах, на которые ориентируются кредиторы и инвесторы [Батаева, Кокурина, Карпов, 2021].

В одной из работ проанализированы такие источники нефинансовой информации, как ежегодные отчеты, отчеты устойчивого развития, социальные и экологические отчеты, отчеты по стандартам GRI российских публичных компаний за период 2010–2018 гг. Результаты исследования показали существенное влияние ряда тематических групп нефинансовой отчетности на основные показатели российских компаний, подтвердив важность раскрытия информации об экологии для рыночных агентов и управленческой информации для внутренних заинтересованных сторон [Федорова и др., 2020].

Исследователи используют контент-анализ и анализ тональностей для выявления потенциальных факторов, определяющих объемы текста и тональности обращений президентов компаний к акционерам [Bonsón, Perea, Azevedo, 2021]. Количественные методы контент-анализа информации, представленной в нефинансовых отчетах, широко применяются для выявления факторов и трендов устойчивого развития компаний [Asif et al., 2013; Saber, Weber, 2019].

Авторы предлагают индикатор настроений, влияющих на биржевые характеристики, с широким спектром прокси-переменных, и доказывают, что увеличение числа сообщений как позитивной, так и негативной тональности способствует росту торговой активности, однако воздействие последних более су-

щественно [Теплова и др., 2022]. Зафиксировано усиление влияния тональности новостей в зарубежных СМИ на поведение участников композитного фондового индекса Мосбиржи и его динамику после введения санкций, и сделан вывод об асимметрии в реакции рынка на позитивные и негативные высказывания в СМИ [Афанасьев и др., 2020]. Текстологический анализ финансовых новостей с использованием распределения Дирихле, методов BERT и предварительного обучения позволил выявить характер влияния нарративов, тематики общественных дискуссий и новостных настроений на занижение цены IPO [Fedorova, Druchok, Drogovoz, 2022].

К отдельному направлению исследований можно отнести изучение воздействия способов и качества раскрытия нефинансовой информации о работе компаний на их инвестиционную привлекательность. Анализ деятельности группы компаний в период 2004–2014 гг. показал, что выпуск отчетов в соответствии с принципами GRI положительно влияет на отношение рыночной стоимости капитала к балансовой в течение года и через год после выпуска [Кузубов, Евдокимова, 2017]. С помощью построения панельной регрессии была эмпирически доказана значимость полноты раскрытия нефинансовой информации, в том числе по критериям стандартов GRI, для формирования инвестиционной привлекательности российских компаний в период 2009–2018 гг. [Федорова, Хрустова, Демин, 2020].

Нефинансовая отчетность отечественных предприятий металлургической и горнодобывающей отраслей рассматривалась в качестве ключевого инструмента мониторинга реализации стандартов достойного труда. Обнаружена положительная связь между полнотой раскрытия информации о достойном труде и финансовыми показателями деятельности компаний, а также производительностью труда [Нехода и др., 2022].

Обзор публикаций позволяет сделать вывод об актуальности изучения взаимосвязи информации нечисловой природы с экономической деятельностью хозяйствующих субъектов и о перспективности выявления латентных факторов и тенденций организационных изменений и стратегического развития с учетом динамики трансформации деловой лексики предприятий. Несмотря на то, что ряд авторов обращается к проблеме динамики изучаемых показателей [Lahat, Adali, Jutten, 2015; Федорова и др., 2020; Теплова и др., 2022; Федорова и др., 2022], сегодня она является наименее проработанной, имеет перспективы развития и представляет большой научно-практический интерес.

**Гипотеза исследования и основное допущение метода.** С учетом цели исследования и результатов анализа публикаций была выдвинута гипотеза о том, что Data Mining ключевых лексем в годовых отчетах компаний позволяет выявить устойчивую динамику

изменения их частотности, которая совместно с итогами анализа основных контекстов их использования указывает на наличие явных или латентных факторов и тенденций организационных изменений и стратегического развития. Предполагается, что динамика и основные контексты ключевых лексем обладают индикаторными свойствами, позволяющими считать их перспективными показателями нечисловой природы в рамках применения инструментария Data Mining для осуществления многоаспектного организационно-экономического анализа деятельности компаний вне зависимости от их отраслевой принадлежности и других признаков. Эти показатели отражают наиболее общие, в том числе латентные, тенденции развития, выявление которых традиционными методами относится к нетривиальным задачам.

Основным допущением метода является предположение о том, что тексты отчетов достоверно отражают состояние дел в компаниях, а сочетание роста или спада частотности контекстно взаимосвязанных лексем свидетельствует о возрастании или снижении внимания менеджмента к соответствующим проблемам развития компаний.

## МАТЕРИАЛЫ И МЕТОДЫ

**Общая схема и ключевые понятия исследования.** Исследование предусматривало проведение текстового анализа (Data Mining) русскоязычных годовых отчетов крупнейших российских компаний с привлечением методов и инструментов анализа больших объемов слабоструктурированных данных, кластерного и регрессионного анализа. Объектом исследования послужил корпус отчетов за 2018–2020 гг. российских компаний, вошедших в рейтинг Forbes 2021 г. (табл. 1).

*Лексика* в данной работе понимается как словарный состав русскоязычных отчетов – совокупность лексем, допускающих анализ частотности и основных контекстов их использования, а *лексема* (Token) – как слово во всех его словоформах, встречающееся в текстах с лексическими значениями, позволяющими выявлять их контексты. Например, лексема «проект» объединяет слова «проект», «проекты», «проекту», «проектом» и др. и может употребляться в контекстах «Предприятие реализует инновационный проект...», «В проекте применено современное оборудование...» и др.

*Ключевой лексемой* (КЛ) (Key Token) называется лексема, частотность использования которой в годо-

Таблица 1 – Выборка исследования  
Table 1 – Study sample

Компания	Данные за 2021 г., млрд долл.		
	Выручка	Активы	Рыночная стоимость
En+ Group	10,40	22,10	5,90
X5 Group	27,40	15,90	8,60
АФК «Система»	9,60	19,20	4,70
Банк ВТБ	17,10	245,30	8,10
Газпром	90,50	294,90	73,50
Группа НЛМК	9,20	9,90	20,90
Интер РАО	13,60	12,30	4,80
Лукойл	71,80	81,50	52,30
Магнит	21,50	13,00	6,90
Магнитогорский металлургический комбинат	6,40	7,60	10,10
Московская биржа	0,70	66,70	5,20
Новатэк	9,60	27,90	58,40
Норникель	15,70	20,70	53,20
ПИК	2,70	8,90	8,30
Полюс	5,00	7,30	26,70
Роснефть	70,80	207,50	77,70
Россети	13,80	35,20	3,80
Сбербанк России	47,30	486,90	85,70
Северсталь	7,30	7,50	19,60
Сургутнефтегаз	18,80	79,40	16,70
Татнефть	10,20	17,10	17,30

Составлено по данным официального сайта Forbes (<https://www.forbes.ru>).

*Примечание.* Помимо указанных в таблице в рейтинг вошли еще три ПАО: «Транснефть», «Русгидро» и Московский кредитный банк, однако отсутствие в открытом доступе отчетов этих обществ за анализируемый период не позволило включить их в выборку исследования.



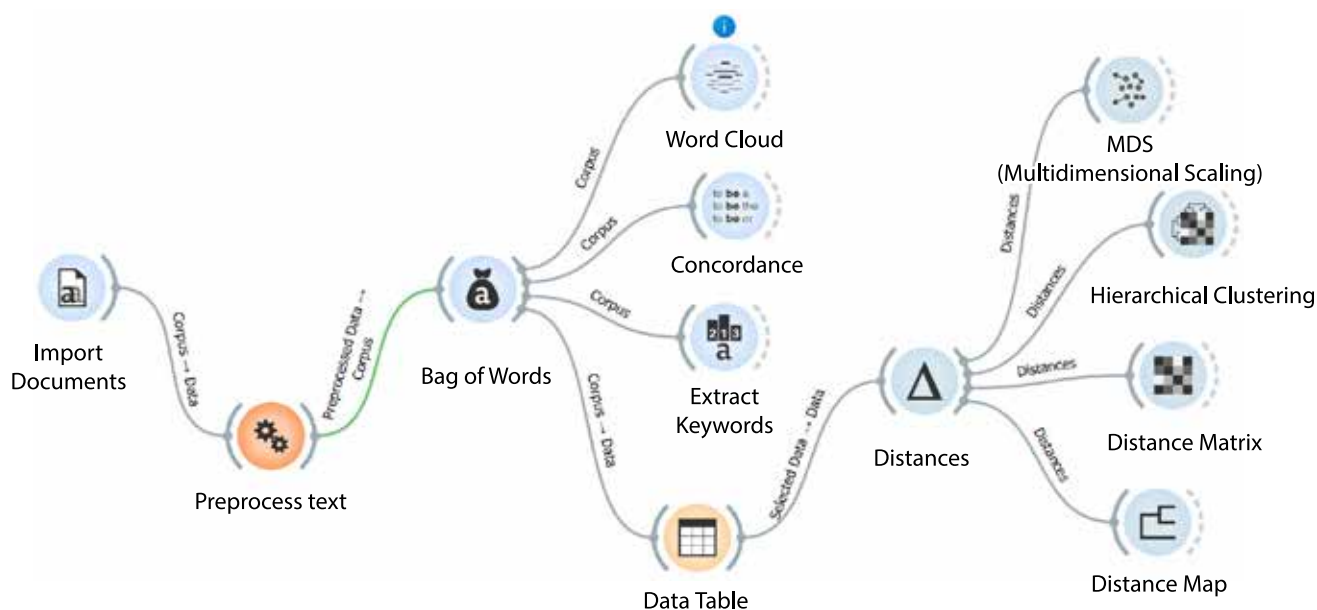


Рис. 2. Анализ данных нечисловой природы: модель Workflow

Fig. 2. Workflow model of non-numeric data analysis

$$\Delta f_{ij} = f_{ij} - f_{i,j-1}; \Delta f_{i,j+1} = f_{i,j+1} - f_{ij} \quad (2)$$

где  $f_{ij} = n_{ij}/m_{ij} \times 100\%$ ;  $n_{ij}$  – общее количество лексем в ГО  $i$ -й компании за  $j$ -й год;  $m_{ij}$  – количество использований анализируемой лексемы в отчетах  $i$ -й компании за  $j$ -й год.

Помимо оценки динамики важным показателем КЛ является ее среднегодовая частотность:

$$\bar{f}_i = \frac{1}{J} \sum_{j=1}^J f_{ij}, \quad (3)$$

где  $J$  – количество лет анализируемого периода.

Например, годовым частотностям КЛ «развитие» в корпусе отчетов ПАО «Лукойл» соответствуют значения среднегодовой частотности  $\bar{f}_i = 0,53\%$  и динамики КЛ  $d_i = 0,28\%$  (рост); КЛ «программа» –  $\bar{f}_i = 0,23\%$ ,  $d_i = 0,006\%$  (стабильность); КЛ «сумма» –  $\bar{f}_i = 0,16\%$ ,  $d_i = -0,16\%$  (спад).

Фильтрация дубликатов формулировок КЛ позволила сформировать корпус уникальных ключевых лексем (УКЛ) (Unique Key Token, УКТ), представляющий собой совокупность неповторяющихся формулировок КЛ ( $x_i$ ), каждая из которых описывается кортежем  $\{x_i, f_{i,j-1}, f_{ij}, f_{i,j+1}, \bar{f}_i, d_i\}$ , полностью характеризующим эту лексему в рамках корпуса УКЛ конкретной компании за три года.

Сортировка УКЛ по значениям  $\bar{f}_i$  и  $d_i$  позволила провести предварительный анализ корпуса УКЛ. Например, среди лексем ПАО «Сбербанк» максимальный темп роста  $d_i = 0,52\%$  продемонстрировал термин «Сбер» ( $\bar{f}_i = 0,35\%$ ), в то время как понятие «банк»<sup>1</sup> показало максимальную динамику спада ( $d_i = -0,24\%$ ) с высокой среднегодовой частотностью ( $\bar{f}_i = 1,00\%$ ).

<sup>1</sup> В данном исследовании УКЛ «банк» намеренно не была включена в корпус стоп-слов для выявления динамики указанного замещения наиболее частотной лексемы ПАО «Сбербанк».

Причем в данном случае аналитический интерес представляют не только величины среднегодовой частотности и динамики УКЛ, но и годовые значения частотности. Например, заметный спад частотности УКЛ «банк» пришелся лишь на 2020 г., что сопровождается появлением в этом году новой УКЛ «Сбер», отличающейся высокой частотностью в связи с началом ребрендинга банка. Соотношение динамики этих двух УКЛ может указывать на начало процесса переориентации представлений о компании с понятия «банк» на понятие «экосистема Сбер». При этом максимальную среднегодовую частотность  $\bar{f}_i = 0,74 \dots 0,91\%$  в корпусе УКЛ ПАО «Сбербанк» сохраняют традиционные для большинства корпораций понятия «управление», «совет», «корпоративный», «клиент» и т. п. с близкой к нулю динамикой ( $d_i = -0,04 \dots +0,01\%$ ), свидетельствующей о стабильности их использования в течение анализируемого периода.

В соответствии с целью данного исследования наибольший интерес представляют максимально динамичные и высокочастотные УКЛ как наиболее релевантные индикаторы содержательной трансформации лексики отчетов. На рис. 3 показано, насколько различна динамика УКЛ в отчетах ПАО «Лукойл».

Для оценки степени устойчивости тренда динамики УКЛ в работе использовался показатель достоверности линейной аппроксимации  $R^2$  ряда их частотностей по годам. Устойчивыми названы УКЛ, для которых  $R^2 \geq 0,7$ .

**Предварительный анализ и структура данных.** Целью предварительного анализа отчетов было выявление общих структурных признаков УКЛ, а также селекция наиболее динамично убывающих и растущих УКЛ. На рис. 4 показана средняя динамика спада и роста лексем по всем компаниям за 2018–2020 гг.

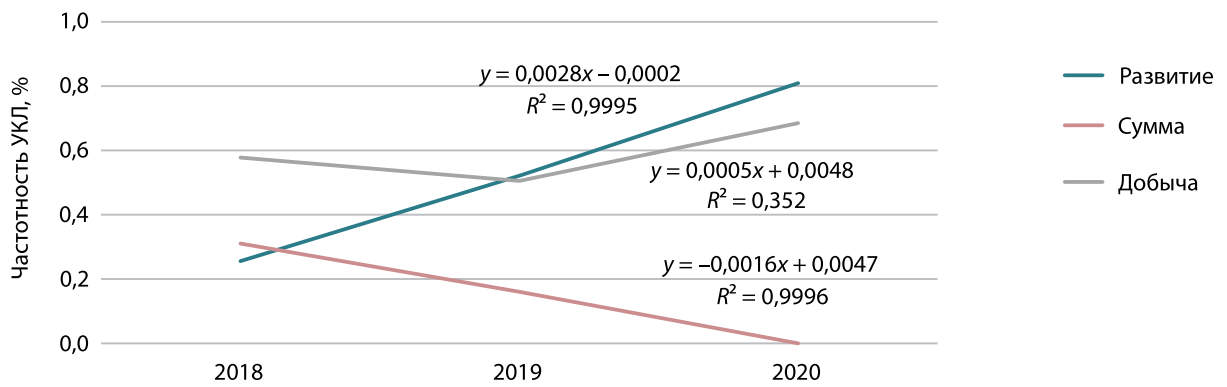


Рис. 3. Примеры динамики УКЛ в годовых отчетах ПАО «Лукойл», 2018–2020

Fig. 3. Dynamics of UKTs in PAO Lukoil annual reports in 2018–2020

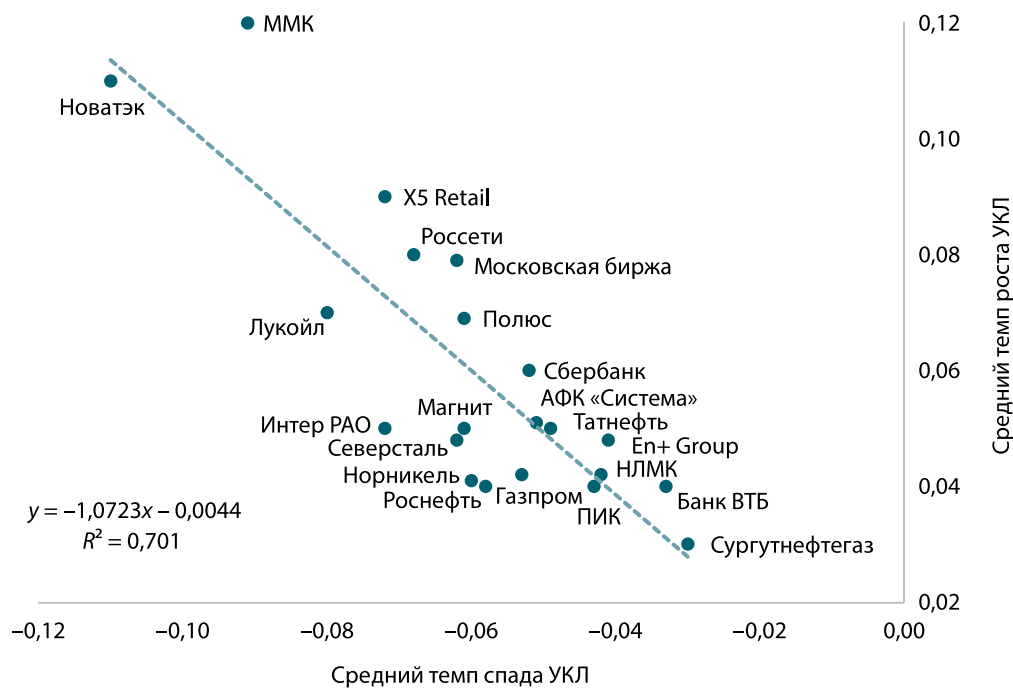


Рис. 4. Средняя динамика спада и роста уникальных ключевых лексем, процент в год

Fig. 4. Average decline and growth dynamics of unique key tokens, percentage per year

Согласно представленным данным, средний темп роста УКЛ всех компаний приблизительно равен среднему темпу спада. Это свидетельствует о трансформации лексики отчетов – обновлении корпуса уникальных ключевых лексем.

Заметная линейная регрессия ( $R^2 = 0,701$ ) средних темпов спада и роста УКЛ позволила выделить предприятия, лидирующие по динамике трансформации лексики. Максимальную динамику продемонстрировали отчеты ПАО «Новатэк» ( $-0,10; 0,11$  %). Также к группе лидеров по обновлению лексики можно отнести Магнитогорский металлургический комбинат, X5 Retail, «Россети», Московскую биржу. Относительная стабильность лексического состава ( $-0,03; 0,03$  %) характерна для компаний «Сургутнефтегаз», «ПИК», «Роснефть», «Банк ВТБ» и др.

В табл. 2 приведены УКЛ с максимальной динамикой спада и роста для всех компаний.

Примером наиболее заметной трансформации является динамика спада лексемы «банк» и динамичный рост лексемы «Сбер» в отчетах ПАО «Сбербанк», о чем упоминалось выше. Однако столь заметные изменения продемонстрировали не все предприятия. Наибольший интерес представляет изменение лексического состава отчетов ПАО «Лукойл», X5 Retail, «Магнит», «Северсталь», «Татнефть» и «Банк ВТБ». Следует отметить, что самой быстрорастущей УКЛ в корпусе отчетов ПАО «Интер РАО» является акроним GRI – Global Reporting Initiative, что указывает на тенденцию декларирования принципов Глобальной инициативы по отчетности и раскрытия нефинансовых показателей деятельности.

Очевидно, что анализ отдельных УКЛ с максимальной динамикой роста и спада не позволяет делать выводы об общих тенденциях в трансформации лексики отчетов. Поэтому, согласно методике данного исследо-

Таблица 2 – Максимальная динамика роста и спада уникальных ключевых лексем  
 Table 2 – Maximum growth and decline dynamics of unique key tokens

Компания	Убывающие УКЛ	Темп спада, %	Растущие УКЛ	Темп роста, %
Сбербанк	Банк	-0,24	«Сбер»	0,52
Россети	Рубль	-0,21	Отчет	0,39
Новатэк	Директор	-0,61	Рубль	0,35
Полюс	Финансовый	-0,14	Общество	0,31
Лукойл	Соблюдаться	-0,25	Развитие	0,28
X5 Retail	Совет	-0,39	Стратегический	0,25
Интер РАО	Рубль	-0,38	GRI	0,22
ММК	Рост	-0,27	Период	0,22
ПИК	Компания	-0,56	Застройщик	0,20
Магнит	Директор	-0,23	Развитие	0,19
Московская биржа	Соблюдаться	-0,24	Годовой	0,17
АФК «Система»	Годовой	-0,18	Управление	0,15
НЛМК	Деятельность	-0,11	Годовой	0,15
Газпром	Отчет	-0,18	Ввод	0,14
Норникель	Рынок	-0,17	Новый	0,14
En+ Group	Завод	-0,16	Область	0,13
Роснефть	Финансовый	-0,19	Группа	0,13
Северсталь	Актив	-0,16	Развитие	0,13
Татнефть	Финансовый	-0,13	Проект	0,12
Банк ВТБ	Группа	-0,13	Онлайн	0,10
Сургутнефтегаз	База	-0,10	Эксплуатация	0,10

вания, следующим этапом предварительной обработки данных стала классификация УКЛ с учетом направления (знака) их динамики.

1. Все анализируемые компании.
2. Компании в разбивке по укрупненным сферам деятельности (табл. 3).
3. Кластеры компаний (рис. 5).
4. Категория «развитие» – корпус УКЛ, отражающих процессы организационных изменений и стратегического развития предприятий.

Таблица 3 – Группы компаний согласно укрупненным сферам деятельности  
 Table 3 – Companies consolidated according to the enlarged areas of activity

Сфера деятельности	Компании
Банки и инвестиции	Сбербанк, Банк ВТБ, АФК «Система», Московская биржа
Металлургия и энергетика	Норникель, Россети, Интер РАО, En+ Group, НЛМК, Северсталь, ММК
Сырьевой сектор	Газпром, Лукойл, Роснефть, Сургутнефтегаз, Татнефть, Новатэк, Полюс
Торговля и строительство	X5 Retail, Магнит, ПИК



Рис. 5. Кластеры исследуемых компаний  
 Fig. 5. Clusters of the companies under study

Кластеризация исследуемых компаний осуществлялась на основе признака лексической близости отчетов с использованием виджетов Hierarchical Clustering, Distance Matrix и Distance Map модели Workflow



(см. рис. 2). Кластеры не тождественны группам компаний по укрупненным сферам деятельности.

Далее в работе релевантной называется УКЛ, соответствующая одному из классификационных признаков и описываемая кортежем  $\{x, m, F, D, R^2\}$ , где  $x$  – формулировка УКЛ;  $m$  – показатель консолидации, а именно процент компаний в рамках классификационного признака, в корпусе УКЛ которых используется данная лексема;  $F$  – средняя по числу компаний (в рамках классификационного признака) частотность общей УКЛ;  $D$  – средняя по числу компаний (в рамках классификационного признака) динамика общей УКЛ;  $R^2$  – показатель достоверности линейной аппроксимации ряда частотностей УКЛ по годам как оценка устойчивости тренда УКЛ.

Например, в группе из семи металлургических и энергетических предприятий растущая лексема «риск» является общей для шести из них, т. е. показатель консолидации  $m = 86\%$ , средняя частотность  $F = 0,53\%$ , средняя динамика  $D = +0,04\%$  с достоверностью линейной аппроксимации  $R^2 = 0,94$ . В той же группе УКЛ «финансовый» является общей только для трех предприятий ( $m = 43\%$ ) со средней частотностью  $F = 0,43\%$ , средней динамикой  $D = +0,11\%$  и  $R^2 = 0,52$ . Низкое значение показателя консолидации не позволяет рассматривать УКЛ «финансовый» в качестве релевантной растущей лексемы в данной классификационной группе. Кроме того, в зависимости от величины  $R^2$  показатель  $D$  с разной точностью характеризует динамику УКЛ: оценка средней динамики лексемы «риск» точнее характеризует ее динамику в данной группе, в то время как оценка средней динамики УКЛ «финансовый» не является достаточно достоверной.

Далее релевантность каждой УКЛ оценивалась по трем критериям – двум количественным и одному качественному:

- *первый количественный критерий* – показатель консолидации  $m$ , демонстрирующий степень един-

ства компаний данной классификационной группы в использовании конкретной УКЛ;

- *второй количественный критерий* – показатель достоверности линейной аппроксимации ряда частотностей УКЛ по годам;

- *качественный критерий* – степень соответствия задаче исследования основных контекстов данной УКЛ в отчетах компаний в классификационной группе.

Степень релевантности УКЛ задаче исследования – выявлению тенденций организационных изменений – во многом определяется на основании качественного критерия, предполагающего анализ контекстов использования УКЛ в отчетах. Методика и результаты контекстного анализа описаны ниже.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

**Результаты идентификации и классификации релевантных УКЛ.** В этом подразделе представлены результаты анализа релевантных УКЛ по указанным классификационным признакам. Основной целью данного этапа стала идентификация наиболее релевантных УКЛ и их классификация с учетом динамики за исследуемый период.

Для визуализации результатов анализа использовались карты релевантных УКЛ, на которых вертикальная ось соответствует средней частотности лексем, их размещение слева от этой оси указывает на снижение (отрицательную динамику), а справа – на рост (положительную динамику) частоты их применения. Размер шрифта приблизительно соответствует величине показателя консолидации, полужирным шрифтом выделены лексемы с максимальной степенью консолидации.

Карта, построенная для всех исследуемых компаний, позволяет сделать вывод об уровне консолидации лексем выше среднего ( $m = 62\text{--}71\%$ ), что с учетом существенных различий по видам экономической деятельности представляется достаточно показательным (рис. 6).

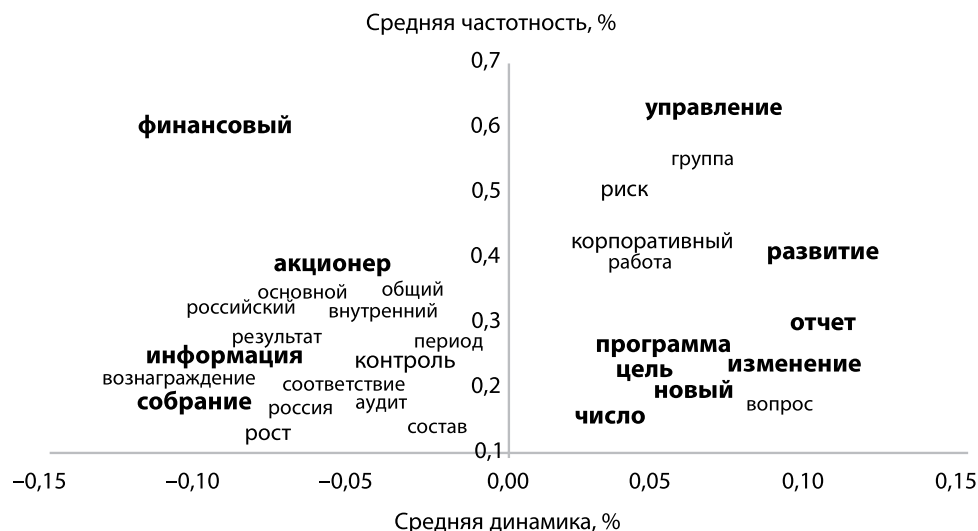


Рис. 6. Карта релевантных УКЛ для всех исследуемых компаний, 2018–2020

Fig. 6. Map of relevant UKTs for the companies under study in 2018–2020

Представленные данные показывают, что в отчетах крупнейших российских компаний за 2018–2020 гг. сформировался заметный тренд на спад терминов «финансовый», «акционер», «собрание», «контроль» и т. п., отражающих формальные паттерны и традиционные формулировки. Этот процесс сопровождался ростом частотности понятий «управление», «развитие», «программа», «цель» и «новый», указывающих на трансформацию лексики в направлении обсуждения организационных изменений и развития. Причем статус УКЛ с максимальной средней частотностью ( $F \geq 0,6\%$ ) за три года перешел от слова «финансовый» к слову «управление».

Примечательно, что для условной выборки, включающей одиннадцать предприятий, парная регрессия частотности лексем «контроль» и показателя «стоимость компании в 2021 г. (млрд долл.)» демонстрирует их сильную стохастическую связь ( $R^2 = 0,850$ ): высокие значения частотности данной УКЛ соответствуют низким значениям рыночной стоимости (рис. 7а).

Высокая частотность УКЛ «контроль» в отчетах компаний с низкой рыночной стоимостью может косвенно свидетельствовать о поиске административно-управленческих решений проблемы повышения величины этого показателя.

Для другой условной выборки тринадцати компаний характерна сильная стохастическая зависимость ( $R^2 = 0,867$ ) темпа роста УКЛ «изменение» и показателя «выручка компании в 2021 г. (млрд долл.)»: высокая динамика этой лексики соответствует низким значениям выручки (рис. 7б). Эта стохастическая связь может от-

ражать возрастающее внимание менеджмента компаний, указанных в правой части рисунка, к обсуждению направлений изменений с целью повышения объемов выручки.

Разумеется, эти и другие частные примеры регрессий не позволяют обоснованно идентифицировать причинно-следственные связи, однако могут быть использованы для уточнения общих выводов о факторах и направлениях трансформации лексики крупнейших российских компаний.

Анализ карты релевантных УКЛ (см. рис. 6) позволяет предположить, что одним из факторов высокой экономической результативности и эффективности компаний является устойчивый тренд на планирование и обсуждение направлений деятельности, связанных с постановкой актуальных целей и управлением программами развития.

На рис. 8 представлены результаты анализа релевантных УКЛ по укрупненным сферам деятельности. Выделение компаний по этому признаку позволило идентифицировать релевантные УКЛ с более высоким показателем консолидации (от 67 до 100%) ввиду меньшего количества предприятий в каждой выборке.

Для сферы «Банки и инвестиции» характерно наличие большого числа убывающих УКЛ, причем динамика спада понятий «система», «соблюдаться» и «вознаграждение» наблюдалась в отчетах всех компаний этой группы ( $m = 100\%$ ). Однако существенный консолидированный рост в данной выборке не продемонстрировала ни одна УКЛ.

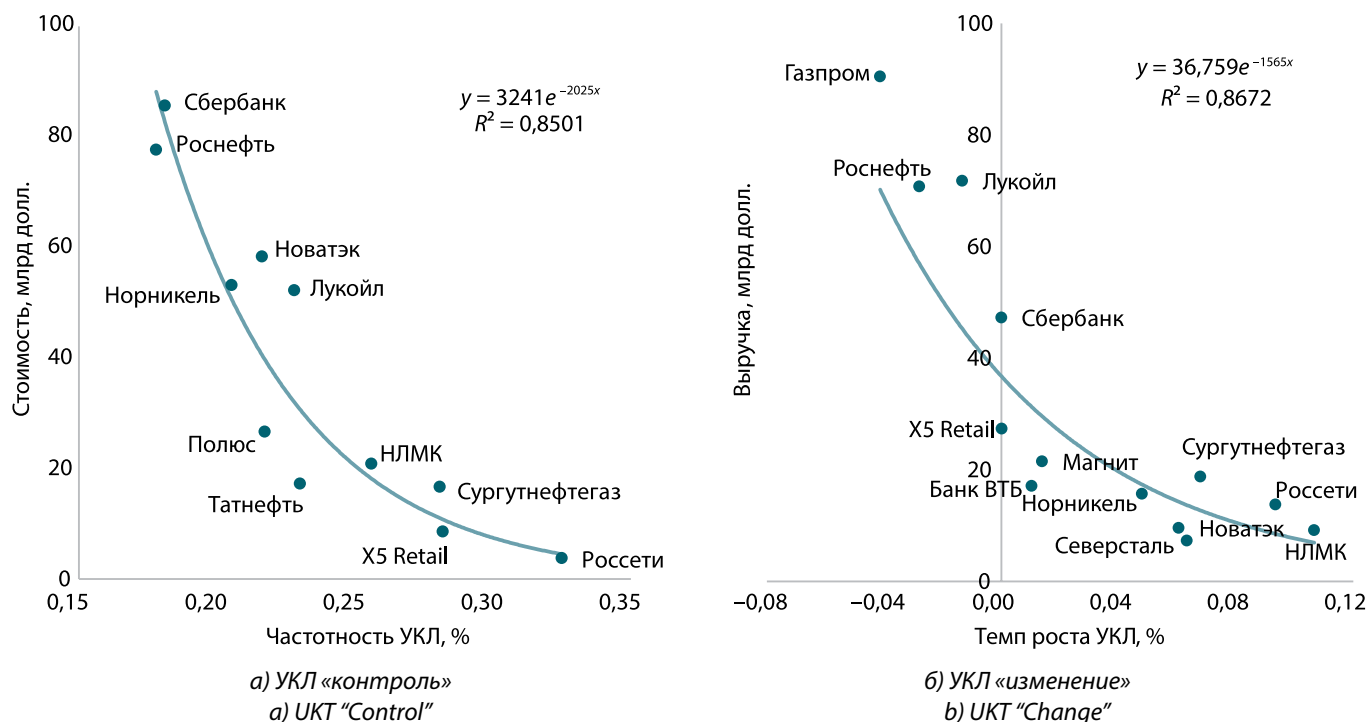


Рис. 7. Парные регрессии:

а) частотности УКЛ «контроль» и стоимости компаний в 2021 г.; б) темпа роста УКЛ «изменение» и выручки в 2021 г.

Fig. 7. Paired regressions:

a) occurrence of the UKT «Control» and the value of the companies in 2021;

b) the growth rate of the UKT «Change» and revenue in 2021

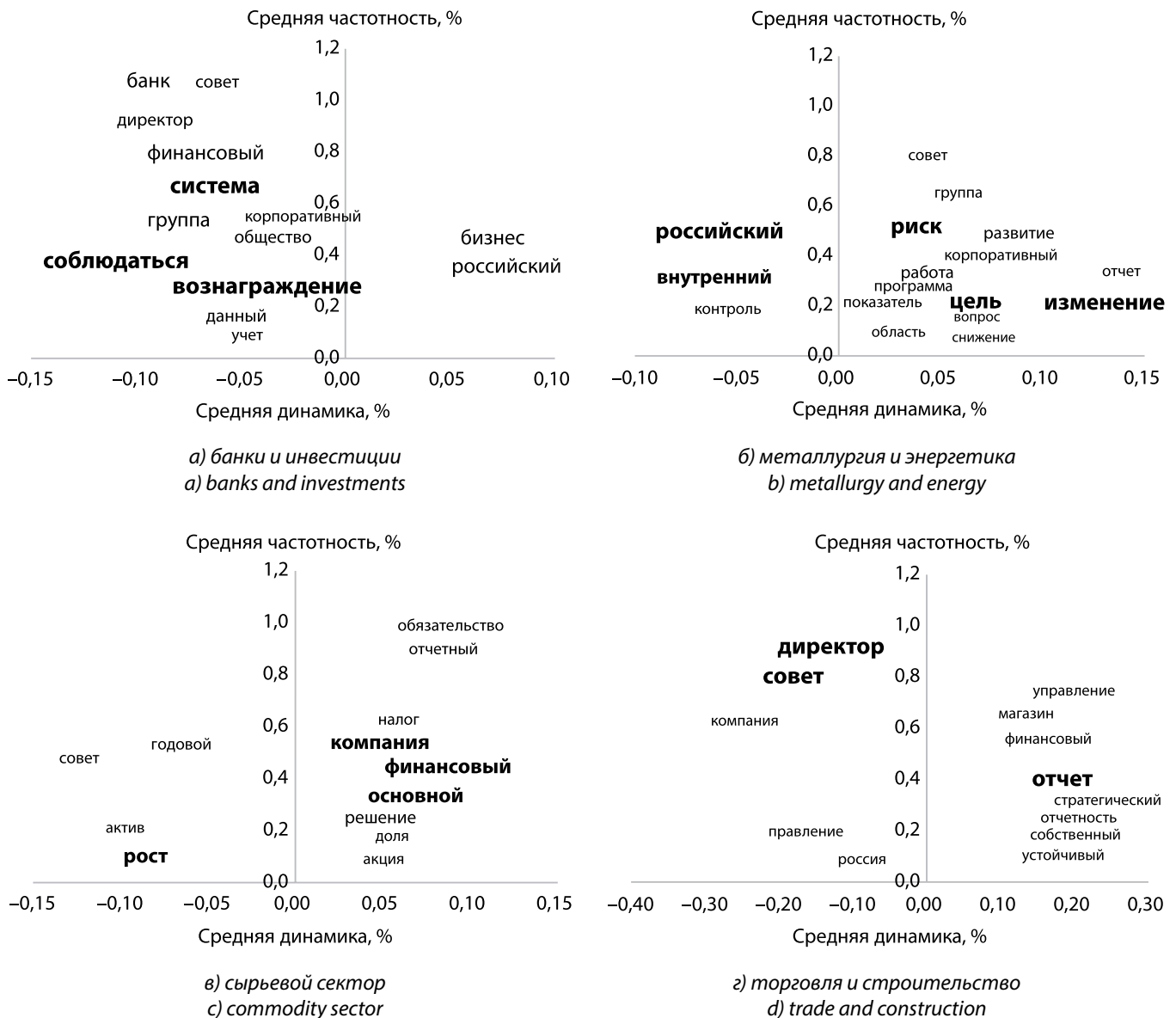


Рис. 8. Карты релевантных УКЛ по укрупненным сферам деятельности компаний  
Fig. 8. Maps of relevant UKTs by the enlarged areas of the companies' activities

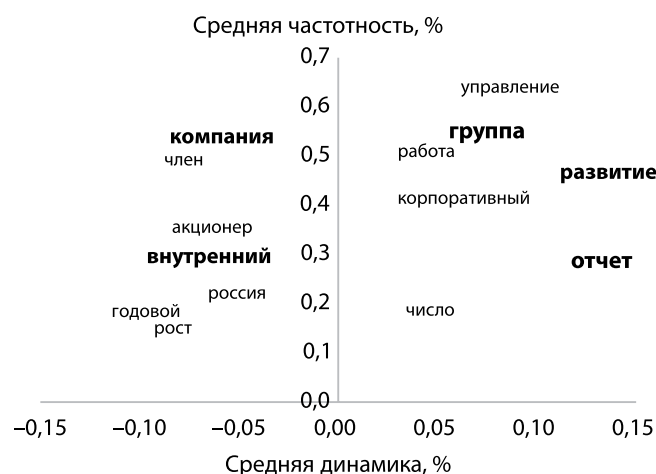
Тексты отчетов металлургических и энергетических предприятий заметно трансформировались в направлении спада частотности лексем «российский» и «внутренний» и положительной динамики УКЛ «цель», «риск» и «изменение». Последняя из них, имея низкую среднюю частотность (около 0,2 %), показала максимальный для выборки темп роста – около 0,13 % в год. Характер изменения корпуса релевантных УКЛ в этой сфере подтверждает общий тренд на отражение факторов развития, дополняя его проблемой риска.

В сырьевом секторе шесть из семи компаний ( $m = 86$  %) продемонстрировали медленный спад частотности УКЛ «рост». Имеются основания полагать, что этот факт свидетельствует не о снижении интереса предприятий данного сектора к развитию, а о формирующейся тенденции отказа от описания его параметров исключительно количественными значениями роста производственных, экономических и финансо-

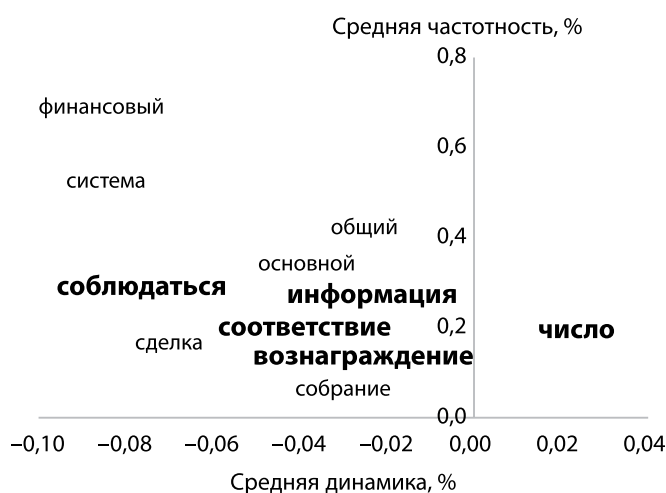
вых показателей и переходе к отражению характеристик качественных изменений.

Для предприятий сферы торговли и строительства характерен консолидированный спад высокочастотных УКЛ «директор» и «совет» (совместно составляющих термин «совет директоров»). Однако он сопровождался быстрым консолидированным ростом не менее формального понятия «отчет». Нельзя не обратить внимание на то, что общей «перегруппировке» формальных лексем сопутствовал в отдельных компаниях быстрый рост частотности терминов «управление», «стратегический» и «устойчивый».

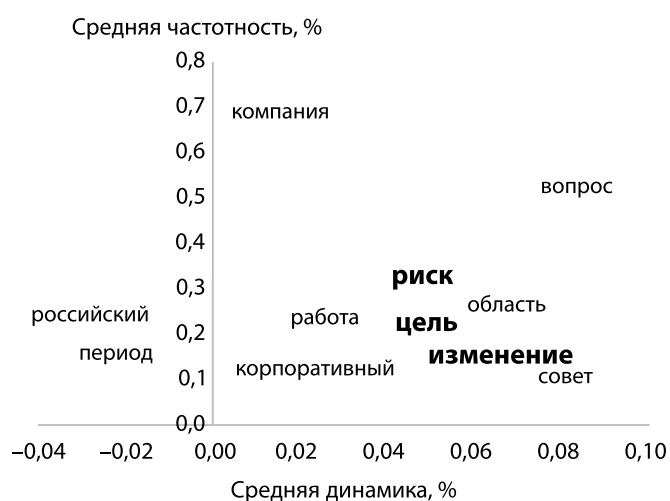
На рис. 9 показаны карты релевантных УКЛ по кластерам. Кластер «А» составили компании, продемонстрировавшие снижение интереса к описанию внутренних проблем и рост внимания к проблемам развития.



а) кластер «А»  
а) cluster "A"



б) кластер «Б»  
б) cluster "B"



в) кластер «В»  
в) cluster "C"

Рис. 9. Карты релевантных УКЛ по кластерам компаний  
Fig. 9. Maps of relevant UKTs by clusters of the companies

Отчеты компаний, вошедших в кластер «Б», характеризовались консолидированным спадом частотности традиционной формальной лексики: «соблюдаться», «информация», «соответствие», «вознаграждение», а также ростом частотности лексемы «число»<sup>1</sup>.

В компаниях, которые сформировали кластер «В», обнаружился отчетливый тренд на использование терминов развития «риск», «цель» и «изменение», совпадающих с набором релевантных УКЛ в сфере деятельности «Металлургия и энергетика». Это объясняется составом данного кластера – пять из шести вошедших в него предприятий представляют эту сферу. Данная группа релевантных УКЛ была признана наиболее перспективным объектом для контекстного анализа, результаты которого описаны в следующем подразделе.

Наибольший аналитический интерес в соответствии с целью исследования представляют результаты картирования УКЛ, отражающих процессы развития предприятий (рис. 10).

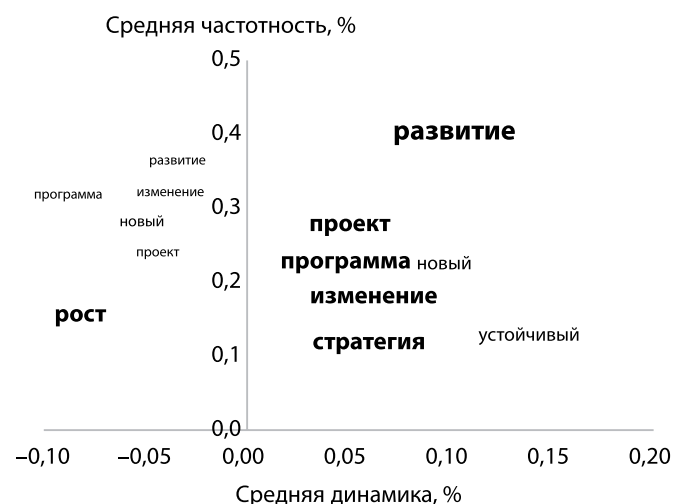


Рис. 10. Карта релевантных УКЛ в категории «Развитие»  
Fig. 10. Map of relevant UKTs in the category "Development"

За исследуемый период в лексике отчетов крупнейших российских компаний сформировался устойчивый тренд на замещение понятия «рост» лексемой «развитие» в сочетании с терминами «проект», «программа», «изменение» и «стратегия». Текущие позиции лексем позволяют предположить, что в деловой лексике большинства компаний снизилась актуальность терминологии количественного роста показателей производственной и финансово-экономической деятельности и усилилось внимание к категориям качественной трансформации на основе стратегических проектов и программ. Примечателен и тот факт, что в лексике отчетов малой части компаний наметился

<sup>1</sup> Отметим, что последующий контекстный анализ УКЛ «число» (см. рис. 9б) не подтвердил предположение о содержательной значимости этой лексемы, поскольку в 87 % случаев она использовалась в форме союза «в том числе», что заставляет рассматривать ее в качестве стоп-слова.

тренд на снижение активности использования лексем, характеризующих проблему развития, в то время как большая часть предприятий продемонстрировала заметный рост частотности таких УКЛ.

Имеются основания полагать, что в перспективе научно-практический интерес представляет выявление трендов более длительного временного периода и построение их прогнозов. В качестве частного примера долгосрочных тенденций на рис. 11 приведены графики частотности трех УКЛ в отчетах ПАО «Газпром» за десятилетний период (2011–2020 гг.), а также прогноз динамики УКЛ «развитие» до 2023 г., построенный по методу экспоненциального сглаживания с доверительным интервалом 95 %.

Налицо рост выборки УКЛ, отражающий стремление предприятия к развитию на основе реализации программ в условиях риска.

**Результаты контекстного анализа релевантных УКЛ.** В данной работе широкие лексические контексты релевантных УКЛ определялись с использованием виджета «Concordance» Workflow-модели Data Mining на основе группировки по частотности их лексическо-

го окружения в текстах отчетов. Под широкими лексическими контекстами понимаются контексты лексем в масштабе целого предложения (иногда двух), в отличие от узких контекстов, выявляемых в границах словосочетания.

Результаты контекстного анализа представлены диаграммами «Облако основных контекстов». Размер шрифта на них приближенно соответствует частотности контекстов (в процентах от общего числа выявленных контекстов), а наиболее частотные из них расположены ближе к центру диаграммы.

На рис. 12 показаны облака основных контекстов выбранных релевантных УКЛ для сферы «Металлургия и энергетика». В подрисуночных подписях указан процент выделенных контекстов, а также направление динамики УКЛ – спад или рост.

Облако контекстов растущей УКЛ «риск» интерпретируется следующим образом: в отчетах металлургических и энергетических предприятий за 2018–2020 гг. эта лексема использовалась в разных контекстах – 79 % из них составляют данное облако, остальные не вошли в него по причине заметно меньшей частот-

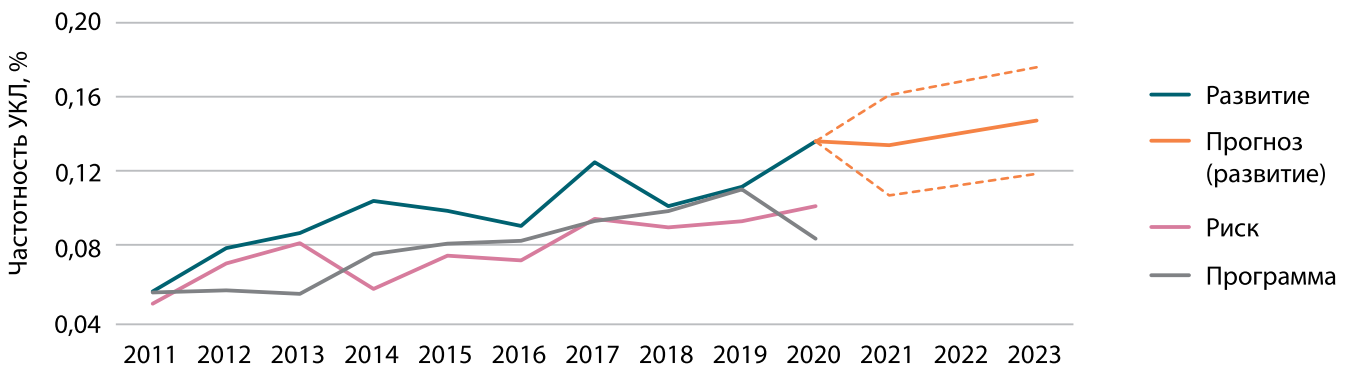
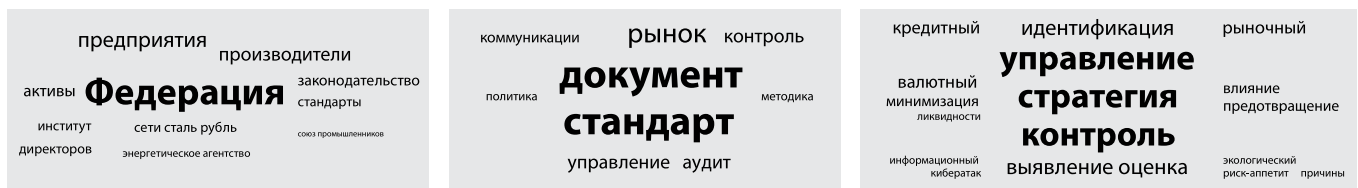


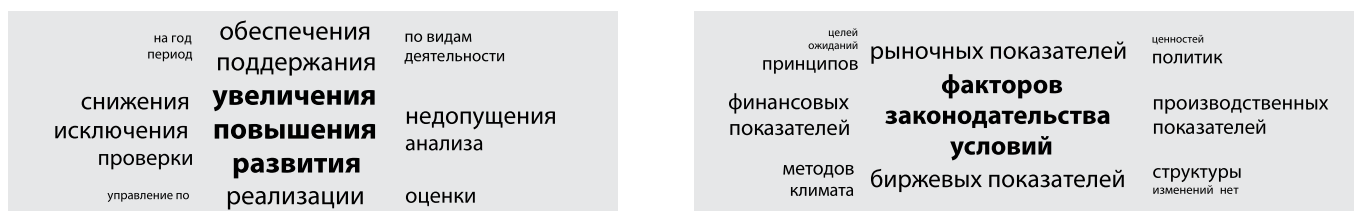
Рис. 11. УКЛ ПАО «Газпром»: динамика в период 2011–2020 гг. и прогноз  
 Fig. 11. UKTs of PAO Gazprom: dynamics in 2011–2020 and forecast



а) УКЛ «российский» (89 %, спад)  
 а) UKT “Russian” (89 %, decline)

б) УКЛ «внутренний» (91 %, спад)  
 б) UKT «Internal» (91 %, decline)

в) УКЛ «риск» (79 %, рост)  
 в) UKT «Risk» (79 %, growth)



г) УКЛ «цель» (96 %, рост)  
 д) UKT «Goal» (96 %, growth)

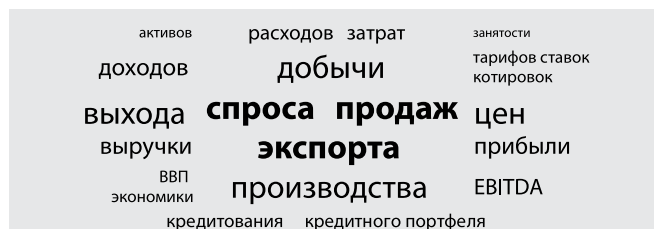
д) УКЛ «изменение» (95 %, рост)  
 е) UKT «Change» (95 %, growth)

Рис. 12. Основные контексты релевантных УКЛ в сфере «Металлургия и энергетика»  
 Fig. 12. Main contexts of relevant UKTs in the field of “Metallurgy and Energy”

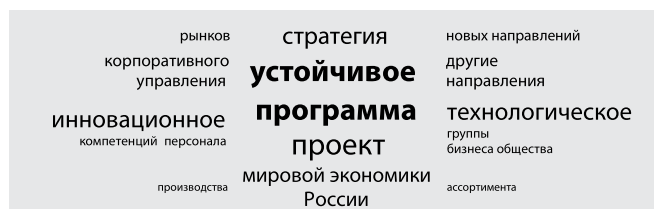
ности. Наиболее частотными оказались контексты «управление рисками», «стратегия управления рисками», «контроль рисков» (например, в предложении «Создание на предприятии системы внутреннего контроля и управления рисками и развитие подходов к риск-менеджменту»). Уступая перечисленным по частотности, в отчетах также использовались другие контексты: «идентификация рисков», «валютный риск», «оценка риска», «влияние риска на (тот или иной вид деятельности)» и т. п.

Анализ и сопоставление основных контекстов позволили выявить общую для металлургических и энергетических предприятий модель трансформации лексики: спад активности в использовании терминов, которые связаны с внутренними документами, разработанными на основе федеральных стандартов, и рост внимания к постановке целей развития в условиях изменений внутренних и внешних факторов, что обуславливает необходимость идентификации рисков и управления ими.

В категории «Развитие» (см. рис. 10) 62 % исследуемых компаний показали спад УКЛ «рост», 57 % – рост УКЛ «развитие». Для восьми из этих компаний («Газпром», «Лукойл», «Роснефть», «Татнефть», «Россети», «Северсталь», «ММК» и «Банк ВТБ») характерен процесс «однозначного замещения» – спад понятия «рост» сопровождается возрастанием частотности лексики «развитие». Дальнейший контекстный анализ релевантных УКЛ «рост» и «развитие» был проведен на основе отчетов данных компаний (рис. 13).



а) УКЛ «рост» (89 %, *спад*)  
а) УКТ «Growth» (89 %, *decline*)



б) УКЛ «развитие» (85,5 %, *рост*)  
б) УКТ «Development» (85,5 %, *growth*)

Рис. 13. Основные контексты УКЛ «рост» и «развитие» в категории «Развитие» для выборки из восьми компаний, 2018–2020

Fig. 13. Main contexts of UKTs «Growth» and «Development» in the category «Development» for the sample of eight companies in 2018–2020

В отчетах компаний из этой выборки в 2018–2020 гг. сокращалось использование понятия «рост» в следующих контекстах: «Рост спроса на продукцию на рынке», «Рост объемов продаж готовой продукции», «Рост цен на ресурсы», «Рост чистой прибыли» и т. п. При этом теми же темпами увеличивалась интенсивность использования понятия «развитие» в контекстах «Устойчивое развитие мировой экономики», «Стратегия развития бизнеса», «Программа развития компании (по направлениям)», «Цели инновационного развития предприятия» и т. п.

Таким образом, результаты контекстного анализа позволили точнее определить основное содержание изменений в лексике отчетов компаний.

### ЗАКЛЮЧЕНИЕ

Деловая лексика годовых отчетов крупнейших российских публичных компаний в период 2018–2020 гг. заметно трансформировалась: снизилась активность обсуждения формальных аспектов отчетности, работы совета директоров, соблюдения внутренних регламентов и увеличилось применение терминов, отражающих постановку новых целей, совершенствование методов управления, разработку программ развития, осуществление изменений, реализацию разнообразных проектов и программ и т. п. Наиболее явно эти тенденции проявились в отчетности металлургических и энергетических предприятий. Кроме того, анализ УКЛ в категории «Развитие» продемонстрировал рост внимания большинства компаний к проблемам организационного развития.

В первом случае основными контекстами стали проблемы постановки целей развития, увеличения основных показателей, разработки и реализации стратегии управления рисками, учета изменений внешних факторов и условий, в том числе законодательства, во втором – контексты роста спроса, продаж, производства, а также реализации концепции устойчивого развития, в том числе за счет выполнения соответствующих программ.

Заметное отражение тенденций развития характерно для компаний, вошедших в кластер «А» («Газпром», «Лукойл», «Роснефть», «X5 Retail», «Магнит», «Сургутнефтегаз», «Россети», «Интер РАО», «Новатэк»). В лексике компаний кластера «В» («Норникель», «En+ Group», «НЛМК», «Северсталь», «ММК», «Полус») наметился тренд на обсуждение проблем целеполагания, изменений и риска. Несмотря на то, что акроним GRI стал в 2018–2020 гг. самой быстрорастущей УКЛ только в отчетах ПАО «Интер РАО», признаки использования основных положений Глобальной инициативы по отчетности заметны в отчетах большинства исследуемых российских компаний – они проявляются в описании принципов экологической, социальной ответственности и корпоративного управления (ESG).

Результаты исследования подтверждают справедливость гипотезы о том, что трансформация лексики годовых отчетов крупнейших российских компаний, выраженная показателями динамики ключевых лексем и их основных контекстов, свидетельствует о наличии факторов и тенденций организационных изменений и стратегического развития. Данные показатели отражают наиболее общие признаки и тенденции развития компаний вне зависимости от их отраслевой принадлежности и других отличительных качеств. Динамика и основные контексты ключевых лексем обладают индикаторными свойствами, позволяющими рассматривать их в качестве перспективных индикаторов нечисловой природы в рамках инструментария мультимодальной бизнес-аналитики для осуществления многоаспектного сравнительного бизнес-анализа.

Предложенный метод анализа нефинансовой информации позволяет выявлять ключевые тенденции изменения содержания годовых отчетов российских компаний. Его оригинальность состоит в оценке динамических аспектов трансформации, а также контекстов использования релевантных ключевых лексем. Полученные результаты и выводы могут применяться для расширения границ сравнительного бизнес-ана-

лиза российских компаний и повышения его аналитической и прогнозной результативности.

К ограничениям метода относится его основное допущение, согласно которому содержание годовых отчетов достоверно отражает состояние дел в компаниях, а сочетание роста или спада частотности контекстно взаимосвязанных лексем свидетельствует о возрастании или снижении внимания менеджмента компаний к проблемам их развития. Устранению этого ограничения может способствовать учет статистически достоверных отклонений содержания отчетности от реальной ситуации.

К направлениям совершенствования предложенного метода можно отнести использование в качестве источников нечисловой информации других форм отчетности и информационных материалов: интегрированных отчетов, текстов программ развития, стратегий развития компаний и т. п.; охват более обширной выборки компаний в сочетании с анализом трендов трансформации их деловой лексики и прогнозированием на более длительные периоды; выявление статистически достоверных регрессий динамики релевантных УКЛ с финансово-экономическими показателями деятельности компаний. ■

#### Источники

- Афанасьев Д.О., Федорова Е.А., Демин И.С., Рогов О.Ю. (2020). О влиянии зарубежных СМИ на российский фондовый рынок: текстовый анализ // Экономика и математические методы. Т. 56, № 2. С. 77–89. DOI: 10.31857/S042473880008527-8.
- Батаева Б.С., Кокурина А.Д., Карпов Н.А. (2021). Влияние раскрытия ESG-показателей на финансовые результаты российских публичных компаний // Управленец. Т. 12, № 6. С. 20–32. DOI: 10.29141/2218-5003-2021-12-6-2.
- Калабихина И.Е., Крикунов А.С. (2018). Новая методика оценки качества нефинансовой отчетности (на примере энергетических компаний) // Вестник СПбГУ. Менеджмент. Т. 17, вып. 3. С. 297–328. DOI: 10.21638/11701/spbu08.2018.303.
- Корнилова А.В., Никоноров С.М. (2017). Методология оценки качества раскрытия информации о корпоративной социальной и экологической ответственности // Вестник Московского университета. Серия 6, Экономика. № 2. С. 3–21. DOI: 10.38050/01300105201721.
- Кузубов С.А., Евдокимова М.С. (2017). Повышает ли стоимость компании публикация нефинансовых отчетов по стандартам GRI (на примере стран БРИКС)? // Учет. Анализ. Аудит. № 2. С. 28–36. DOI: 10.26794/2408-9303-2017-2-28-36.
- Нехода Е.В., Арабов Н.У., Богданов А.Л., Герман М.В., Куклина Т.В. (2022). Достойный труд в нефинансовой отчетности российских компаний: оценка качества раскрытия информации // Управленец. Т. 13, № 2. С. 34–56. DOI: 10.29141/2218-5003-2022-13-2-3.
- Пятов М.Л., Соловей Т.Н., Сорокина А.С., Гусниева А.А. (2018). Нефинансовая отчетность в экономике: опыт XIX – начала XXI в. // Вестник Санкт-Петербургского университета. Экономика. Т. 34, вып. 3. С. 465–492. DOI: 10.21638/spbu05.2018.306.
- Теплова Т.В., Соколова Т.В., Томтосов А.Ф., Бучко Д.В., Никулин Д.Д. (2022). Сентимент частных инвесторов в объяснении различий в биржевых характеристиках акций российского рынка // Журнал Новой экономической ассоциации. № 1 (53). С. 53–84. DOI: 10.31737/2221-2264-2022-53-1-3.
- Федорова Е.А., Афанасьев Д.О., Нерсесян Р.Г., Ледяева С.В. (2020). Влияние нефинансовой информации на основные показатели российских компаний // Журнал Новой экономической ассоциации. № 2 (46). С. 73–96. DOI: 10.31737/2221-2264-2020-46-2-4.
- Федорова Е.А., Пыльцин И.В., Ковальчук Ю.А., Дроговоз П.А. (2022). Новости и социальные сети российских компаний: степень влияния на рынок ценных бумаг // Журнал Новой экономической ассоциации. № 1 (53). С. 32–52. DOI: 10.31737/2221-2264-2022-53-1-2.
- Федорова Е.А., Хрустова Л.Е., Демин И.С. (2020). Влияние качества раскрытия нефинансовой информации российскими компаниями на их инвестиционную привлекательность // Российский журнал менеджмента. Т. 18, № 1. С. 51–72. DOI: 0.21638/spbu18.2020.103.

- Ширяев И.М., Курышева А.А., Вольчик В.В. (2021). Нарративный институциональный анализ и российская инновационная система // *Journal of Institutional Studies*. Т. 13, № 3. С. 81–101. DOI: 10.17835/2076-6297.2021.13.3.081-101.
- Adams C.A., Frost G. (2008). Integrating sustainability reporting into management practices. *Accounting Forum*, no. 32, pp. 288–302. DOI: 10.1016/j.accfor.2008.05.002
- Ahmad K., Han J.G., Hutson E., Kearney C., Liu S. (2015). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, no. 37(C), pp. 152–172. DOI: 10.1016/J.JCORPFIN.2015.12.014
- Asif M., Searcy C., Santos P., Kenshal D. (2013). A review of Dutch corporate sustainable development reports. *Corporate Social Responsibility and Environmental Management*, vol. 20, issue 6, pp. 321–339. DOI: 10.1002/csr.1284
- Baker M., Wurgler J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, vol. LXI, no. 4, pp. 1645–1680. DOI: dx.doi.org/10.2139/ssrn.464843
- Berry M.J., Linoff G.S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management*. John Wiley & Sons.
- Bonsón E., Perea D., Azevedo G. (2021). Tone and content analysis in the president's letters to shareholders: Spanish evidence. *Upravlenets / The Manager*, vol. 12, no. 1, pp. 78–90. DOI: 10.29141/2218-5003-2021-12-1-6.
- Brennan N., Merkl-Davies D. (2013). *Accounting narratives and impression management*. The Routledge Companion to Communication in Accounting. London, Routledge, pp. 109–132. DOI: 10.4324/9780203593493.CH8
- Davis G., Searcy C. (2010). A review of Canadian corporate sustainable development reports. *Journal of Global Responsibility*, no. 1, pp. 316–329. DOI: 10.1108/20412561011079425
- Fedorova E., Druchok S., Drogovoz P. (2022). Impact of news sentiment and topics on IPO underpricing: US evidence. *International Journal of Accounting & Information Management*, vol. 30, no. 1, pp. 73–94. DOI: 10.1108/IJAIM-06-2021-0117
- Lahat D., Adali T., Jutten C. (2015). Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477. DOI: 10.1109/JPROC.2015.2460697
- Li M., Wang F., Jia X., Li W., Li T. (2021). Multi-source data fusion for economic data analysis. *Neural Computing and Applications*, no. 33, pp. 4729–4739. DOI: 10.1007/s00521-020-05531-0
- Piatetsky-Shapiro G. (2007). Data mining and knowledge discovery 1996 to 2005: Overcoming the hype and moving from “university” to “business” and “analytics”. *Data Mining and Knowledge Discovery*, no. 15, pp. 99–105. DOI: 10.1007/s10618-006-0058-2
- Renault T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking and Finance*, no. 84, pp. 25–40. DOI: 10.1016/j.jbankfin.2017.07.002
- Saber M., Weber A. (2019). Sustainable grocery retailing: Myth or reality? – A content analysis. *Business and Society Review*, vol. 124, issue 4, pp. 479–496. DOI: 10.1111/basr.12187
- Siltaoja M. (2006). Value priorities as combining core factors between CSR and reputation – A qualitative study. *Journal of Business Ethics*, no. 68, pp. 91–111. DOI: 10.1007/S10551-006-9042-4
- Skouloudis A., Evangelinos K.I., Kourmoussis F. (2010). Assessing non-financial reports according to the Global Reporting Initiative guidelines: Evidence from Greece. *Journal of Cleaner Production*, no. 18, pp. 426–438. DOI: 10.1016/J.JCLEPRO.2009.11.015
- Tschopp D., Huefner R.J. (2015). Comparing the evolution of CSR reporting to that of financial reporting. *Journal of Business Ethics*, vol. 127, issue 3, pp. 565–577. DOI: 10.1007/s10551-014-2054-6
- Yager R. (2004). A framework for multi-source data fusion. *Information Sciences*, vol. 163, issues 1–3, pp. 75–200. DOI: 10.1016/j.ins.2003.03.018

### References

- Afanasyev D.O., Fedorova E.A., Demin I.S., Rogov O.Yu. (2020). On the influence of foreign media on the Russian stock market: Text analysis. *Ekonomika i matematicheskie metody / Economics and Mathematical Methods*, vol. 56, issue 2, pp. 77–89. DOI: 10.31857/S042473880008527-8. (in Russ.)
- Bataeva B.S., Kokurina A.D., Karpov N.A. (2021). The impact of ESG reporting on the financial performance of Russian public companies. *Upravlenets / The Manager*, vol. 12, no. 6, pp. 20–32. DOI: 10.29141/2218-5003-2021-12-6-2. (in Russ.)
- Kalabikhina I.E., Krikunov A.S. (2018). A new method of assessing the quality of non-financial reporting (on the example of energy companies). *Vestnik SPbGU. Menedzhment / Vestnik of St Petersburg University. Management*, vol. 17, issue 3, pp. 297–328. DOI: 10.21638/11701/spbu08.2018.303. (in Russ.)
- Kornilova A.V., Nikonorov S.M. (2017). Methodology for assessing the disclosure quality of corporate social and environmental responsibility information. *Vestnik Moskovskogo universiteta. Seriya 6: Ekonomika / Moscow University Economics Bulletin*, no. 2, pp. 3–21. DOI: 10.38050/01300105201721. (in Russ.)
- Kuzubov S.A., Evdokimova M.S. (2017). 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*, no. 2, pp. 28–36. <https://doi.org/10.26794/2408-9303-2017--2-28-36>. (in Russ.)
- Nekhoda E.V., Arabov N.U., Bogdanov A.L., German M.V., Kuklina T.V. (2022). Decent work in non-financial reporting of Russian companies: Assessing the disclosure quality. *Upravlenets / The Manager*, vol. 13, no. 2, pp. 34–56. DOI: 10.29141/2218-5003-2022-13-2-3. (in Russ.)
- Pyatov M.L., Solovey T.N., Sorokina A.S., Gusnieva A.A. (2018). Non-financial reporting in the economy: History of 19th – early 21st century. *Vestnik Sankt-Peterburgskogo universiteta. Ekonomika / St Petersburg University Journal of Economic Studies (SUJES)*, vol. 34, issue 3, pp. 465–492. DOI: 10.21638/spbu05.2018.306. (in Russ.)



- Teplova T.V., Sokolova T.V., Tomtosov A.F., Buchko D.V., Nikulin D.D. (2022). 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*, no. 1(53), pp. 53–84. DOI: 10.31737/2221-2264-2022-53-1-3. (in Russ.)
- Fedorova E.A., Afanasev D.O., Nersesyan R.G., Ledyayeva S.V. (2020). Impact of non-financial information on key financial indicators of Russian companies. *Zhurnal Novoy ekonomicheskoy assotsiatsii / Journal of the New Economic Association*, no. 2(46), pp. 73–96. DOI: 10.31737/2221-2264-2020-46-2-4. (in Russ.)
- Fedorova E.A., Pyltsin I.V., Kovalchuk Yu.A., Drogovoz P.A. (2022). News and social networks of Russian companies: Degree of influence on the securities market. *Zhurnal Novoy ekonomicheskoy assotsiatsii / Journal of the New Economic Association*, no. 1(53), pp. 32–52. DOI: 10.31737/2221-2264-2022-53-1-2. (in Russ.)
- Fedorova E.A., Khrustova L.E., Demin I.S. (2020). Completeness of non-financial disclosure by Russian companies: The influence on investment attractiveness. *Rossiyskiy zhurnal menedzhmenta / Russian Management Journal*, vol. 18, no. 1, pp. 51–72. <https://doi.org/10.21638/spbu18.2020.103>. (in Russ.)
- Shiriaeve I.M., Kuryshcheva A.A., Volchik V.V. (2021). Narrative institutional analysis and the national innovation system in Russia. *Journal of Institutional Studies*, vol. 13, no. 3, pp. 81–101. DOI: 10.17835/2076-6297.2021.13.3.081-101. (in Russ.)
- Adams C.A., Frost G. (2008). Integrating sustainability reporting into management practices. *Accounting Forum*, no. 32, pp. 288–302. DOI: 10.1016/j.accfor.2008.05.002
- Ahmad K., Han J.G., Hutson E., Kearney C., Liu S. (2015). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, no. 37(C), pp. 152–172. DOI: 10.1016/J.JCORPFIN.2015.12.014
- Asif M., Searcy C., Santos P., Kensah D. (2013). A review of Dutch corporate sustainable development reports. *Corporate Social Responsibility and Environmental Management*, vol. 20, issue 6, pp. 321–339. DOI: 10.1002/csr.1284
- Baker M., Wurgler J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, vol. LXI, no. 4, pp. 1645–1680. DOI: dx.doi.org/10.2139/ssrn.464843
- Berry M.J., Linoff G.S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management*. John Wiley & Sons.
- Bonsón E., Perea D., Azevedo G. (2021). Tone and content analysis in the president's letters to shareholders: Spanish evidence. *Upravlenets / The Manager*, vol. 12, no. 1, pp. 78–90. DOI: 10.29141/2218-5003-2021-12-1-6.
- Brennan N., Merkl-Davies D. (2013). *Accounting narratives and impression management*. The Routledge Companion to Communication in Accounting. London, Routledge, pp. 109–132. DOI: 10.4324/9780203593493.CH8
- Davis G., Searcy C. (2010). A review of Canadian corporate sustainable development reports. *Journal of Global Responsibility*, no. 1, pp. 316–329. DOI: 10.1108/20412561011079425
- Fedorova E., Druchock S., Drogovoz P. (2022). Impact of news sentiment and topics on IPO underpricing: US evidence. *International Journal of Accounting & Information Management*, vol. 30, no. 1, pp. 73–94. DOI: 10.1108/IJAIM-06-2021-0117
- Lahat D., Adali T., Jutten C. (2015). Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477. DOI: 10.1109/JPROC.2015.2460697
- Li M., Wang F., Jia X., Li W., Li T. (2021). Multi-source data fusion for economic data analysis. *Neural Computing and Applications*, no. 33, pp. 4729–4739. DOI: 10.1007/s00521-020-05531-0
- Piatetsky-Shapiro G. (2007). Data mining and knowledge discovery 1996 to 2005: Overcoming the hype and moving from “university” to “business” and “analytics”. *Data Mining and Knowledge Discovery*, no. 15, pp. 99–105. DOI: 10.1007/s10618-006-0058-2
- Renault T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking and Finance*, no. 84, pp. 25–40. DOI: 10.1016/j.jbankfin.2017.07.002
- Saber M., Weber A. (2019). Sustainable grocery retailing: Myth or reality? – A content analysis. *Business and Society Review*, vol. 124, issue 4, pp. 479–496. DOI: 10.1111/basr.12187
- Siltaoja M. (2006). Value priorities as combining core factors between CSR and reputation – A qualitative study. *Journal of Business Ethics*, no. 68, pp. 91–111. DOI: 10.1007/S10551-006-9042-4
- Skouloudis A., Evangelinos K.I., Kourmoussis F. (2010). Assessing non-financial reports according to the Global Reporting Initiative guidelines: Evidence from Greece. *Journal of Cleaner Production*, no. 18, pp. 426–438. DOI: 10.1016/J.JCLEPRO.2009.11.015
- Tschopp D., Huefner R.J. (2015). Comparing the evolution of CSR reporting to that of financial reporting. *Journal of Business Ethics*, vol. 127, issue 3, pp. 565–577. DOI: 10.1007/s10551-014-2054-6
- Yager R. (2004). A framework for multi-source data fusion. *Information Sciences*, vol. 163, issues 1–3, pp. 75–200. DOI: 10.1016/j.ins.2003.03.018

## Информация об авторе

## Information about the author

## Михненко Павел Александрович

## Pavel A. Mikhnenko

Доктор экономических наук, профессор кафедры предпринимательства и внешнеэкономической деятельности. МГТУ им. Н.Э. Баумана, г. Москва, РФ. E-mail: mpa69@yandex.ru

Dr. Sc. (Econ.), Professor of Entrepreneurship and Foreign Economic Activity Dept. Bauman University, Moscow, Russia. E-mail: mpa69@yandex.ru