

Научная статья

УДК 332.14

DOI: 10.17323/1999-5431-2022-0-2-106-133

ОЦЕНКА СООТВЕТСТВИЯ ПРИОРИТЕТОВ СТРАТЕГИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ, ИХ ОТРАСЛЕВОЙ СПЕЦИАЛИЗАЦИИ НА ОСНОВЕ *TEXT MINING*

¹ **Козоногова Елена Викторовна,**

² **Дубровская Юлия Владимировна,**

³ **Русинова Мария Романовна,**

⁴ **Иванов Павел Витальевич**

^{1, 2, 3} Гуманитарный факультет, Пермский национальный исследовательский политехнический университет; 614990, г. Пермь, Комсомольский пр., 29, корпус Б, к. 307.

¹ Кандидат экономических наук, доцент; elena.kozonogova@gmail.com; ORCID: 0000-0001-9573-7336

² Кандидат экономических наук, доцент; uliadubrov@mail.ru; ORCID: 0000-0002-3205-9264

³ Кандидат экономических наук, младший научный сотрудник; rusinova.mr@gmail.com; ORCID: 0000-0002-3576-3409

⁴ Начальник отдела промышленности Министерства промышленности и торговли Пермского края; 614990, г. Пермь, ул. Петропавловская, д. 56; pvivanov@permkrai.ru; ORCID: 0000-0002-5402-4935

Аннотация. Задача определения уникальности целеполагания регионов в части проверки соответствия текстов стратегий развития регионов их отраслевой специализации решалась в ходе представленного в статье исследования. На основе технологии «умного» бенчмаркинга территорий, а также инструментария *Text Mining*, проанализированы долгосрочные стратегии развития 11 регионов с общим объемом текстового корпуса 415 780 слов. В качестве ключевых слов выбраны основные разделы общероссийского классификатора видов экономической деятельности, характеризующие отраслевые приоритеты развития регионов. Извлечение ключевых понятий из текстов стратегий и проведение их количественного анализа осуществлено с помощью высокоуровневого языка программирования *Python*. Полученные количественные результаты сравнения именованных сущностей стратегий развития субъектов Российской Федерации доказали, что недостаточность уникального целеполагания в части идентификации перспективных специализаций в стратегиях развития субъектов Российской Федерации искажает систему приоритетных направлений развития и поэтому объективно является одной из причин недостижения территориями запланированных показателей. В работе используются методы интеллектуального анализа текста, математической

статистики, группировки и обобщения, а также приемы визуализации анализируемых данных. Авторская методика проведения интеллектуального анализа текстов является универсальной для любой области науки. Разработанные алгоритмы извлечения именованных сущностей из текста и представленные алгоритмы *Text Mining* открывают широкие горизонты для дальнейших исследований в области анализа стратегий как публичных документов, обращенных к заинтересованным субъектам.

Ключевые слова: отраслевая специализация, приоритеты регионального развития, Text Mining, региональные стратегии развития, «умный» бенчмаркинг, интеллектуальный анализ текста.

Для цитирования: Козоногова Е.В., Дубровская Ю.В., Русинова М.Р., Иванов П.В. Оценка соответствия приоритетов стратегического развития регионов, их отраслевой специализации на основе *Text Mining* // Вопросы государственного и муниципального управления. 2022. № 2. С. 106–133. DOI: 10.17323/1999-5431-2022-0-2-106-133.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 21-78-10134, <https://rscf.ru/project/21-78-10134/>.

Благодарность: Статья подготовлена в рамках проекта по поддержке публикаций авторов российских образовательных и научных организаций в научных изданиях НИУ ВШЭ (программа «Университетское партнерство»).

Original article

ASSESSMENT OF COMPLIANCE OF STRATEGIC DEVELOPMENT PRIORITIES OF REGIONS WITH THEIR INDUSTRY SPECIALIZATION BASED ON TEXT MINING

¹ **Elena V. Kozonogova,**

² **Yulia V. Dubrovskaya,**

³ **Maria R. Rusinova,**

⁴ **Pavel V. Ivanov**

^{1, 2, 3} Faculty of Humanities, Perm National Research Polytechnic University;
Of. 307, 29/b Komsomolsky Av., 614990 Perm, Russia

¹ Ph.D. (in economics), associate professor; elena.kozonogova@gmail.com;
ORCID: 0000-0001-9573-7336

² Ph.D. (in economics), associate professor; uliadubrov@mail.ru; ORCID: 0000-0002-3205-9264

³ Ph.D. (in economics), junior research fellow; rusinova.mr@gmail.com; ORCID: 0000-0002-3576-3409

⁴ Head of the Industry Department, Ministry of Industry and Trade of the Perm Territory;
56 Petropavlovskaya St., 614990 Perm, Russia; pvivanov@permkrai.ru; ORCID: 0000-0002-5402-4935

Abstract. The task of determining the correctness of self-positioning of regions in terms of verifying the compliance of texts of regional development strategies with their industry specialization was solved in the course of the research presented in the article. Using the «smart» benchmarking methodology, as well as the Text Mining tools, long-term development strategies of 11 regions with a total text corpus of 415,780 words were analyzed. The main sections of the all-Russian classifier of economic activities that characterize the sectoral priorities of regional development were selected as keywords. The extraction of key concepts from strategy texts, as well as their quantitative analysis, was carried out using the high-level Python programming language. The obtained quantitative results of comparing the named entities of the development strategies of the subjects of the Russian Federation proved that the insufficiency of unique goal-setting in terms of identifying promising specializations in regional development strategies distorts the system of priority development directions. This is objectively one of the reasons why the territories do not achieve the planned indicators. The paper uses methods of Text Mining, mathematical statistics, grouping and generalization, as well as techniques for visualizing the analyzed data. The author's method of conducting intellectual analysis of texts is universal for any field of science. The developed algorithms for extracting named entities from the text and algorithms for quantitative analysis of the text open up wide horizons for further research in the field of strategy analysis, as public documents addressed to interested subjects.

Keywords: industry specialization, regional development priorities, Text Mining, regional development strategies, “smart” benchmarking, intelligent text analysis.

For citation: Kozonogova, E.V., Dubrovskaya, Yu.V., Rusinova, M.R. and Ivanov, P.V. (2022) ‘Assessment of compliance of strategic development priorities of regions with their industry specialization based on Text Mining’, *Public Administration Issues*, 2, pp. 106–133. (in Russian). DOI: 10.17323/1999-5431-2022-0-2-106-133.

Acknowledgement: The study was supported by the Russian Science Foundation Grant No. 21-78-10134, <https://rscf.ru/project/21-78-10134/>.

The article was prepared as part of a project to support publications by authors of Russian educational and scientific organizations in scientific publications of the National Research University Higher School of Economics, Russia (University Partnership Program).

JEL: R58, C83

Введение

В основе территориального стратегирования на современном этапе лежит научно обоснованная идентификация перспективных производственных специализаций, развиваемых ведущими отраслями регионов. Законодательная формализация такого подхода к управлению территориальным развитием была осуществлена Министерством экономического развития РФ согласно Методическим рекомендациям по разработке стратегии социаль-

но-экономического развития субъектов Российской Федерации¹, а также Правительством РФ в процессе утверждения Стратегии пространственного развития Российской Федерации до 2025 г.². Следуя последнему документу, стратегирование регионального развития должно осуществляться в рамках закрепленных законодателем экономических специализаций, перспективность которых обусловлена благоприятным сочетанием конкурентных преимуществ. Так как именно отраслевая привязка традиционно лежит в основе планирования территориального развития, можно объективно заключить, что качество территориального развития напрямую зависит не только от обоснованности идентификации, но и от степени последующей приоритезации перспективных специализаций в стратегиях социально-экономического развития.

В целом вопросам оценки качества регионального стратегирования уделяется достаточно высокое внимание как учеными, так и региональными управлениями. Вместе с тем существующие подходы к оценке эффективности территориальных стратегий по большей части носят качественный характер (Ильина и др., 2016; Морозова и др., 2014; Шпакова, 2019; Комаров и др., 2021). Количественные методы, если и применяются, то в полуручном режиме. Например, ученые используют методику балльной системы оценки или вводят авторские шкалы оценивания (Егоров, 2010; Ильина и др., 2014; Кислицына и др., 2017; Рисин, 2021). При этом часто отечественные исследователи, анализирующие стратегии развития субъектов в России, говорят о спонтанности процесса стратегирования и низком качестве регионального стратегического управления (Батов, 2016; Липина и др., 2017; Климанов и др., 2017; Марголин, 2018).

Как отмечает Е. Куценко, в большинстве случаев выбранные приоритеты развития «слабо обоснованы аналитически и не подкреплены конкретными проектами» (Kutsenko et al., 2018, с. 38). В результате сложившаяся практика регионального стратегирования подчас принимает декларативный характер, что не позволяет заинтересованным субъектам проводить сопоставительные анализы, формулировать выводы и принимать решения (Суховой и др., 2016; Репичев и др., 2018). Очевидно, что неверно заданный вектор развития территории сдерживает реализацию имеющегося потенциала. Таким образом, актуализируется необходимость совершенствования инструментария, позволяющего проводить количественные оценки текстов стратегий развития регионов на предмет уникальности целеполагания, что и явилось целью настоящего исследования.

Для достижения поставленной цели мы обратились к методике интеллектуального анализа текста *Text Mining*. Задача такого анализа – найти общую тенденцию в текстовых данных и выявить новую или не очевидную чи-

¹ Пп. «е» п. 4.1 ст. 4 Приказа Министерства экономического развития Российской Федерации N 132 от 23 марта 2017 г. «Об утверждении Методических рекомендаций по разработке и корректировке стратегии социально-экономического развития субъекта Российской Федерации и плана мероприятий по ее реализации» (с изменениями на 7 сентября 2018 г.).

² Утвержденной Распоряжением от 13 февраля 2019 года N 207-р.

тателю информацию (Türegün, 2019). Данный функционал представляется интересным как с точки зрения научного изучения текстов стратегий социально-экономического развития, так и с точки зрения прикладного использования полученных результатов в практике регионального управления.

Обзор научной литературы

Текст – это хранилище значительного количества информации. При этом применение стандартных алгоритмов *Data Mining* для анализа этой информации не представляется возможным, так как текстовые данные зачастую не структурированы и включают в себя графическую информацию. Поэтому для анализа текстов используется специальный метод интеллектуального анализа текста – *Text Mining*. Учеными детально описаны различные *Text Mining*-системы и реализованные в них функции анализа текста (Пиотровская, 2014; Shobha et al., 2010; Witten et al., 1999; Cunningham, 2002; Bank et al., 2012). Процесс работы с текстами включает в себя несколько последовательных этапов (табл. 1). При этом повышенное внимание уделяется разработке и созданию релевантных текстовых корпусов, под которыми понимается подобранная и обработанная по определенным правилам совокупность текстов, используемых в качестве базы для исследования языка (Гатауллин, 2019).

Таблица 1

Этапы проведения *Text Mining*

N	Этапы	Задачи
1	Формирование текстового корпуса	Сбор коллекции документов, содержащей морфологическую, синтаксическую, семантическую или другую необходимую информацию
2	Предварительная обработка документов	Выполнение необходимых преобразований документов путем удаления стоп-слов, стемминга, N-грамм, приведения регистра
3	Извлечение информации	Определение ключевых понятий для дальнейшего выполнения анализа документа
4	Применение алгоритмов извлечения шаблонов	Объединение частых наборов слов в ключевые понятия с помощью специального алгоритма, идентификация фактов в текстах и извлечение их характеристик
5	Интерпретация результатов	Интерпретация полученных результатов путем их представления на естественном языке, а также визуализации в графическом виде

Источник: Составлена авторами (- и далее, если не указано иное).

Анализ специальной литературы выявил большой потенциал применения интеллектуального анализа текстовых документов (Chetty, 2017). Систематизация возможностей применения инструментария *Text Mining* приведена в Таблице 2.

Таблица 2

Возможности применения инструментария *Text Mining*

N	Задача анализа	Содержательная интерпретация	Авторы
1	Классификация	Определение одной/нескольких заранее заданных категорий, к которым относится документ	Шабанов В., Андреев А. (Шабанов и др., 2003); Кристофер Д. (Christopher et al., 2008); Мэннинг Д. (Manning et al., 2000); Алексеев А. (Алексеев и др., 2016); Фомин В. (Фомин и др., 2017); Котеесваран С. (Koteeswaran et al., 2012)
2	Кластеризация	Автоматическое выявление семантически похожих групп документов среди заданного фиксированного множества	Эндрюс Н. (Andrews et al., 2007); Некрестьянов И. (Некрестьянов, 2000); Сэлтон Г. (Salton, 1989); Чернышова Г., Овчинников А. (Чернышова и др., 2015); Абакумов О. (Абакумов, 2010); Аллахяри М. (Allahyari et al., 2017); Майнер Г. (Miner et al., 2012)
3	Автоматическое аннотирование	Сокращение текста при сохранении его смысла (результат – наиболее значимые предложения в тексте)	Хан У., Мани И. (Hahn et al., 2000); Сэлтон Г. (Salton, 1997); Круенгкрай, Ярускулчай Ц. (Kruengkrai et al., 2003)
4	Извлечение ключевых понятий	Идентификация фактов и отношений в тексте при помощи использования словарей (имена существительные и нарицательные: имена и фамилии людей, названия организаций и др.)	Гришман Р. (Grishman, 1997); Кузык М., Симачев Ю., Власенко А. (Kuzyk et al., 2019); Эон Й., Чое Й. (Jeon et al., 2009); Литвинова Т., Литвинова О., Заварзина Г. (Litvinova et al., 2019); Бейкер П., Вессей Р. (Baker et al., 2018)
5	Навигация по тексту	Обеспечение пользователям возможности перемещаться по документам относительно тем и значимых терминов (за счет идентификации отношений между ключевыми понятиями)	Каннинггэм Х. (Cunningham, 2002); Бэнк М., Счиерле М. (Bank et al., 2012)
6	Анализ трендов	Идентификация трендов в наборах документов	Хуртадо Й., Агарвал А. (Hurtado et al., 2016); Ким Мин-Эонг (Kim et al., 2017)
7	Поиск ассоциаций	Идентификация ассоциативных отношений между ключевыми понятиями	Тиен Дунг До (Tien et al., 2006); Рагхаван П., Тсапарас П. (Raghavan et al., 2002)

Важно отметить, что в большинстве рассмотренных нами работ, авторы которых применяют инструментарий *Text Mining*, в качестве исследовательских задач заявлены классификация и кластеризация текста.

Вместе с тем применительно к объекту исследования настоящей работы – текстовому корпусу стратегий социально-экономического развития –

основной задачей *Text Mining* является извлечение ключевых понятий. Несмотря на то что решение данной задачи путем интеллектуального анализа текста не получило широкого распространения в экономических публикациях, отметим ряд работ из разных областей науки, оказавших непосредственное влияние на настоящее исследование. Так, в работах П. Бейкер и Р. Вессей (Baker et al., 2018) и Т. Литвиновой, О. Литвиновой и Г. Завариной (Litvinova et al., 2019) анализируется исламистский экстремистский дискурс методом сравнения текстовых корпусов на английском, французском и русском языках. Выявление ключевых терминов, призывающих к насилию, призвано бороться с пропагандой экстремизма. В работе Й. Эон, Й.-Щ. Чое (Jeon et al., 2009) анализируют текстовые корпуса мужской и женской речи с целью сравнения частоты использования интенсифицирующих наречий. Одним из ключевых вопросов указанного исследования стал выбор наиболее подходящего речевого корпуса с помощью проверки разных совокупностей текстов с точки зрения размера и состава.

Отдельно выделим две наиболее близкие к нашей теме работы, объектом исследования в которых выступили нормативно-правовые акты в области территориального развития. Так, Р. Гришман на основе «сквозного» анализа принятых стратегий регионального развития выделяет, с одной стороны, общие акценты государства в отраслевом стратегическом планировании и их изменение с течением времени, с другой – специфику подходов в отношении отдельных отраслей и их групп (Grishman, 1997). Методической особенностью указанного исследования является сочетание качественной экспертной оценки стратегий с применением численных методов анализа текстов (*Text Mining*) и обработки естественного языка (*natural language processing*). В работе И. Е. Калабихиной на основе методов качественного и количественного контент-анализа текстов с использованием программного обеспечения MAXQDA были выявлены «новации в документах в области социально-демографической политики, связанные с различными изменениями в стране в период проведения данной политики» (Калабихина, 2019, с. 21). Проведенный анализ позволил автору сделать выводы об отсутствии преемственности концептуальных подходов в разных анализируемых документах. Вместе с тем в указанных работах интеллектуальный анализ текстов стратегий регионального развития не использовался для оценки соответствия отраслевой специализации территорий, отражающей реальный сектор экономики исследуемых субъектов РФ, приоритетам развития.

Методика исследования

Разработка универсального количественного инструментария определения уникальности целеполагания регионов в текстах стратегий развития относительно их отраслевой специализации на основе *Text Mining* проводилась путем реализации трех последовательных этапов на основе предварительно сформированных данных в виде текстового корпуса. Совокупность текстов, используемых в качестве базы для исследования языка, включила стратегии социально-экономического развития субъектов Российской Федерации.

На первом этапе происходит извлечение ключевых понятий – именованных сущностей из текста стратегий анализируемых регионов. Извлечение ключевых понятий выполняется отдельно для каждой стратегии. Для этого, во-первых, текст стратегии анализируемого региона загружается в машиночитаемом формате (*PDF*, *Word*). Во-вторых, извлекается текст из стратегии с помощью фильтра. В-третьих, извлекаются ключевые слова из текста стратегии с помощью библиотеки *Rutermextract*³ и приводятся в начальную форму русского языка с применением библиотеки *Pymorphy*⁴. Ключевые слова фильтруются таким образом, что остаются словосочетания без числительных и некоторых стоп-слов. Для каждого ключевого слова делается запрос в поисковую систему для получения связанных с ним слов, что позволяет в дальнейшем отождествлять такие понятия, как, например, «кожуун»⁵ и «район». В результате формируется список объектов типа *Term*, в котором исключены текстуальные повторы и словоформы. В объекте *Term* содержится информация о списке слов, количестве слов, частоте употреблений. Таким образом, на первом этапе получается список именованных сущностей (*E*) текста стратегии развития каждого рассмотренного региона (*R*).

На втором этапе проводится количественный анализ текста стратегий. Каждый текст стратегии анализируется с учетом списка ключевых слов *W*, заданных пользователем. Каждое из заданных пользователем ключевых слов $w_j \in W$ ($j = \overline{1, n}$, (n – количество ключевых слов, заданных пользователем) сравнивается с каждой сущностью $e_i \in E$ каждого региона $r_k \in R$ и рассчитывается декартово расстояние d между ними. Далее оценивается степень выраженности p ключевого слова w_j как $p = 1 - i/L$, где i – позиция в списке сущностей (чем ближе к 0, тем значимее), L – мощность множества сущностей текста стратегии развития данного региона. Таким образом, выраженность сущностей в начале списка стремится к 1, в конце – к 0. Все вхождения p помещаются в список *F*. Если список вхождений *F* не пустой по окончании обработки, то производится расчет среднего значения $m = \overline{F}$, которое помещается в список *M* показателей выраженности ключевых слов, заданных пользователем, в стратегии исследуемого региона. Таким образом, в результате второго этапа получается список выраженности показателей ключевых слов, заданных пользователем, в стратегиях исследуемых регионах.

На заключительном третьем этапе проводится количественный и качественный сравнительный анализ текстов стратегий выбранных регионов и осуществляется интерпретация полученных результатов.

Описанные этапы анализа текстовых корпусов с использованием инструментария *Text Mining* являются универсальными и могут быть применены к любым текстовым объектам анализа. Вместе с тем актуальной методической проблемой оценивания текстов стратегий территориального развития является выбор объектов сравнения в виде регионов. В релевант-

³ Term extraction for Russian language. 2018. URL: <https://github.com/igor-shevchenko/rutermextract>.

⁴ Морфологический анализатор pymorphy2. URL: <https://pymorphy2.readthedocs.io/en/latest/>.

⁵ Наименование муниципальных районов в Республике Тыва.

ных отечественных исследованиях, посвященных анализу эффективности территориальных стратегий, выбор авторами субъектов Российской Федерации, нормативно-правовых стратегических документов, которые в дальнейшем оцениваются, обосновывается различными характеристиками: географическими, когда авторы анализируют стратегии регионов определенного федерального округа (Жихаревич и др., 2011; Селиверстов и др., 2011); отраслевыми, когда анализируются стратегии развития инновационного, инвестиционного, промышленного секторов региональной экономики (Kutsenko et al., 2018; Николаев и др., 2016; Шулекина, 2014); хронологическими, когда анализируются региональные законы о стратегическом планировании, принятые в разные периоды времени (Жихаревич и др., 2015). Отдельно стоит отметить научные работы, включающие сплошной анализ региональных стратегий по всей совокупности субъектов Российской Федерации (Ильина и др., 2016; Шпакова, 2019).

В настоящей работе, целью которой явилось определение уникальности целеполагания регионов в текстах стратегий развития, выбор регионов производился на основе гипотетического предположения о том, что для территорий с непохожими институциональными условиями развития (включая преобладание разных отраслевых секторов в экономике), характерны различные приоритеты развития. Для проверки данной гипотезы нами были отобраны стратегии десяти максимально отличающихся субъектов Российской Федерации от анализируемого региона, в качестве которого выступила Республика Татарстан, занимающая лидирующие позиции согласно большинству ведущих рейтингов развития субъектов Российской Федерации⁶.

Отличие отобранных территорий от Республики Татарстан обуславливается значениями индексов структурных расстояний, рассчитанных на основе методики «умного» бенчмаркинга регионов (более подробно методика «умного» бенчмаркинга представлена в публикациях: Dubrovskaya et al., 2018; Kurushin et al., 2019; Dubrovskaya et al., 2019; Dubrovskaya et al., 2019; Русинова, 2020). Отметим, что указанная методика основана на универсальном подходе к планированию деятельности, объединяющем постулаты нескольких научных школ стратегического менеджмента: дизайна, обучения, конфигурации, бенчмаркинга (более подробно систематизация подходов ведущих научных школ стратегического менеджмента, основанная на результатах исследования (Минцберг, 2015), представлена в публикации (Дубровская, 2018)). В процессе проведения «умного» бенчмаркинга для каждо-

⁶ Рейтинг инновационных регионов России. 2018. URL: <http://i-regions.org/images/files/airr18.pdf>. Национальный рейтинг состояния инвестиционного климата в субъектах Российской Федерации. URL: <https://asi.ru/investclimate/rating/>.

Рейтинг социально-экономического положения субъектов Российской Федерации. 2018. URL: <https://riarating.ru/infografika/20180523/630091878.html>.

Уровень развития науки и технологий в регионах России. 2018. URL: <https://riarating.ru/infografika/20181017/630109152.html>.

Индекс кредитоспособности регионов. 2018. URL: http://vid1.rian.ru/ig/ratings/region_credit_2018.pdf.

Качество жизни в российских регионах. 2018. URL: <https://riarating.ru/infografika/20190219/630117422.html>. Выборка показателей была произведена за 2017 г.

го субъекта Российской Федерации рассчитываются индексы структурных расстояний на основе системы показателей, отражающих институциональные условия развития территории, ключевую роль в которых играет отраслевая структура экономики (формула 1).

$$d(i, i') = \sum_{p=1}^P W_p (P_{re} - P_{ri})^2 \quad (1)$$

где $d(i, i')$ – индекс структурного расстояния i – региона;

P_{re} – значение r -показателя e -исходного региона;

P_{ri} – значение r -показателя i – «другого» региона;

W_p – весовой коэффициент.

Система оценки институциональных условий территорий включает 34 статистических количественных показателя, объединенных в семь критериев регионального развития (геодемография, образование, инновации, отраслевая структура, инвестиционный климат, открытость, общественные ценности). Автоматическая реализация процедуры «умного» бенчмаркинга доступна в режиме реального времени на сайте <http://ruclusters.ru>. Программное средство размещено в публичном доступе и зарегистрировано в Реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности под номером 2020616804⁷.

Результаты исследования

Для оценки соответствия приоритетов стратегического развития регионов их отраслевой специализации, как отмечалось выше, нами были выбраны стратегии десяти максимально отличающихся субъектов Российской Федерации от Республики Татарстан. Выбор производился на основе результатов проведения процедуры «умного» бенчмаркинга для Республики Татарстан (Tat). В результате нами были выявлены следующие наиболее отличные от Республики Татарстан территории: Республика Ингушетия (Ing), Кабардино-Балкарская Республика (Kab), Республика Крым (Cr), Республика Тыва (Tuv), Республика Адыгея (Adu), Республика Калмыкия (Kai), Республика Дагестан (Dag), Карачаево-Черкесская Республика (Kao), Чеченская Республика (Cha), Республика Алтай (Ali). Отметим, что большинство проанализированных стратегий были приняты на период до 2030 г. (Adu, Ing, Kai, Cr, Tuv, Tat), стратегия социально-экономического развития Dag и Cha – до 2025 г., Ali и Kao – до 2035 г. и Kab – до 2034 г.

Указанные десять регионов имеют наибольший индекс структурного расстояния относительно Республики Татарстан (см. технологию «умного» бенчмаркинга в (Dubrovskaya et al., 2018)), т.е. являются самыми непохожими на нее по институциональным параметрам развития. Несхожесть полученных

⁷ Русинова М.Р., Курушин Д.С., Дубровская Ю.В., Козоногова Е.В., Ларионов А.А. Программа для проведения процедуры «умного» бенчмаркинга регионов Российской Федерации («Умный» бенчмаркинг регионов России): Свидетельство о государственной регистрации программы для ЭВМ № 2020616804 от 23.06.2020 (РФ).

регионов относительно анализируемого субъекта Российской Федерации (Республика Татарстан) подтверждается всей выборкой статистических показателей: геодемографическими параметрами (площадь территории, плотность расселения, структура населения и т.д.); уровнем инвестиционной привлекательности, инновационности, образованности населения и пр.

Отдельно остановимся на различиях в отраслевой структуре регионов. Именно данный показатель является принципиальным аспектом настоящего исследования, обуславливающим дифференциацию регионального стратегирования. Так, в выбранных регионах прослеживаются значимые различия в отраслевой специализации. Республика Татарстан специализируется на добыче полезных ископаемых: эта отрасль занимает более 30% в общей структуре валового продукта региона и составляет 753 615 млн руб. В Республике Тыва данный вид экономической деятельности также является наиболее развитым (доля в валовом региональном продукте (ВРП) – 25%), однако показатель ВРП по добыче полезных ископаемых составляет всего 3 681 млн руб. В автономной Республике Крым и Республике Дагестан развит такой вид экономической деятельности, как оптовая и розничная торговля, ремонт автотранспортных средств, мотоциклов, бытовых изделий и предметов личного пользования (их доли в совокупном ВРП составляют 17% и 23% соответственно); в Кабардино-Балкарской, Карачаево-Черкесской республиках, а также в Республике Калмыкия лидирует сельское хозяйство. В остальных регионах не прослеживается характерно выраженного вида экономической деятельности: показатель совокупного валового регионального продукта распределен между отраслями в пределах 10–15%.

Описанные различия исследуемых регионов обуславливают объективную возможность проверки сформулированной нами гипотезы исследования: *для территорий с преобладанием различных отраслевых секторов в экономике характерны различные приоритеты развития.*

Итоговый сформированный текстовый корпус по 11 региональным стратегиям развития включил 415 780 слов. Далее в результате определения уникальности целеполагания регионов в текстах стратегий развития относительно их отраслевой специализации с помощью *Text Mining* нами были получены списки именованных сущностей текста стратегии развития для каждого региона (всего получено 11 списков).

На основе полученных данных было проведено сравнение текстов анализируемых стратегий. Авторами было посчитано количество пересекающихся именованных сущностей стратегий по каждой паре анализируемых регионов. Результаты сравнительного анализа приведены в Таблице 3. В ячейках таблицы содержатся выраженные в процентах значения, каждое из которых равно отношению числа пересекающихся именованных сущностей двух регионов к общему числу именованных сущностей в регионе. По строкам таблицы содержится информация о том, на сколько процентов список именованных сущностей региона совпадает со списками других регионов, по столбцам – на сколько процентов список именованных сущностей стратегии рассматриваемого региона выражен в списках других регионов.

Таблица 3

**Результаты сравнения именованных сущностей стратегий
развития регионов, %**

Регион	Ady	Tat	Ali	Dag	Ing	Kab	Kai	Kao	Cr	Tuv	Cha
Ady	X	30,6	22,8	11,6	28,6	14,3	30,8	17,1	29,1	18,3	26,3
Tat	19,9	X	16,3	7,2	20,6	7,9	19,6	10,8	20,1	13,0	17,1
Ali	22,6	25,0	X	9,4	37,0	14,1	30,6	20,0	28,6	23,0	27,1
Dag	35,9	34,3	29,4	X	41,7	22,0	38,8	26,2	36,6	30,7	49,2
Ing	12,0	13,3	15,6	5,6	X	8,3	16,8	11,0	18,8	13,2	18,0
Kab	34,3	29,3	34,1	17,0	47,6	X	40,6	29,1	38,3	30,8	39,8
Kai	21,2	20,8	21,2	4,9	27,7	11,6	X	14,0	23,9	18,3	21,0
Kao	24,4	23,8	28,8	12,1	37,6	17,3	29,2	X	30,4	25,3	30,3
Cr	11,0	11,7	10,9	4,5	17,0	6,0	13,2	8,0	X	10,0	12,7
Tuv	19,5	21,3	24,6	10,5	33,6	13,6	28,3	18,8	28,3	X	23,8
Cha	16,4	16,5	17,1	9,9	26,9	10,4	19,2	13,2	21,1	14,0	X

Так, согласно полученным результатам, список именованных сущностей Республики Дагестан на 49,2% совпадает со списком именованных сущностей Чеченской Республики. В свою очередь, список именованных сущностей стратегии Чеченской Республики лишь на 9,9% совпадает со списком Республики Дагестан. Кроме того, список именованных сущностей стратегии Республики Дагестан на 34,3% совпадает со списком Республики Татарстан, а список именованных сущностей Республики Татарстан в большей степени совпадает со списком стратегии Адыгеи (19,9%). Наиболее «уникальной» является стратегия Республики Крым: ее список именованных сущностей лишь на 13,2% совпадает со списком стратегии Республики Калмыкия.

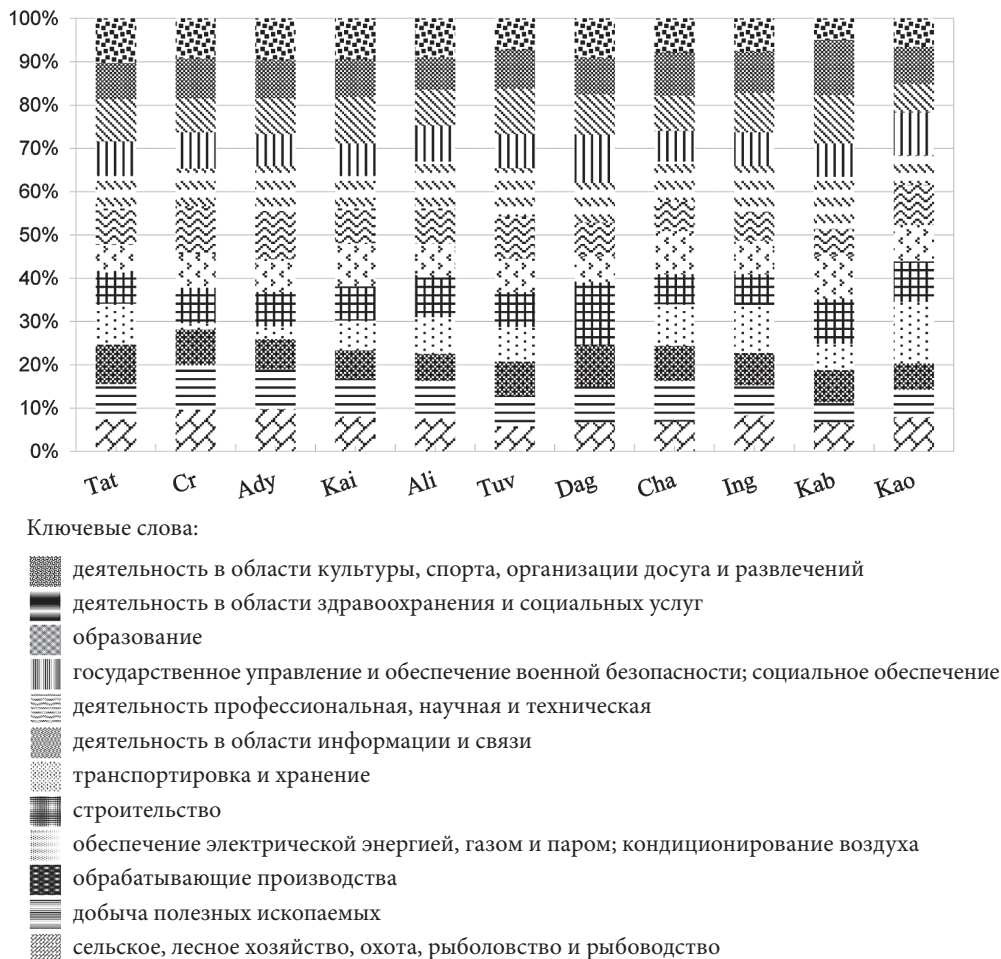
Далее для проверки гипотезы о том, что для территорий с преобладанием различных отраслевых секторов в экономике характерны различные приоритеты развития, нами были выбраны ключевые слова в виде 12 основных разделов общероссийского классификатора видов экономической деятельности, объективно характеризующих отраслевые приоритеты развития регионов⁸. Выбор показателя ОКВЭД обусловлен, во-первых, тем, что именно отраслевая привязка традиционно лежит в основе планирования территориального развития; во-вторых, разделы ОКВЭД позволяют сопоставить полученные результаты с фактической отраслевой структурой ВРП анализируемых регионов; в-третьих, 12 основных разделов ОКВЭД являются объективно достаточной выборкой для проведения настоящего исследования, т.к. увеличение количества ключевых слов приведет к «размыванию» результатов.

⁸ Общероссийский классификатор видов экономической деятельности ОК 029-2014. 2020. URL: http://www.consultant.ru/document/cons_doc_LAW_163320/.

На основе степени выраженности ключевых слов – видов экономической деятельности в списках именованных сущностей – авторами были созданы портреты стратегий социально-экономического развития регионов (рис. 1).

Рисунок 1

Портрет документов стратегий социально-экономического развития регионов

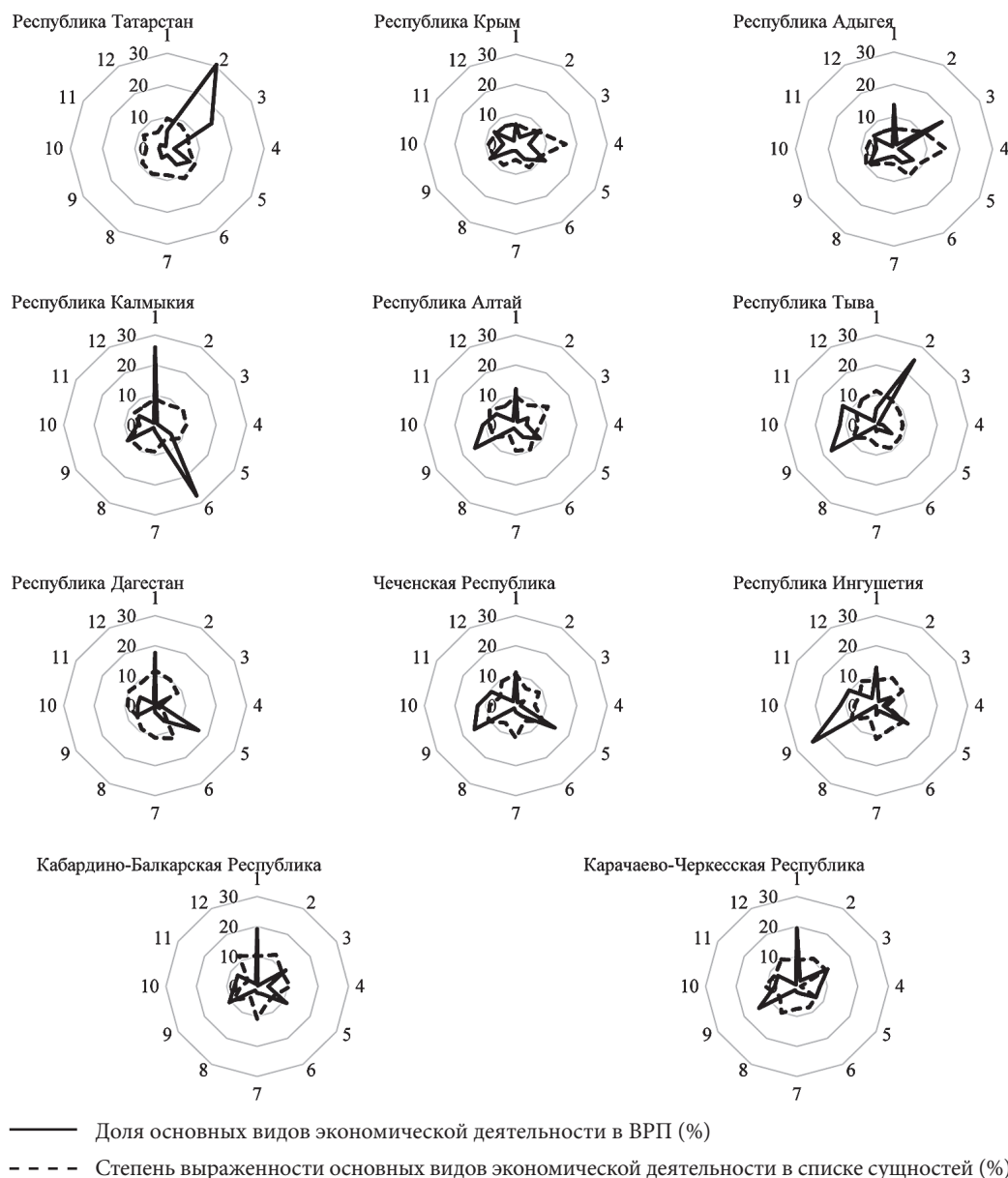


Как видно из рисунка, в целом структура портретов текстовых документов анализируемых стратегий регионов совпадает, что говорит об отсутствии выраженной приоритезации в документах отдельных видов экономической деятельности. Возможно, низкая спецификация территориальных особенностей в текстах стратегий регионов может быть следствием распространенной практики привлечения экспертной группы для подготовки региональной стратегии развития. В результате текст стратегии, подготовленный для одного региона, выступает черновиком для другого.

Далее мы сравнили распределение значений степени выраженности основных видов экономической деятельности в списке сущностей текста стратегий развития регионов, наиболее непохожих на Республику Татарстан (рис. 2).

Рисунок 2

Распределение степени выраженности основных видов экономической деятельности в списке сущностей текста стратегий развития регионов, наиболее непохожих на Республику Татарстан



1 – Сельское, лесное хозяйство, охота, рыболовство и рыбоводство; 2 – Добыча полезных ископаемых; 3 – Обрабатывающие производства; 4 – Обеспечение электрической энергией, газом и паром, кондиционирование воздуха; 5 – Строительство; 6 – Транспортировка и хранение; 7 – Деятельность в области информации и связи; 8 – Деятельность профессиональная, научная и техническая; 9 – Государственное управление и обеспечение военной безопасности; социальное обеспечение; 10 – Образование; 11 – Деятельность в области здравоохранения и социальных услуг; 12 – Деятельность в области культуры, спорта, организации досуга и развлечений.

Источник: Составлен авторами на основе результатов проведения процедуры *Text Mining*.

Согласно результатам анализа степени выраженности основных видов экономической деятельности в списке сущностей, полученным на основе *Text Mining* (%) и отраженным на Рисунке 2, можно сделать вывод о том, что в целом приоритеты развития основных видов деятельности, обозначенные в Стратегиях социально-экономического развития регионов, *не соответствуют структуре их отраслевой специализации*. А так как отраслевая специализация Республики Татарстан и специализация наименее идентичных по отношению к ней регионов принципиально отличаются, то мы можем объективно сделать вывод об отсутствии ярко выраженной приоритезации в территориальных стратегиях развития перспективных видов деятельности.

Например, таким видам деятельности, как сельское хозяйство, охота и рыболовство, традиционно занимающим наибольшие удельные веса в ВРП Республики Адыгея (13,7%), Республика Калмыкия (25,9%) и Карачаево-Черкесская Республика (19,4%), не уделено должного внимания в стратегиях их социально-экономического развития. Степень выраженности основных видов экономической деятельности в списке сущностей, полученных на основе *Text Mining*, составила соответственно 6,4; 8,7 и 8,8%.

К ряду отраслей, наоборот, наблюдается повышенное внимание. Это может быть вполне оправданно в случае с такими видами экономической деятельности, как «Обрабатывающие производства» или «Строительство», когда органы региональной власти совершенно объективно и четко обозначают свое намерение развивать данные направления на этапе подготовке стратегий. Вместе с тем по вполне очевидным причинам считаем нелогичным сверхприоритезацию такого вида деятельности, как «Добыча полезных ископаемых», при сравнительно умеренных уровнях развития и разработки минерально-сырьевой базы. Так, превышение степени выраженности данного вида экономической деятельности в списке сущностей, полученных на основе *Text Mining*, над их удельными весами в ВРП наблюдается в Кабардино-Балкарской Республике – в 40,7 раза; в Республике Дагестан – в 20 раз; в Чеченской Республике – в 5,5 раза.

Сделанный вывод требует отдельного дополнительного исследования, что связано с возможными положительными перспективами развития данной отрасли в будущем. Именно по этой причине для объективной интерпретации полученных результатов нами была дополнительно использована Стратегия пространственного развития Российской Федерации на период до 2025 г., утвержденная распоряжением Правительства Российской Федерации от 13 февраля 2019 г. N 207-р⁹. Указанная стратегия закрепила за каждым субъектом Российской Федерации так называемые «перспективные экономические специализации», под которыми законодателем понимается совокупность укрупненных видов экономической деятельности (отраслей), обусловленных благоприятным сочетанием конкурентных преимуществ (пространственных факторов размещения видов экономической деятель-

⁹ Распоряжение Правительства Российской Федерации от 13 февраля 2019 г. N 207-р (ред. от 31.08.2019) «Об утверждении Стратегии пространственного развития Российской Федерации на период до 2025 года».

ности). При этом отмечается, что стратегирование регионального развития должно осуществляться в рамках реализации закрепленных перспективных экономических специализаций субъектов Российской Федерации.

Так, проведенный нами анализ экономических специализаций, установленных для каждого региона Стратегией пространственного развития Российской Федерации на период до 2025 г. в качестве приоритетных, выявил, что, например, за Кабардино-Балкарской Республикой «Добыча полезных ископаемых» закреплена в качестве перспективной отрасли в связи с наличием значительных запасов полезных ископаемых. Вместе с тем результаты аналогичного сравнения по таким регионам, как Чеченская Республика и Республика Дагестан, не опровергли наш вывод о необоснованности приоритезации в данных субъектах Российской Федерации добычи полезных ископаемых.

Продолжая рассуждение по этому, на наш взгляд, крайне важному вопросу, мы не можем не отметить, что указанное несоответствие имеемого и желаемого (по крайней мере в части добычи полезных ископаемых) может иметь успешное решение только в результате интенсификации развития науки и техники, что напрямую прописывается в качестве основной задачи в Стратегии социально-экономического развития Кабардино-Балкарской Республики до 2034 г.: «Задача 2. Совершенствование технологии добычи, транспортировки и повышение глубины переработки углеводородного сырья. Недостаточная степень технической перевооруженности мощностей по переработке углеводородных ресурсов ведет к невозможности достижения глубины переработки нефти, при которой произведенные нефтепродукты смогут успешно продаваться на рынке»¹⁰. Вместе с тем степень выраженности профессиональной, научной и технической деятельности в списке сущностей, полученных на основе *Text Mining*, составила для Кабардино-Балкарской Республики всего лишь 4,3%. Очевидно, что актуализированные расхождения требуют анализа более широкого перечня документов и повышенного внимания как со стороны исследователей, так и со стороны соответствующих органов власти.

Отдельно остановимся на результатах, полученных нами в ходе анализа экономической деятельности в области информации и связи. Так, по всем анализируемым регионам степень выраженности данного вида деятельности в списке сущностей, полученных на основе *Text Mining*, в разы превышает удельный вес указанной отрасли в общем объеме ВРП. Вместе с тем проведенный нами анализ экономических специализаций, установленных для каждого региона Стратегией пространственного развития Российской Федерации на период до 2025 г.¹¹ в качестве приоритетных, выявил, что деятельность в области информации и связи в качестве перспективной отрасли

¹⁰ Распоряжение Правительства Кабардино-Балкарской Республики от 30 декабря 2016 года N 727-рп (с изменениями на 16 февраля 2018 г.) «Об утверждении Стратегии социально-экономического развития Кабардино-Балкарской Республики до 2034 года».

¹¹ Распоряжение Правительства Российской Федерации от 13 февраля 2019 г. N 207-р (ред. от 31.08.2019) «Об утверждении Стратегии пространственного развития Российской Федерации на период до 2025 года».

закреплена только за Республиками Крым, Дагестан и Татарстан. Отметим, что полученные нами результаты согласуются с опубликованными ранее Е. Куценко, Е. Исланкиной, А. Киндрась (Kutsenko et al., 2018) оценками стратегий развития семи субъектов Российской Федерации, показавшими необоснованность выбора большинством регионов отрасли информационно-телекоммуникационных технологий (далее – ИКТ) в качестве приоритетной. Как отмечают ученые, «только в одном из пяти регионов, выбравших ИКТ в качестве инновационного приоритета, уровень развития данной отрасли выше среднероссийского значения (Республика Татарстан). В других регионах уровень развития ИКТ оказался ниже среднего по стране, а в ряде субъектов – даже среднего по федеральному округу» (Kutsenko et al., 2018). Понимая специфику функционирования системы ИКТ, обусловленную отсутствием территориальных и географических ограничений деятельности, присущих, например, добывающим отраслям, отметим, что ИКТ являются залогом инновационности развития любой территории. Поэтому в целом в долгосрочных стратегиях развития приоритезация деятельности в области информации и связи объективно является актуальной задачей органов власти, при решении которой все же должны учитываться возможности и уровень развития ИКТ конкретной территории.

В заключение остановимся на результатах анализа, полученных по группе видов экономической деятельности, финансируемых, по большей части, из бюджетов бюджетной системы Российской Федерации: 1. Государственное управление и обеспечение военной безопасности; социальное обеспечение; 2. Образование; 3. Деятельность в области здравоохранения и социальных услуг. Важно отметить, что в результате интеллектуального анализа текстов стратегий по данным направлениям практически не было выявлено расхождений между степенью выраженности этих видов деятельности в списке сущностей и их удельными весами в ВРП. Считаем, что это обусловлено примерно одинаковыми уровнями значимости социальных направлений деятельности во всех субъектах Российской Федерации. Таким образом, ничуть не умаляя важность указанных направлений социального развития, мы можем сделать заключение о корректности применяемого нами инструментария *Text Mining* для определения уникальности целеполагания регионов в текстах стратегий развития. Действительно, *недостаточное внимание к вопросу приоритезации развития ведущих отраслей экономической специализации в стратегических планах субъектов Российской Федерации отражается на качестве формирования системы перспективных направлений развития региона, что может являться одной из причин недостижения запланированных целевых индикаторов развития территории.*

Полученные в исследовании выводы добавляют новые эмпирические свидетельства в систему научных работ, посвященных вопросам анализа осмысленности текстов регионального стратегирования, и таким образом поддерживают дискуссию о важности уникального целеполагания в части идентификации перспективных специализаций в стратегиях развития субъектов Российской Федерации.

Заключение

В настоящей работе проведена оценка соответствия приоритетов стратегического развития регионов их отраслевой специализации на основе *Text Mining*. После детального анализа релевантной литературы нами было выявлено, что технология интеллектуального анализа текста *Text Mining* может применяться для анализа текстов стратегий территориального развития. Также было обосновано, что функционал *Text Mining* представляется интересным не только с научной точки зрения, но и в части прикладного использования полученных результатов в практике регионального управления.

При этом выбор регионов для анализа текстов стратегий проводился с помощью инструментария «умного» бенчмаркинга, основанного на универсальном подходе к планированию деятельности и объединяющего постулаты различных научных школ стратегического менеджмента. Данный инструментарий позволяет идентифицировать эффективные направления инвестирования и объекты первоочередного бюджетного финансирования. Таким образом, теоретическая значимость исследования заключается в развитии концептуальных подходов ведущих научных школ стратегического менеджмента применительно к вопросам регионального стратегирования.

Практическая значимость работы заключается в разработке универсального для любой области науки инструментария проведения интеллектуального анализа текстов, включающего разработку программного кода реализации процедуры *Text Mining*. В результате интеллектуального анализа текстов стратегий развития регионов было представлено эмпирическое свидетельство того, что недостаточность уникального целеполагания в части приоритезации развития ведущих отраслей экономической специализации, представляющего собой процесс выбора и обоснования целей социально-экономического стратегирования и основанного на предварительном анализе исходных институциональных условий функционирования региона, негативно влияет на качество формирования системы перспективных направлений развития региона. Данный вывод был подтвержден количественными результатами сравнения именованных сущностей стратегий развития 11 субъектов Российской Федерации. Выявлено, что анализируемые тексты стратегий регионального развития не отражают приоритеты развития, согласующиеся с отраслевой специализацией территорий.

Говоря о ценности полученных результатов для практики стратегического управления, отметим, что разработанные алгоритмы извлечения именованных сущностей из текста и представленные алгоритмы *Text Mining* являются частью универсальной методики количественной оценки приоритетов текстовых документов любого содержания и открывают широкие горизонты дальнейших исследований в области анализа текстов программ и стратегий территориального развития, в том числе в части проверки их соответствия федеральным приоритетам развития.

СПИСОК ИСТОЧНИКОВ

1. Абакумов О. Использование кластеризации в text mining // Новые информационные технологии в автоматизированных системах. 2010. № 13. С. 128–129.
2. Алексеев А.А., Катасев А.С., Кириллов А.Е., Кирпичников А.П. Классификация текстовых документов на основе технологии Text Mining // Вестник Казанского технологического университета. 2016. Т. 18, № 19. С. 116–119.
3. Батов Г.Х. Стратегическое планирование как инструмент стратегического управления региональной социально-экономической системой // МИР (Модернизация. Инновации. Развитие). 2016. Т. 7, № 2. С. 212–218.
4. Боровская Л.В. Сравнительный анализ результатов социально-экономического развития республики Крым на основе метода аналогий // Вестник ВГУ. Серия: Экономика и управление. 2019. № 3. С. 51–58.
5. Гатауллин Р.Р. Методы, модели и программный инструментарий разрешения многозначности в текстах: автореферат дис. ... кандидата технических наук. Казан. (Приволж.) федер. ун-т. Казань, 2019.
6. Дубровская Ю.В., Кудрявцева М.Р. Систематизация теоретических подходов к формированию стратегии регионального развития // Государственное управление. Электронный вестник. 2018. № 68. С. 155–171.
7. Егоров К.Ю. Оценка эффективности реализации стратегии социально-экономического развития региона // Вестник Самарского государственного экономического университета. 2010. Т. 72, № 10. С. 10–16.
8. Жихаревич Б.С., Крыловский А.Б. Опыт сравнительной оценки качества стратегического планирования в регионах Северного Кавказа // Экономика развития региона: проблемы, поиски, перспективы. 2011. № 12. С. 217–228.
9. Жихаревич Б.С., Лебедева Н.А. Трансляция идей трансформации регионального пространства в документы стратегического планирования // Экономика Северо-Запада: проблемы и перспективы развития. 2015. Т. 49, № 4. С. 58–74.
10. Ильина И.Н., Кириллова А.Н., Плисецкий Е.Е., Копыченко Г.С., Рыбина Е.Г. Оценка эффективности реализации региональных целевых программ (на примере Тюменской области) // Региональная экономика: теория и практика. 2014. № 46. С. 37–55.
11. Ильина И.Н., Плисецкий Е.Е., Копыченко Г.С., Рыбина Е.Г. Оценка качества разработки региональных стратегий развития в России // Региональная экономика: теория и практика. 2016. Т. 427, № 4. С. 178–196.
12. Калабихина И.Е. Современная социально-демографическая политика в России: есть ли преемственность в концептуальных подходах в документах 2007–2017 гг. // Женщина в российском обществе. 2019. № 4. С. 14–28.

13. Кислицына В.В., Чеглакова Л.С., Караулов В.М., Чикишева А.Н. Формирование комплексного подхода к оценке социально-экономического развития регионов // Экономика региона. 2017. Т. 13, № 2. С. 369–380.
14. Климанов В.В., Будаева К.В., Чернышова Н.А. Промежуточные итоги стратегического планирования в регионах России // Экономическая политика. 2017. Т. 12, № 5. С. 104–127.
15. Комаров В.М., Акимова В.В., Коцюбинский В.А., Земцов С.П. Сравнительный анализ подходов к разработке долгосрочных государственных стратегий в России и мире // Вопросы государственного и муниципального управления. 2021. № 1. С. 56–74.
16. Липина С.А., Смирнова О.О. Стратегическое планирование в субъектах Российской Федерации: методологические основы и методические рекомендации // Региональная экономика. Юг России. 2017. Т. 15, № 1. С. 25–35.
17. Марголин А.М. Пути совершенствования методов оценки эффективности государственных программ // Экономическая политика. 2018. № 6. С. 54–81.
18. Минцберг Г., Альстранд Б., Лампель Ж. Стратегическое сафари. Экскурсия по дебрям стратегического менеджмента. 2-е изд. М: Альпина Паблишер, 2015.
19. Морозова Н.В., Бондаренко Н.В. Подходы к оценке эффективности реализации региональных стратегий социально-экономического развития // Вестник Чувашского университета. 2014. № 3. С. 178–182.
20. Некрестьянов И.Г. Тематико-ориентированные методы информационного поиска. Диссертация. СПб.: 2000.
21. Николаев М.А., Махотаева М.Ю. Анализ стратегий развития промышленного комплекса региона // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Экономические науки. 2016. Т. 246, № 4. С. 29–41.
22. Пиотровская К.Р. Текст-майнинг: перспективы развития // Известия Российского государственного педагогического университета им. А.И. Герцена. 2014. № 168. С. 128–134.
23. Репичев А.И., Тугачева Л.В., Воробьева А.В., Авдеева Д.А. Возможности разработки региональных инновационных стратегий на принципах «Умной специализации» // Вопросы управления. 2018. Т. 32, № 2. С. 37–45.
24. Рисин И.Е. Методический подход к оценке качества региональной стратегии // Регион: системы, экономика, управление. 2021. Т. 55, № 4. С. 68–73. DOI: 10.22394/1997-4469-2021-55-4-68-73.
25. Русинова М.Р. Алгоритмизация процедуры проведения «умного» бенчмаркинга регионов Российской Федерации // Вестник Пермского национального исследовательского политехнического университета. Социально-экономические науки. 2020. № 2. С. 286–292.

26. Селиверстов В.Е., Мельникова Л.В. Анализ состояния стратегического планирования в регионах Сибирского федерального округа // Регион: Экономика и Социология. 2011. № 3. С. 3–21.
27. Суховой А.Ф., Голова И.М. Обоснование трансформации приоритетов инновационно-технологического развития регионов Российской Федерации в условиях глобального кризиса // Экономика региона. 2016. Т. 12, № 3. С. 911–923.
28. Фомин В.В., Флегонтов А.В., Осочкин А.А. Метод частотно-морфологической классификации текстов // Программные продукты и системы. 2017. Т. 3, № 30. С. 478–486.
29. Чернышова Г.Ю., Овчинников А.Н. Применение методов интеллектуального анализа данных для кластеризации текстовых документов // Информационная безопасность регионов. 2015. Т. 21, № 4. С. 5–12.
30. Шабанов В.И., Андреев А.М. Метод классификации текстовых документов, основанный на полнотекстовом поиске // Труды РОМИП'2003. СПб: НИИ Химии СПбГУ, 2003. С. 52–71.
31. Шпакова Р.Н. Стратегии социально-экономического развития регионов: от оценки достижения целей – к оценке эффективности // Вестник Евразийской науки. 2019. Т. 11, № 2. С. 59.
32. Шулекина Е.Н. К вопросу о соответствии инвестиционных стратегий регионов требованиям инвестиционного Стандарта // Сибирская финансовая школа. 2014. Т. 106, № 5. С. 154–159.
33. Allahyari M., Pouriyeh S., Assefi M., Safaei S., Trippe E.D., Gutierrez, J.B., Kochut, K.A. Brief survey of text mining: classification, clustering and extraction techniques // KDD Bigdas. 2017. URL: https://www.researchgate.net/publication/318336890_A_Brief_Survey_of_Text_Mining_Classification_Clustering_and_Extraction_Techniques/references.
34. Andrews N.O., Fox E.A. Recent Developments in Document Clustering. Department of Computer Science. Virginia Tech, Blacksburg, VA 24060 in October 16, 2007.
35. Baker P., Vessey R. A corpus-driven comparison of English and French Islamist extremist texts // International Journal of Corpus Linguistics. 2018. No. 23. P. 255–278.
36. Bank M., Schierle M. A Survey of Text Mining Architectures and the UIMA Standard // Proceedings of the Eight International Conference on Language Resources and Evaluation LREC'12. 2012. P. 3480–3486.
37. Chetty P. Text mining as a better solution for analyzing unstructured data. 2017. URL: <https://www.projectguru.in/publications/text-mining-analyzing-unstructured-data/>.
38. Manning C.D., Raghavan P., Schütze H. An Introduction to Information Retrieval. NY, Cambridge University Press New York, 2008.

39. Cunningham H. GATE, a General Architecture for Text Engineering // Computers and the Humanities. 2002. Vol. 36, no. 2. P. 223–254.
40. Dubrovskaya J., Rusinova M., Zhulanov E. Smart Benchmarking as a Tool for the Implementation of the Industrial Policy of the Territory // *Ekonomicke, politicke a pravne otazky medzinarodnych vzťahov 2019* (Economic, Political and Legal Issues of International Relations 2019). Zbornik vedeckych prac. 2019. P. 107–114.
41. Dubrovskaya J.V., Rusinova M.R., Kozonogova E.V. Advantages of Modeling a Regional Development Strategy Based on «Smart» Benchmarking // *Education Excellence and Innovation Management through Vision 2020: Proceedings of the 33rd International Business Information Management Association Conference (IBIMA)*. USA, IBIMA Publishing LLC. 2019. P. 3976–3982.
42. Dubrovskaya Yu.V., Kudryavtseva M.R., Kozonogova E.V. “Smart” benchmarking as a basis for strategic planning in regional development // *Economic and Social Changes: Facts, Trends, Forecast*. 2018. Vol. 11, no. 3. P. 100–116.
43. Grishman R. Information extraction: Techniques and challenges // *Information extraction a multidisciplinary approach to an emerging information technology*. 1997. Vol. 1299. P. 10–27.
44. Hahn U., Mani I. The challenges of automatic summarization // *IEEE Computer*. 2000. Vol. 33, no. 11. P. 29–36.
45. Hurtado J.L., Agarwal A., Zhu X. Topic discovery and future trend forecasting for texts // *Journal of Big Data*. 2016. Vol. 3, no. 7. URL: <https://link.springer.com/article/10.1186/s40537-016-0039-2>.
46. Jeon J., Choe J.-W. A Key Word Analysis of English Intensifying Adverbs in Male and Female Speech // *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*. 2009. Vol. 1. P. 210–219.
47. Kim M.-J., Ohk K., Moon Ch.-S. Trend analysis by using text mining of journal articles regarding consumer policy // *New Physics: Sae Mulli*. 2017. Vol. 67, no. 5. P. 555–561.
48. Koteeswaran S., Visu P., Janet J. A review on clustering and outlier analysis techniques in Data Mining // *American Journal of Applied Sciences*. 2012. No. 9. P. 254–258.
49. Kruengkrai C., Jaruskulchai C. Generic text summarization using local and global properties of sentences // *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence*. Washington, DC, USA. 2003.
50. Kurushin D., Dubrovskaya J., Rusinova M., Kozonogova E. Automation and Algorithmization of Smart Benchmarking of Territories Based on Data Parsing // *Digital Science 2019: Conf. on Digital Science (DSIC 2019)*. 2019. P. 126–134.
51. Kutsenko E., Islankina E., Kindras A. Smart by Oneself? An Analysis of Russian Regional Innovation Strategies within the RIS3 Framework // *Foresight and STI Governance*. 2018. Vol. 12, no. 1. P. 25–45.

52. Kuzyk M., Simachev Yu., Vlasenko A. Strategies for the development of Russian industries and sectors: is there room for structural change? (in Russian) // XX April International Academic Conference on Economic and Social Development. 2019. URL: https://www.researchgate.net/publication/333783636_Strategies_for_the_development_of_Russian_industries_and_sectors_is_there_room_for_structural_change_in_Russian.
53. Litvinova T., Litvinova O., Zavarzina G. Two Views on the 2010 Moscow Metro Bombings: Corpus-Based Contrastive Keyword Analysis // 4th International Conference DTGS: International Conference on Digital Transformation and Global Society. 2019. P. 393–404.
54. Manning D., Schutze H. Foundations of statistical natural language processing. Cambridge, MIT Press, 2000.
55. Miner G., Elder J. IV, Hill T. Practical text mining and statistical analysis for non-structured text data applications. Academic Press, 2012.
56. Raghavan P., Tsaparas P. Mining significant associations in large scale text corpora // Proceedings of IEEE International Conference on Data Mining. 2002. P. 402–409.
57. Salton G. Automatic Text Processing the Analysis and retrieval of Information by Computer. Addison Wesley, 1989.
58. Salton G. Automatic Text Structuring and Summarization // Information Processing & Management. 1997. Vol. 33, no. 2. P. 193–207.
59. Shobha S.R., Thakore D. Text mining using keyphrase extraction // Indian Journal of Computer Science and Engineering. 2010. Vol. 1, no. 2. P. 82–85.
60. Tien D.D., Siu Ch. H., Alvis C.M. Fong Associative Feature Selection for Text Mining // International Journal of Information Technology. 2006. Vol. 12, no. 4. P. 59–68.
61. Türegün N. Text Mining in Financial Information // Current Analysis on Economics & Finance. 2019. Vol. 1. P. 18–26.
62. Witten I.H., Paynter G.W., Frank E., Gutwin C., Nevill-Manning C.G. KEA: Practical Automatic Keyphrase Extraction // ACM DL. 1999. No. 8. P. 254–255.

REFERENCES

1. Abakumov, O. (2010) 'New information technologies in automated systems', 13, pp. 128–129. (In Russian).
2. Alekseev, A.A., Katasev, A.S., Kirillov, A.E. and Kirpichnikov, A.P. (2016) 'Classification of text documents based on Text Mining technology', *Vestnik Kazanskogo tekhnologicheskogo universiteta* [Bulletin of the Technological University], 19(18), pp. 116–119. (In Russian).

3. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K. (2017) 'A brief survey of text mining: Classification, clustering and extraction techniques', *KDD Bigdas*. Available at: https://www.researchgate.net/publication/318336890_A_Brief_Survey_of_Text_Mining_Classification_Clustering_and_Extraction_Techniques/references (accessed 20 October 2021).
4. Andrews, N.O. and Fox, E.A. (2007) 'Recent developments in document clustering', *Department of Computer Science*. Virginia Tech, Blacksburg, VA 24060 in October 16.
5. Baker, P. and Vessey, R. (2018) 'A corpus-driven comparison of English and French Islamist extremist texts', *International Journal of Corpus Linguistics*, 23, pp. 255–278.
6. Bank, M. and Schierle, M. (2012) 'A Survey of Text Mining Architectures and the UIMA Standard', *Proceedings of the Eight International Conference on Language Resources and Evaluation LREC'12*, pp. 3480–3486.
7. Batov, G.Kh. (2016) 'Strategic planning as a tool of strategic management of regional socio-economic system', *MIR (Modernizatsiya. Innovatsii. Razvitie)*, 7(2), pp. 212–218. (In Russian).
8. Borovskaya, L.V. (2019) 'The comparative analysis of results of social and economic development of the republic of Crimea on the basis of the analog method', *Vestnik VGU. Seriya: Ekonomika i upravlenie*, 3, pp. 51–58. (In Russian).
9. Chernyshova, G.Yu. and Ovchinnikov, A.N. (2015) 'Using data mining methods for text document cluster analysis', *Informatsionnaya bezopasnost' regionov*, 4 (21), pp. 5–12. (In Russian).
10. Chetty, P. (2017) *Text mining as a better solution for analyzing unstructured data*. Available at: <https://www.projectguru.in/publications/text-mining-analyzing-unstructured-data/> (accessed 20 October 2021).
11. Christopher, D. (2008) 'Manning, Prabhakar Raghavan, Hinrich Schütze', *An Introduction to Information Retrieval*. NY: Cambridge University Press New York.
12. Cunningham, H. (2002) 'GATE, a General Architecture for Text Engineering', *Computers and the Humanities*, 36, pp. 223–254.
13. Dubrovskaya, J., Rusinova, M. and Zhulanov, E. (2019) 'Smart benchmarking as a tool for the implementation of the industrial policy of the territory', *Ekonomicke, politicke a pravne otazky medzinarodnych vzťahov 2019 [Economic, Political and Legal Issues of International Relations 2019]*. In: *Zbornik vedeckych prac*, pp. 107–114.
14. Dubrovskaya, J.V., Rusinova, M.R. and Kozonogova, E.V. (2019) 'Advantages of modeling a regional development strategy based on «smart» benchmarking', *Education Excellence and Innovation Management through Vision 2020: Proceedings of the 33rd International Business Information Management Association Conference (IBIMA)*, pp. 3976–3982.
15. Dubrovskaya, Yu.V. and Kudryavtseva, M.R. (2018) 'Systematization of theoretical approaches to the formation of a regional development strategy', *Gosudarstvennoe upravlenie. Elektronnyy vestnik [Public Administration. E-journal.]*, 68, pp. 155–171. (In Russian).

16. Dubrovskaya, Yu.V., Kudryavtseva, M.R. and Kozonogova, E.V. (2018) “Smart” benchmarking as a basis for strategic planning in regional development’, *Economic and Social Changes: Facts, Trends, Forecast*, 11(3), pp. 100–116.
17. Egorov, K.Yu. (2010) ‘Evaluating the efficiency of implementing the strategy of social and economic development of the region’, *Vestnik Samarskogo gosudarstvennogo ekonomicheskogo universiteta*, 72(10), pp. 10–16. (In Russian).
18. Fomin, V.V., Flegontov, A.V. and Osochkin, A.A. (2017) ‘Method of frequency-morphological classification of texts’, *Programmnye produkty i sistemy*, 30(3), pp. 478–486. (In Russian).
19. Gataullin, R.R. (2019) *Methods, models and software tools for resolving polysemy in texts*. PhD Thesis. Kazan. (Privolzh.) feder. un-t. Kazan’. (In Russian).
20. Grishman, R. (1997) ‘Information extraction: Techniques and challenges’, *Information extraction a multidisciplinary approach to an emerging information technology*, 1299, pp. 10–27.
21. Hahn, U. and Mani, I. (2000) ‘The challenges of automatic summarization’, *IEEE Computer*, 33(11), pp. 29–36.
22. Hurtado, J.L., Agarwal, A. and Zhu, X. (2016) ‘Topic discovery and future trend forecasting for texts’, *Journal of Big Data*, 3(7). Available at: <https://link.springer.com/article/10.1186/s40537-016-0039-2> (accessed 20 October 2021).
23. Il’ina, I.N., Kirillova, A.N., Plisetskii, E.E., Kopychenko, G.S. and Rybina, E.G. (2014) ‘Evaluation of the effectiveness of regional target programs implementation: The Tyumen region case study’, *Regional’naya ekonomika: teoriya i praktika*, 46, pp. 37–55. (In Russian).
24. Il’ina, I.N., Plisetskii, E.E., Kopychenko, G.S. and Rybina, E.G. (2016) ‘Quality assessment of regional strategies for development in Russia’, *Regional’naya ekonomika: teoriya i praktika*, 427(4), pp. 178–196. (In Russian).
25. Jeon, J. and Choe, J.-W. (2009) ‘A key word analysis of English intensifying adverbs in male and female speech’, *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, 1, pp. 210–219.
26. Kalabikhina, I.E. (2019) ‘Modern socio-demographic policy in Russia: Is there any continuity in conceptual approaches in the documents of 2007–2017?’, *Zhenshchina v rossiyskom obshchestve*, 4, pp. 14–28. (In Russian).
27. Kim, M.-J., Ohk, K. and Moon, Ch.-S. (2017) ‘Trend analysis by using text mining of journal articles regarding consumer policy’, *New Physics: Sae Mulli*, 67(5), pp. 555–561.
28. Kislicyna, V.V., Cheglakova, L.S., Karaulov, V.M. and Chikisheva, A.N. (2017) ‘Formation of the integrated approach to the assessment of the socio-economic development of regions’, *Ekonomika regiona*, 13(2), pp. 369–380. (In Russian).
29. Klimanov, V.V., Budaeva, K.V. and Chernyshova, N.A. (2017) ‘Preliminary results of strategic planning in Russian regions’, *Ekonomicheskaya politika*, 12(5), pp. 104–127. (In Russian).

30. Komarov, V.M., Akimova, V.V., Kotsyubinskiy, V.A. and Zemtsov, S.P. (2021) 'Comparative analysis of the development approaches to long-term government strategies in Russia and in the world', *Public Administration Issues*, 1, pp. 56–75. (In Russian).
31. Koteeswaran, S., Visu, P. and Janet, J. (2012) 'A review on clustering and outlier analysis techniques in Data Mining', *American Journal of Applied Sciences*, 9, pp. 254–258.
32. Kruengkrai, C. and Jaruskulchai, C. (2003) 'Generic text summarization using local and global properties of sentences', *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence*. Washington, DC: USA.
33. Kurushin, D., Dubrovskaya, J., Rusinova, M. and Kozonogova, E. (2019) 'Automation and algorithmization of smart benchmarking of territories based on data parsing', *Digital Science 2019: Conf. on Digital Science (DSIC 2019)*, pp. 126–134.
34. Kutsenko, E., Islankina, E. and Kindras, A. (2018) 'Smart by oneself? An analysis of Russian regional innovation strategies within the RIS3 framework', *Foresight and STI Governance*, 12(1), pp. 25–45.
35. Kuzyk, M., Simachev, Yu. and Vlasenko, A. (2019) 'Strategies for the development of Russian industries and sectors: Is there room for structural change?', *XX April International Academic Conference on Economic and Social Development*. Available at: https://www.researchgate.net/publication/333783636_Strategies_for_the_development_of_Russian_industries_and_sectors_is_there_room_for_structural_change_in_Russian (accessed 20 October 2021). (In Russian).
36. Lipina, S.A. and Smirnova, O.O. (2017) 'Strategic planning in subjects of the Russian Federation', *Regional'naya ekonomika. Yug Rossii*, 15(1), pp. 25–35. (In Russian).
37. Litvinova, T., Litvinova, O. and Zavarzina, G. (2019) 'Two views on the 2010 Moscow metro bombings: Corpus-based contrastive keyword analysis', *4th International Conference DTGS: International Conference on Digital Transformation and Global Society*, pp. 393–404.
38. Manning, D. and Schutze, H. (2000) *Foundations of statistical natural language processing*. Cambridge: MIT Press.
39. Margolin, A.M. (2018) 'Why government program evaluation methodologies require improvement', *Ekonomicheskaya politika*, 6, pp. 54–81. (In Russian).
40. Miner, G., Elder, J. IV and Hill, T. (2012) *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
41. Mintzberg, G., Alstrand, B. and Lampel, J. (2015) *Strategic safari. Excursion through the wilds of electoral management*. Moscow: Alpina Publisher. (In Russian).
42. Morozova, N.V. and Bondarenko, N.V. (2014) 'Approaches to assessing the effectiveness of the regional strategy socio-economic development', *Vestnik Chuvashskogo universiteta*, 3, pp. 178–182. (In Russian).

43. Nekrest'yanov, I.G. (2000) Topic-oriented information retrieval methods. St. Petersburg. (In Russian).
44. Nikolaev, M.A. and Mahotaeva, M.Yu. (2016) 'The analysis of development strategy of the region's industrial complex', *Nauchno-texnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politexnicheskogo universiteta. E'konomicheskie nauki*, 246(4), pp. 29–41. (In Russian).
45. Piotrovskaya, K.R. (2014) 'A Survey of Text mining', *Izvestiya Rossiyskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena*, 168, pp. 128–134. (In Russian).
46. Raghavan, P. and Tsaparas, P. (2002) 'Mining significant associations in large scale text corpora', *Proceedings of IEEE International Conference on Data Mining*, pp. 402–409.
47. Repichev, A.I., Tugacheva, L.V., Vorob'eva, A.V. and Avdeeva, D.A. (2018) 'Opportunities for the development of regional innovation strategies based on the principles of «smart specialization»', *Voprosy upravleniya*, 32(2), pp. 37–45. (In Russian).
48. Risin, I.E. (2021) 'Methodological approach to quality assessment regional strategy', *Region: sistemy, ekonomika, upravlenie*, 55 (4), pp. 68–73. DOI: 10.22394/1997-4469-2021-55-4-68-73. (in Russian).
49. Rusinova, M.R. (2020) 'Algorithmization of the procedure for carrying out “smart” benchmarking of regions of the Russian Federation', *Vestnik Permskogo natsional'nogo issledovatel'skogo politehnicheskogo universiteta. Sotsial'no-ekonomicheskie nauki*, 2, pp. 286–292. (In Russian).
50. Salton, G. (1989) *Automatic text processing the analysis and retrieval of information by computer*. Addison Wesley.
51. Salton, G. (1997) 'Automatic text structuring and summarization', *Information Processing & Management*, 33(2), pp. 193–207.
52. Seliverstov, V.E. and Mel'nikova, L.V. (2011) 'Analyzing the regional strategic planning in the Siberian federal district', *Region: E'konomika i Sociologiya*, 3, pp. 3–21. (In Russian).
53. Shabanov, V.I. and Andreev, A.M. (2003) 'A method for classifying text documents based on full-text search', *Trudy ROMIP'2003yu. SPb: NII Khimii SPbGU*, pp. 52–71. (In Russian).
54. Shobha, S.R. and Thakore, D. (2010) 'Text mining using keyphrase extraction', *Indian Journal of Computer Science and Engineering*, 1(2), pp. 82–85.
55. Shpakova, R.N. (2019) 'Strategies of socio-economic development of regions: from assessment of achievement of goals to assessment of effectiveness', *Vestnik Evraziiskoi nauki*, 11(2), pp. 59 (in Russian).
56. Shulekina, E.N. (2014) 'To the issue of regional investment strategies compliance with the requirements of investment Standards', *Sibirskaya finansovaya shkola*, 106(5), pp. 154–159. (In Russian).

57. Sukhovey, A.F. and Golova, I.M. (2016) 'Substantiation of the transformation of the priorities of innovation and technological development of Russian regions in the global crisis', *Ekonomika regiona*, 12(3), pp. 911–923. (In Russian).
58. Tien, D.D., Siu, Ch.H. and Alvis, C.M. (2006) 'Fong associative feature selection for Text Mining', *International Journal of Information Technology*, 12(4), pp. 59–68.
59. Türegün, N. (2019) 'Text Mining in financial information', *Current Analysis on Economics & Finance*, 1, pp. 18–26.
60. Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C. and Nevill-Manning, C.G. (1999) 'KEA: Practical Automatic Keyphrase Extraction', *ACM DL*, 8, pp. 254–255.
61. Zhiharevich, B.S. and Krylovskiy, A.B. (2011) 'Experience in comparative assessment of the quality of strategic planning in the regions of the North Caucasus', *E'konomika razvitiya regiona: problemy, poiski, perspektivy*, 12, pp. 217–228. (In Russian).
62. Zhiharevich, B.S. and Lebedeva, N.A. (2015) 'Translyatsiya idej transformatsii regional'nogo prostranstva v dokumenty strategicheskogo planirovaniya', *E'konomika Severo-Zapada: problemy i perspektivy razvitiya*, 4(49), pp. 58–74.

Статья поступила в редакцию 27.10.2021;
одобрена после рецензирования 12.12.2022;
принята к публикации 27.05.2022.