

Т.В. МИТРОФАНОВА, А.В. ХРИСТОФОРОВА

ТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГРАНТОВЫХ ПРОЕКТОВ КАК ИНСТРУМЕНТ УПРАВЛЕНИЯ СОЦИАЛЬНЫМ РАЗВИТИЕМ РЕГИОНА

Ключевые слова: социальные системы, анализ грантовых проектов, Чувашская Республика, тематическое моделирование, LDA, BERTopic, управление развитием, обработка естественного языка (NLP).

Управление социальным развитием региона в условиях цифровизации требует эффективных аналитических инструментов и точных рейтинговых систем. В качестве одного из таких инструментов может выступать тематическое моделирование, которое позволяет проводить автоматическое обобщение и структуризацию больших текстовых массивов. Статья посвящена применению методов тематического моделирования для анализа грантовых проектов в Чувашской Республике. Актуальность исследования обусловлена необходимостью анализа тематик грантовых проектов для определения ключевых направлений общественно значимой деятельности и их соответствия региональным стратегическим целям.

Цель исследования – выявление ключевых тематических направлений поддержанных грантовых проектов с использованием методов тематического моделирования (LDA и BERTopic) для повышения эффективности управления социальным развитием региона.

Материалы и методы. Для анализа использовался датасет, содержащий описания проектов-победителей грантовых конкурсов в Чувашской Республике за 2023–2025 годы. Тексты прошли предобработку, включая очистку от стоп-слов, нормализацию и лемматизацию с использованием библиотеки Stanza. Методы LDA и BERTopic были применены для выявления тематических кластеров, а визуализация результатов выполнена с помощью библиотеки pyLDAvis.

Результаты. В результате анализа выявлено пять основных тематических направлений: поддержка и социальная помощь уязвимым группам населения, спортивные и масовые мероприятия, развитие и здоровье детей, в том числе с ограниченными возможностями здоровья, обучение и развитие молодежи, психологическая и социальная поддержка участников программ. Выявлены ключевые направления развития региона, что позволяет повысить эффективность стратегического планирования грантовой поддержки будущих заявок. Исследование подтвердило, что темы имеют высокую степень соответствия целям Стратегии развития Чувашской Республики до 2035 г. и федеральным национальным проектам. Результаты проведенного исследования могут быть использованы грантодателями для изменения и адаптации конкурсных условий и критерии оценки, а также для введения новых актуальных запросов направлений. Кроме того, организации могут применять результаты исследования для подготовки более релевантных заявок, отвечающих актуальным социальным потребностям региона.

Выводы. Выявлено пять ключевых направлений грантовой поддержки в Чувашской Республике, что предоставляет органам управления объективную и структурированную основу для корректировки региональной политики и оценки результативности механизмов ресурсного распределения. Исследование подтвердило высокую эффективность применения методов тематического моделирования для анализа корпуса текстовых описаний грантовых проектов. Выявленные авторами с помощью методов LDA и BERTopic тематические направления поддержки грантов в Чувашской Республике за 2023–2025 гг. соответствуют целям, заявленным в стратегических региональных документах. Результаты работы могут служить основой для совершенствования конкурсных механизмов грантовой поддержки в области управления социальным развитием региона.

Введение. Существующий разрыв между номинальными целями региональной стратегии и фактическим распределением ресурсов на микроуровне создает системную методологическую проблему в управленческом цикле. Традиционные подходы к мониторингу, базирующиеся на агрегированных статистических показателях, не позволяют осуществить содержательный анализ качественных параметров развития, выраженных в текстовых массивах проектной документации. Возникает необходимость в инструментарии, способном обеспечить верифицируемое соотнесение дискретных инициатив некоммерческого сектора со стратегическими ориентирами территориального развития. Отсутствие формализованных процедур для выявления латентных тематических структур в корпусе поддерживаемых проектов ограничивает возможность корректной оценки результативности таких механизмов региональной политики, как грантовая поддержка.

Разрешение указанной методологической проблемы лежит в плоскости применения алгоритмов обработки естественного языка к анализу проектных заявок. Получаемая тематическая модель позволяет не только верифицировать степень соответствия реализуемых проектов стратегическим приоритетам, но и идентифицировать латентные диспропорции в распределении ресурсов между различными направлениями развития.

Таким образом, разработка и апробация специализированного методического инструментария для контент-анализа документов регионального развития являются актуальными с точки зрения решения фундаментальной задачи региональной экономики – создания релевантных механизмов обратной связи между стратегическим планированием и практикой ресурсного распределения на субнациональном уровне. Полученная тематическая карта позволит органам власти не только констатировать текущие приоритеты некоммерческого сектора, но и выявлять стратегически важные для экономики региона направления, оказавшиеся на периферии грантовой активности, и целенаправленно стимулировать подачу заявок по данным темам через адаптацию конкурсных условий.

Современные методы анализа текстов, основанные на алгоритмах обработки естественного языка (Natural Language Processing, NLP), находят все большее применение в различных прикладных задачах, включая анализ содержательных направлений грантовой и проектной деятельности. Одним из эффективных инструментов является тематическое моделирование, позволяющее выявлять скрытые смысловые структуры в корпусах текстов. Это особенно актуально в контексте анализа грантовых заявок, где тексты проектов отражают приоритеты и направления общественно значимой деятельности.

Методы тематического моделирования находят широкое применение в социальных науках, предоставляя мощные инструменты для анализа больших текстовых коллекций и извлечения значимой информации [6]. Латентное размещение Дирихле, являющееся одним из наиболее распространенных методов тематического моделирования, направлено на выявление скрытых тем в наборе документов на основе анализируемых слов [7, 8]. Этот подход основывается на принципе, согласно которому каждый документ представляет собой

смесь тем, а каждая тема, в свою очередь, является распределением слов [9]. Как неконтролируемый метод, LDA автоматически выявляет и маркирует темы в больших текстовых наборах, не требуя предварительной аннотации данных [10]. Тематическое моделирование ориентировано на анализ текстов с целью выявления значений терминов в их контексте на естественном языке [11]. Таким образом, скрытое распределение Дирихле позволяет моделировать документ как комбинацию скрытых тем.

Более поздняя разработка BERTopic использует возможности языковых моделей на основе transformer, таких как BERT, для создания плотных вложений документов, а затем применяет методы кластеризации для определения тем. BERTopic – это подход к моделированию тем, который использует языковые модели на основе трансформаторов для получения плотных вкраплений документов, которые впоследствии кластеризуются для поиска тем. Этот подход, по сравнению с традиционными методами, часто приводит к получению более связных и интерпретируемых тем.

В сфере исследований психического здоровья методы тематического моделирования используются для выявления преобладающих тем в материалах о психическом здоровье [12]. Например, исследователи использовали LDA для анализа онлайн-форумов, посвященных психическому здоровью, выявляя такие темы, как тревога, депрессия и суицидальные мысли. Применение тематического моделирования распространяется и на другие области, включая политологию, медиаисследования и социологию, предлагая ценное понимание сложных социальных явлений [1].

Тематическое моделирование может выявлять скрытые закономерности и доминирующие темы в текстовых данных, связанных с предложениями о грантах и отчетами по проектам, предоставляя информацию, которая может быть неочевидна при традиционном статистическом анализе или проверке вручную [13]. Такой подход позволяет проводить всестороннюю оценку деятельности по предоставлению грантов на основе данных, способствуя более обоснованному и стратегическому распределению ресурсов в будущем.

Оценка социально-экономического развития такого региона, как Чувашская Республика, требует детального подхода, учитывающего различные взаимосвязанные факторы [14]. Анализ грантовых проектов дает ценное представление о приоритетах развития и стратегиях распределения ресурсов в регионе. Система распределения грантов ведущим грантовым оператором России – Фондом президентских грантов анализируется в работах [2-4].

Способность эффективно интерпретировать и использовать огромные объемы информации стала критически важным фактором, определяющим стратегическое преимущество и операционную эффективность, и в этом контексте тематическое моделирование и кластеризация становятся мощными аналитическими инструментами, способными превратить необработанные данные в действенные знания, способствующие принятию более обоснованных и эффективных управленческих решений.

Тематическое моделирование – набор алгоритмов, предназначенных для выявления тематической структуры, лежащей в основе коллекции документов, –

является мощным средством обобщения и понимания больших объемов текстовых данных [15].

Латентное распределение Дирихле [16], широко используемый метод моделирования тем, предполагает, что каждый документ представляет собой смесь нескольких тем и что каждая тема характеризуется распределением по словам. Этот подход позволяет менеджерам автоматически определять ключевые темы и подтемы, присутствующие в их данных, предлагая ценные сведения о предпочтениях клиентов, рыночных тенденциях и конкурентной среде.

Цель исследования – выявление ключевых содержательных направлений поддержанных проектов на основе автоматизированного анализа текстов с использованием инструментов NLP и тематического моделирования.

Материалы и методы. Объектом исследования выступает корпус текстовых описаний проектов-победителей грантовых конкурсов для некоммерческих организаций, реализуемых на территории Чувашской Республики в период 2023–2025 гг. Эмпирическую базу исследования составили данные, сформированные на основе информации, размещенной в открытом доступе на официальном портале Фонда президентских грантов¹. На основе этих данных был подготовлен датасет общим объемом 35 поддержанных проектов, включающий информацию о названии проекта, его кратком описании, где проведения, размере поддержки и тематике (табл. 1).

Таблица 1
Основные поля датасета

Поле	Описание
Название проекта	Полное название поддержанного проекта
Организация	Организация, получившая финансирование
Сумма гранта	Размер предоставленного гранта в рублях
Год	Год проведения конкурса (2023, 2024, 2025)
Тематика	Тематическое направление проекта
Краткое описание	Краткое описание содержания проекта

Для анализа текстов использовались методы обработки естественного языка, включая очистку текстов от стоп-слов и пунктуации, нормализацию (приведение к нижнему регистру), векторизацию с помощью CountVectorizer.

Для лингвистической нормализации текстов использовалась лемматизация с применением библиотеки Stanza, разработанная Стэнфордским университетом. Выбор был обусловлен тем, что, в отличие от простых словарных подходов ruymorphy2, Mystem, Stanza использует обученные нейросетевые модели, демонстрирующие более высокую точность в разрешении морфологических амбиграмм, особенно в контексте социальных и проектных текстов.

Тематическое моделирование проводилось на основе комбинированного подхода, сочетающего классический метод LDA и современный нейросетевой подход BERTopic. Такой синтез методов позволил максимально полно выявить

¹ Фонд президентских грантов: сайт. URL: <https://президентскиегранты.рф> (дата обращения: 16.05.2025).

тематическую структуру описаний грантовых проектов. LDA был выбран благодаря своей прозрачности, хорошей интерпретируемости результатов и удобству визуализации с использованием pyLDAvis. В то же время BERTopic, основанный на трансформерных моделях, обеспечивает более глубокий семантический анализ и лучше учитывает контекстные взаимосвязи между словами.

Результаты исследования. В результате тематического моделирования было выявлено пять основных тем, каждая из которых характеризуется набором ключевых слов, отражающих содержание соответствующей группы проектов (табл. 2).

Таблица 2
Выявленные темы в результате использования метода LDA

Тема	Ключевые слова	Интерпретация
Поддержка и социальная помощь	Помощь, ребенок, центр, республика, инвалид, регион	Тема связана с социальной поддержкой уязвимых групп населения, особенно детей с ограниченными возможностями здоровья (ОВЗ) или инвалидностью. Упоминаются такие аспекты, как помощь на уровне региона (Чувашия), участие центров реабилитации и комплексного сопровождения. Особое внимание уделяется значимости региональных программ и межведомственных коалиций
Спортивные и массовые мероприятия	Спортивный, мероприятие, чувашский, школьник, проведение, участие	Содержание темы составляют спортивные и образовательные мероприятия, направленные на привлечение школьников, студентов и других групп населения. Основное внимание уделяется проведению мероприятий в рамках республиканских программ, таких как спортивные соревнования, форумы или тренинги. Выявлена связь с популяризацией здорового образа жизни и вовлечением молодежи
Дети, развитие и здоровье	Ребенок, занятие, спорт, развитие, родитель, ОВЗ, здоровье	Тема посвящена развитию детей, особенно тех, кто имеет ограничения по здоровью (ОВЗ). Основное внимание уделяется занятиям, направленным на улучшение физического и психического здоровья, а также на поддержку родителей. Обсуждаются адаптированные программы для детей с расстройствами аутистического спектра (PAC) и другие формы реабилитации
Молодежь и обучение	Молодой, проведение, развитие, обучение, ребенок, помощь, взрослый	Тема связана с обучением и развитием молодых людей, включая подростков и молодых взрослых. Обсуждаются вопросы профессионального роста, подготовки к самостоятельной жизни и предоставления возможностей для обучения, а также затрагиваются аспекты помощи молодым людям с ОВЗ
Психологическая и социальная поддержка	Человек, участник, поддержка, жизнь, психологический, физический, социальный	Тема затрагивает аспекты психологической и социальной поддержки участников различных программ. Основное внимание уделяется улучшению качества жизни через физические и психологические практики. Упоминаются групповые занятия, консультации и другие формы работы с участниками проектов

Визуализация результатов тематического моделирования с использованием библиотеки pyLDAvis позволила более детально интерпретировать структуру полученных тем и их взаимосвязи (рис. 1). На карте межтематических расстояний

(Intertopic Distance Map) каждая тема представлена в виде круга, площадь которого пропорциональна доле соответствующей темы в корпусе текстов.

Было выявлено пять основных тем, различающихся по объему и степени взаимосвязи. Темы 1 и 3 частично пересекаются, что свидетельствует о наличии общих лексических маркеров, отражающих родственные направления в описаниях проектов. Тема 2 располагается в непосредственной близости к этим кластерам, демонстрируя тематическую сопряженность при сохранении собственной семантической автономии. Тема 4 находится ближе к теме 1, что может свидетельствовать о пересечении отдельных аспектов содержания. Тема 5 выделяется наиболее отчетливо, занимая изолированное положение на карте, что указывает на ее семантическую уникальность и тематическую специфичность (рис. 1).

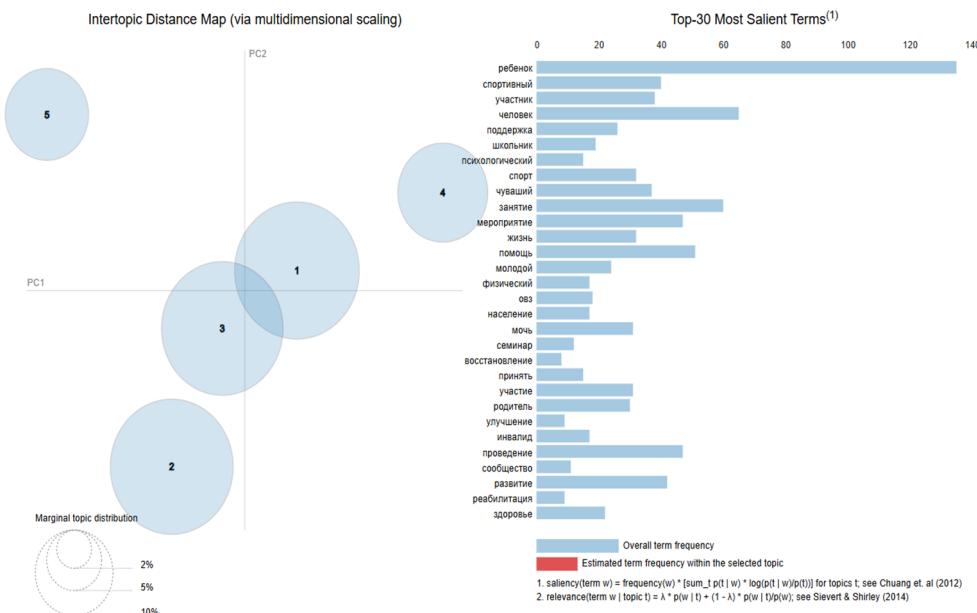


Рис. 1. Визуализация результатов тематического моделирования LDA:
 1 – поддержка и социальная помощь; 2 – спортивные и массовые мероприятия;
 3 – дети, развитие и здоровье; 4 – молодежь и обучение;
 5 – психологическая и социальная поддержка

На правой панели (рис. 1) визуализированы 30 наиболее значимых (salient) терминов, определенных с учетом их частоты и различимости между темами. К числу ключевых слов, характеризующих корпус в целом, относятся: «ребенок», «человек», «спорт», «здравье», «поддержка», «участник», «ОВЗ», «психологический», «семинар», «реабилитация», «восстановление», что подтверждает ориентированность проектов преимущественно на социальную, оздоровительную и инклюзивную повестку.

Для углубленного анализа тематической структуры был дополнительно применен метод BERTopic, основанный на трансформерных языковых моделях (рис. 2). Темы, выделенные в результате моделирования BERTopic, обладают интерпретируемостью и логической связностью. Повторяющееся слово «проект» отражает специфику корпуса, ориентированного на описание инициатив.

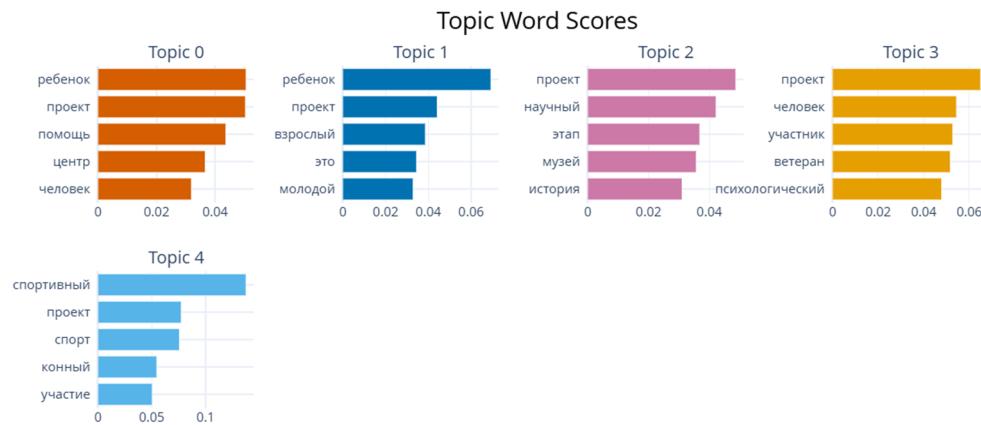


Рис. 2. Визуализация результатов тематического моделирования BERTopic

Сравнение результатов, полученных с помощью LDA и BERTopic, показало как совпадения, так и различия в структуре тематических кластеров (табл. 3).

Таблица 3
Сопоставление результатов тематического моделирования
методами LDA и BERTopic

Тематика	Ключевые слова LDA	Ключевые слова BERTopic
Социальная помощь и поддержка	Помощь, ребенок, центр, инвалид, регион	Ребенок, проект, помощь, центр, человек
Спортивные и массовые мероприятия	Спортивный, мероприятие, школьник, участие	Спортивный, спорт, проект, участие
Развитие и здоровье детей	Ребенок, занятие, спорт, развитие, здоровье	Ребенок, здоровье, развитие
Молодежь и обучение	Молодой, развитие, обучение, помощь	Проект, взрослый, молодой
Психологическая и социальная поддержка	Человек, участник, поддержка, жизнь, психологический	Участник, ветеран, психологический, человек

Таким образом, тематическое моделирование позволило структурировать содержательное пространство грантовых проектов и выделить доминирующие направления, что дает основание говорить о приоритетах региональной поддержки в социальной сфере.

Для повышения надежности результатов тематического моделирования текстов описаний социальных проектов были применены два метода – BERTopic и LDA. Оба алгоритма позволили выявить релевантные тематические кластеры, однако в силу различий в подходах к построению тем наблюдаются как совпадения, так и расхождения.

Совпадающие темы включают социальную помощь и поддержку. В BERTopic это тема с ключевыми словами «ребенок», «помощь», «проект», «центр», а в LDA – тема «Поддержка и социальная помощь» с ключевыми словами «ребенок», «инвалид», «центр», «регион». Интерпретация идентична – оказание социальной поддержки уязвимым категориям, в первую очередь

детям с ОВЗ. Еще одно совпадение касается темы спорта и мероприятий. В BERTopic присутствует тема с ключевыми словами «спортивный», «спорт», «проект», а в LDA – «спортивные и массовые мероприятия». Оба подхода фиксируют активное участие детей и молодежи в спортивной и просветительской активности. Кроме того, при применении моделей BERTopic и LDA была зафиксирована конвергенция результатов, выражаясь в выделении темы, связанной с детским развитием, физическим и психологическим здоровьем. Анализ показывает терминологическое перекрытие, включающее лексемы «ребенок», «здравье», «развитие», «занятие», «ОВЗ».

Наблюдаемые расхождения заключаются в следующем: с помощью BERTopic удалось достичь более отчетливого тематического разделения между кластерами «ветераны и участники» и «научные проекты». В то же время при использовании LDA основной акцент смещен в сторону вопросов обучения молодежи и психологической поддержки. Тематические кластеры, полученные по результатам LDA, интерпретируются как более обобщенные категории, такие как «молодежь и обучение» и «психологическая поддержка», тогда как на основе BERTopic были сформированы более сжатые, но конкретные лексические ядра.

Проведенный анализ позволил установить, что ключевые темы описаний социальных проектов коррелируют с целевыми ориентирами Стратегии развития Чувашской Республики до 2035 г. (далее – Стратегия) [5]. Выявленная проблематика – развитие человеческого капитала, укрепление социальной сплоченности и создание условий для активного долголетия – непосредственно соответствует формулировкам цели 4 «Развитие человеческого капитала и социальной сферы в Чувашской Республике. Повышение уровня и качества жизни населения» указанного документа (табл. 4).

Таблица 4

Социальные задачи и тематика проектов

Задачи социального развития Чувашии	Проектные акценты
Задача 4.1. Демографическое развитие, улучшение здоровья населения и поддержание его долголетней активной жизни	Акцент на оздоровительных, профилактических, спортивных проектах и проектах реабилитации
Задача 4.3. Создание конкурентоспособного образования, кадровое обеспечение реального сектора экономики и приоритетные направления работы с молодежью	Проекты, связанные с обучением, занятиями для школьников, молодежными инициативами и дополнительным образованием
Задача 4.5. Развитие социальной защиты населения	Тематика поддержки инвалидов, семей с детьми, пожилых людей, оказание помощи в трудной жизненной ситуации
Задача 4.7. Развитие культуры, укрепление единства российской нации и этнокультурное развитие народов Чувашской Республики	Темы патриотического воспитания, региональной идентичности и сохранения культурного наследия
Задача 4.9. Развитие физической культуры и спорта	Проекты, направленные на расширение доступа населения к спорту, особенно в детской и школьной среде

Проведенный тематический анализ позволяет утверждать, что тематика поддержанных проектов соответствует стратегическим ориентирам региона в социальной политике.

Результаты исследования свидетельствуют также и об ограниченности тематического спектра грантовых проектов в аспекте отражения в них полного перечня стратегических целей развития Чувашской Республики. Так, например, некоторые направления, закрепленные в Стратегии, практически не нашли отражения в грантовой активности.

В корпусе проектов отсутствуют инициативы, напрямую связанные с первой целью Стратегии, касающейся роста конкурентоспособности экономики, а именно: развитие высокотехнологичных производств и ИТ-кластера и создание инфраструктуры для инноваций. Кроме того, не представлены проекты, направленные на достижение цели совершенствования институциональной среды (привлечение инвестиций, улучшение предпринимательского климата), а также на реализацию пятой цели Стратегии, касающейся сбалансированного пространственного развития территорий. Тематический анализ не выявил в массиве поддержанных проектов, ориентированных на развитие туристической отрасли. Выявление стратегических направлений, не представленных в тематике грантов, демонстрирует, что тематическое моделирование является не просто инструментом классификации, а аналитическим ресурсом для диагностики дисбалансов в распределении ресурсов. Полученные результаты указывают на то, что механизм грантовой поддержки в его текущем виде не в полной мере задействован для стимулирования проектов, которые напрямую ориентированы на диверсификацию и рост конкурентоспособности региональной экономики.

Приоритетные темы поддержанных проектов грантов также согласуются с целями таких национальных проектов, как «Демография» (развитие здравоохранения, поддержка семей с детьми), «Образование» (дополнительное образование, развитие профессиональных навыков молодежи), «Культура» (сохранение историко-культурного наследия, развитие культурной идентичности), «Спорт – норма жизни» (популяризация массового спорта среди детей и молодежи). Такое соответствие усиливает стратегическую значимость выявленных направлений проектной деятельности и подтверждает эффективность региональных инициатив в контексте общенациональных задач.

Полученные результаты тематического моделирования могут быть использованы для динамической адаптации грантовых программ под изменяющиеся социальные потребности региона. Регулярный анализ тематики проектов позволит выявлять новые приоритеты и тренды в гражданских инициативах, оперативно реагировать на социальные вызовы. Это даст возможность органам власти и операторам грантов обновлять условия конкурсов, вводить специальные грантовые направления и корректировать критерии оценки заявок в соответствии с актуальной социальной повесткой.

Выводы. Выявлены пять основных тематических направлений, которые отражают ключевые приоритеты региональной поддержки: это социальная помощь, развитие и поддержка детей и молодежи, вопросы здоровья и инклюзии, а также просветительская и образовательная деятельность. Данная типологизация предоставляет органам управления объективную и структурированную основу для корректировки региональной политики и оценки результативности механизмов ресурсного распределения. Установлено, что тематики проектов

соответствуют целям Стратегии социально-экономического развития Чувашской Республики до 2035 г. Это свидетельствует о целенаправленном и эффективном использовании грантовых механизмов для решения стратегических задач региона.

Проведенное исследование подтвердило высокую эффективность применения комбинации методов тематического моделирования, таких как LDA и BERTopic, для анализа содержания грантовых проектов, реализованных в Чувашской Республике в 2023–2025 гг. Метод LDA обеспечил ясную интерпретацию тем, а метод BERTopic раскрыл дополнительные семантические аспекты грантовых тематик.

Результаты исследования обладают практической значимостью как для грантодавающих организаций, осуществляющих экспертизу проектов, так и для некоммерческих структур, планирующих участие в конкурсах. На основе выявленных тематических дисбалансов могут быть разработаны целевые меры управленческого воздействия, например, установление специальных номинаций, повышающих коэффициентов для заявок, соответствующих «непокрытым» стратегическим задачам экономического развития, а также проведение информационных кампаний для привлечения организаций из соответствующих секторов. Это позволит перейти от пассивной модели грантовой поддержки, адаптирующейся к сложившимся запросам, к активной модели, выступающей в роли катализатора структурных преобразований в экономике региона в соответствии со стратегическими приоритетами.

Кроме того, работа формирует методологическую основу для дальнейших научных изысканий. В будущем планируется расширение исследования за счет увеличения объема данных, применения дополнительных методов обработки естественного языка (например, анализа тональности) и изучения взаимосвязей между тематикой проектов и их социально-экономическими результатами.

Литература

1. Методы анализа письменной коммуникации скулштеров для прогнозирования предупреждающего поведения / А.О. Савельев, А.Ю. Карпова, Д.А. Третьяков, Н.Г. Максимова // Управление развитием крупномасштабных систем (MLSD'2024): труды Семнадцатой междунар. конф., 24–26 сентября 2024 г. / под общ. ред. С.Н. Васильева, А.Д. Цвиркуна; Ин-т проблем упр. им. В.А. Трапезникова Рос. акад. наук. М.: ИПУ РАН, 2024. С. 1399–1405.
2. Смирнов В.А. Эффективность грантовой поддержки российских социально-ориентированных некоммерческих организаций (на примере Фонда президентских грантов) // Социологические исследования. 2020. № 9. С. 79–89.
3. Старшинова А.В., Бородкина О.И. Стратегии устойчивости социально ориентированных НКО: механизм грантовой поддержки // Экономические и социальные перемены: факты, тенденции, прогноз. 2022. № 5. С. 221–236.
4. Тарасова А.Н., Ижик А.В. Грантовая поддержка проектов социально ориентированных некоммерческих организаций Свердловской области // Теория и практика общественного развития. 2021. № 11. С. 78–85.
5. Стратегия развития Чувашской Республики до 2035 года [Электронный ресурс] // Министерство экономического развития и имущественных отношений Чувашской Республики: офиц. сайт. URL: <https://minec.cap.ru/action/activity/soc-econom-razvitiie/strategiya-socialjno-ekonomicheskogo-razvitiya-chu/strategiya-socialjno-ekonomicheskogo-razvitiya-chu> (дата обращения: 16.02.2025).
6. Egger R., Yu J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 2022, vol. 7, 886498. DOI: 10.3389/fsoc.2022.886498.

7. Ghanem R., Erbay H. Context-dependent model for spam detection on social networks. *SN Applied Sciences*, 2020, vol. 2, 1587. DOI: 10.1007/s42452-020-03374-x.
8. Lancichinetti A., Sicer M.I., Wang J.X. et al. A high-reproducibility and high-accuracy method for automated topic classification. *ArXiv*. 2014. arXiv: 1402.0422. Available at: <https://arxiv.org/pdf/1402.0422.pdf>.
9. Sangaraju V.R., Bolla B.K., Nayak D.K., Kh J. Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings. *ArXiv*. 2022. arXiv:2205.07259. Available at: <https://arxiv.org/abs/2205.07259>.
10. Wang L., Miloslavsky E., Stone J.H. et al. Topic modeling to characterize the natural history of ANCA-Associated vasculitis from clinical notes: A proof of concept study. *Seminars in Arthritis and Rheumatism*, 2021, vol. 51, no. 1, pp. 150–157.
11. Bellaouar S., Bellaouar M.M., Ghada I.E. Topic Modeling: Comparison of LSA and LDA on Scientific Publications. In: Proc. of the 4th International Conference on Data Storage and Data Engineering (DSDE'21). New York, Association for Computing Machinery, 2021, pp. 59–64. DOI: 10.1145/3456146.3456156.
12. Gao X., Sazara C. Discovering Mental Health Research Topics with Topic Modeling. *ArXiv*. 2023. arXiv:2308.13569. Available at: <https://arxiv.org/abs/2308.13569>.
13. Tohalino J.A.V., Quispe L.V.C., Amancio D.R. Analyzing the relationship between text features and grants productivity. *Scientometrics*, 2021, vol. 126, pp. 4255–4275.
14. Gorbunova P., Denisov G., Perfilova E., Zakharova A. Assessment of the socio-economic development of the region (on the example of the Chuvash Republic). *SHS Web of Conferences*, 2021, vol. 110, 01027. DOI: 10.1051/shsconf/202111001027.
15. Knorr K., Hein-Pensel F. Since Albert and Whetten: the dissemination of Albert and Whetten's conceptualization of organizational identity. *Management Review Quarterly*, 2024, vol. 74, pp. 597–625.
16. Oprea S.-V., Bâra A. Is Artificial Intelligence a Game-Changer in Steering E-Business into the Future? Uncovering Latent Topics with Probabilistic Generative Models. *Journal of Theoretical and Applied Electronic Commerce Research*, 2025, vol. 20, no. 1, 16. DOI: 10.3390/jtaer20010016.

МИТРОФАНОВА ТАТЬЯНА ВАЛЕРЬЕВНА – кандидат физико-математических наук, доцент кафедры математического и аппаратного обеспечения информационных систем, Чувашский государственный университет, Россия, Чебоксары (mitrofanova_tv@mail.ru; ORCID: <https://orcid.org/0000-0002-5750-7991>).

ХРИСТОФОРОВА АНАСТАСИЯ ВЛАДИМИРОВНА – кандидат физико-математических наук, доцент кафедры математического и аппаратного обеспечения информационных систем, Чувашский государственный университет, Россия, Чебоксары (kristofav@mail.ru; ORCID: <https://orcid.org/0000-0003-3534-8747>).

Tatiana V. MITROFANOVA, Anastasiia V. KHRISTOFOROVA

THEMATIC MODELING OF GRANT PROJECTS AS A TOOL FOR MANAGING THE SOCIAL DEVELOPMENT OF A REGION

Key words: social systems, grant project analysis, the Chuvash Republic, thematic modeling, LDA, BERTopic, development management, natural language processing (NLP).

Managing the social development of a region in the context of digitalization requires effective analytical tools and accurate rating systems. One of these tools can be thematic modeling, which enables to carry out automatic generalization and structuring of large text arrays. The article is devoted to applying thematic modeling methods for the analysis of grant projects in the Chuvash Republic. The relevance of the research is due to the need to analyze the topics of grant projects in order to identify key areas of socially significant activities and their compliance with regional strategic goals.

The purpose of the study is to identify the key thematic areas of supported grant projects using thematic modeling methods (LDA and BERTopic) to improve the effectiveness of managing the social development of the region.

Materials and methods. For the analysis, a dataset containing descriptions of the winning projects of grant competitions in the Chuvash Republic for 2023–2025 was used. The texts were pre-processed, including purification from stop words, normalization and lemmatization using the Stanza library. The LDA and BERTopic methods were applied to identify thematic clusters, and visualization of the results was performed using the pyLDAvis library.

Results. As a result of the analysis, five main thematic areas were identified: support and social assistance to vulnerable groups of the population, sports and mass events, development and health of children, including those with disabilities, education and youth development, psychological and social support for program participants. The key directions of the region's development were identified, which makes it possible to increase the effectiveness of strategic planning for grant support for future applications. The study confirmed that the topics have a high degree of compliance with the goals of the Development Strategy of the Chuvash Republic until 2035 and federal national projects. The results of the conducted research can be used by grant-givers to change and adapt the competitive conditions and evaluation criteria, as well as to introduce new relevant research areas. In addition, organizations can apply the results of the study to prepare more relevant applications that meet current social needs of the region.

Conclusions. Five key areas of grant support in the Chuvash Republic were identified, which provides the governing bodies with an objective and structured basis for adjusting regional policy and evaluating the effectiveness of resource allocation mechanisms. The study confirmed high efficiency of using thematic modeling methods to analyze the corpus of text descriptions of grant projects. The thematic areas of grant support in the Chuvash Republic in 2023–2025 identified by the authors using LDA and BERTopic methods correspond to the goals stated in the strategic regional documents. The results of the work can serve as a basis for improving competitive grant support mechanisms in the field of managing the social development of the region.

References

1. Savel'ev A.O., Karpova A.Yu., Tret'yakov D.A., Maksimova N.G. *Metody analiza pis'mennoi kommunikatsii skulshuterov dlya prognozirovaniya preduprezhdayushchego povedeniya* [Methods for analyzing written communication of school shooters to predict warning behavior]. In: Vasil'eva S.N., Tsvirkuna A.D., ed. *Upravlenie razvitiem krupnomasshtabnykh sistem (MLSD'2024): trudy Semnadtsatoi mezhdunarodnoi konferentsii* [Proc. of 17th Int. Sci. Conf. "Management of Large-Scale Systems Development (MLAD'2024)"]. Moscow, 2024, pp. 1399–1405.
2. Smirnov V.A. *Efektivnost' grantovoi podderzhki rossiiskikh sotsial'no-orientirovannykh nekommerseskikh organizatsii (na primere Fonda prezidentskikh grantov)* [The Effectiveness of Grant Support for Russian Socially-Oriented Non-Profit Organizations (Based on the Example of the Presidential Grants Fund)]. *Sotsiologicheskie issledovaniya*, 2020, no. 9, pp. 79–89.
3. Starshinova A.V., Borodkina O.I. *Strategii ustoichivosti sotsial'no orientirovannykh NKO: mehanizm grantovoi podderzhki* [Sustainability Strategies for Socially Oriented Non-Profit Organizations: A Grant Support Mechanism]. *Ekonomicheskie i sotsial'nye peremeny: fakty, tendentsii, prognoz*, 2022, no. 5, pp. 221–236.
4. Tarasova A.N., Izhik A.V. *Grantovaya podderzhka proektorov sotsial'no orientirovannykh nekommerseskikh organizatsii Sverdlovskoi oblasti* [Grant support for projects of socially oriented non-profit organizations in the Sverdlovsk region]. *Teoriya i praktika obshchestvennogo razvitiya*, 2021, no. 11, pp. 78–85.
5. *Strategiya razvitiya Chuvashskoi Respubliki do 2035 goda* [Development Strategy of the Chuvash Republic until 2035]. Available at: [https://minec.cap.ru/action/activity/soc-econom-razvitiye-strategiya-socialjno-ekonomicheskogo-razvitiya-chu](https://minec.cap.ru/action/activity/soc-econom-razvitiye-strategiya-socialjno-ekonomicheskogo-razvitiya-chu/strategiya-socialjno-ekonomicheskogo-razvitiya-chu) (Access Date: 2025, Feb. 16).
6. Egger R., Yu J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 2022, vol. 7, 886498. DOI: 10.3389/fsoc.2022.886498.
7. Ghanem R., Erbay H. Context-dependent model for spam detection on social networks. *SN Applied Sciences*, 2020, vol. 2, 1587. DOI: 10.1007/s42452-020-03374-x.
8. Lancichinetti A., Sicer M.I., Wang J.X. et al. A high-reproducibility and high-accuracy method for automated topic classification. *ArXiv*. 2014. arXiv: 1402.0422. Available at: <https://arxiv.org/pdf/1402.0422.pdf>.

9. Sangaraju V.R., Bolla B.K., Nayak D.K., Kh J. Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings. ArXiv. 2022. arXiv:2205.07259. Available at: <https://arxiv.org/abs/2205.07259>.
10. Wang L., Miloslavsky E., Stone J. H. et al. Topic modeling to characterize the natural history of ANCA-Associated vasculitis from clinical notes: A proof of concept study. *Seminars in Arthritis and Rheumatism*, 2021, vol. 51, no. 1, pp. 150–157.
11. Bellaouar S., Bellaouar M.M., Ghada I.E. Topic Modeling: Comparison of LSA and LDA on Scientific Publications. In: Proc. of the 4th International Conference on Data Storage and Data Engineering (DSDE'21). New York, Association for Computing Machinery, 2021, pp. 59–64. DOI: 10.1145/3456146.3456156.
12. Gao X., Sazara C. Discovering Mental Health Research Topics with Topic Modeling. ArXiv. 2023. arXiv:2308.13569. Available at: <https://arxiv.org/abs/2308.13569>.
13. Tohalino J.A.V., Quispe L.V.C., Amancio D.R. Analyzing the relationship between text features and grants productivity. *Scientometrics*, 2021, vol. 126, pp. 4255–4275.
14. Gorbunova P., Denisov G., Perfilova E., Zakharova A. Assessment of the socio-economic development of the region (on the example of the Chuvash Republic). *SHS Web of Conferences*, 2021, vol. 110, 01027. DOI: 10.1051/shsconf/202111001027.
15. Knorr K., Hein-Pensel F. Since Albert and Whetten: the dissemination of Albert and Whetten's conceptualization of organizational identity. *Management Review Quarterly*, 2024, vol. 74, pp. 597–625.
16. Oprea S.-V., Bâra A. Is Artificial Intelligence a Game-Changer in Steering E-Business into the Future? Uncovering Latent Topics with Probabilistic Generative Models. *Journal of Theoretical and Applied Electronic Commerce Research*, 2025, vol. 20, no. 1, 16. DOI: 10.3390/jtaer20010016.

TATIANA V. MITROFANOVA – Candidate of Physics and Mathematics Sciences, Associate Professor, Department of Mathematical and Hardware Support for Information Systems, Chuvash State University, Russia, Cheboksary (mitrofanova_tv@mail.ru; ORCID: <https://orcid.org/0000-0002-5750-7991>).

ANASTASIIA V. KHRISTOFOROVA – Candidate of Physics and Mathematics Sciences, Associate Professor, Department of Mathematical and Hardware Support for Information Systems, Chuvash State University, Russia, Cheboksary (khristofav@mail.ru; ORCID: <https://orcid.org/0000-0003-3534-8747>).

Формат цитирования: Митрофанова Т.В., Христофорова А.В. Тематическое моделирование грантовых проектов как инструмент управления социальным развитием региона [Электронный ресурс] // Oeconomia et Jus. 2025. № 4. С. 25–37. URL: <http://oeconomia-et-jus.ru/single/2025/4/3>. DOI: 10.47026/2499-9636-2025-4-25-37.