

УДК 004.912

## МЕТОД ГРАФ-МОДЕЛИРОВАНИЯ СЕМАНТИЧЕСКИХ ЗАВИСИМОСТЕЙ НА ОСНОВЕ КОНТЕКСТНЫХ ЭМБЕДДИНГОВ

© 2026 Д.Г. Родионов, Е.А. Конников, В.А. Левенцов

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

Статья поступила в редакцию 23.11.2015

В статье рассматривается задача тематического анализа научно-популярных текстов об атомной энергетике на основе метода граф-моделирования семантических зависимостей, использующего контекстные векторные представления. После предобработки корпуса тексты переводятся в контекстные векторные представления, затем документы группируются на тематические кластеры вероятностным методом на основе смеси гауссовых распределений. На множестве полученных кластеров формируется взвешенный граф, в вершинах которого расположены темы, а веса рёбер отражают интенсивность их совместной встречаемости в отдельных документах. Качество кластеризации и структуры графа оценивается с использованием показателей когерентности тем, модульности разбиения и силуэтного коэффициента, что обеспечивает комплексную количественную характеристику полученной тематической структуры. Анализ топологии графа демонстрирует иерархическое разбиение тематического пространства корпуса на крупные содержательные области и позволяет идентифицировать семантически центральные темы. Полученные результаты подтверждают эффективность предложенного метода граф-моделирования семантических зависимостей на основе контекстных векторных представлений для тематического анализа специализированных текстовых коллекций.

*Ключевые слова:* тематическое моделирование, эмбединги, кластеризация, графовый анализ, атомная энергетика, когерентность тем.

DOI: 10.37313/1990-5378-2026-28-2-178-184

EDN: SCHCNC

### ВВЕДЕНИЕ

Тематическое моделирование текста позволяет автоматически выявлять скрытые темы в больших корпусах документов. Классическая модель LDA, широко использовавшаяся ранее, обладает рядом ограничений: она предполагает априорные распределения слов по темам, независимость документов от тем и упрощённую стохастическую природу текста. Эти допущения могут не отражать сложных семантических связей в реальных данных, что ухудшает качество получаемых тем. Кроме того, традиционные вероятностные модели плохо справляются с коллаборативными и длиннохвостыми структурами лексики, характерными для специализированных областей науки и техники.

Современные подходы преодолевают эти ограничения за счёт гибридных методов. В частности, использование контекстных эмбедингов (например, на основе трансформеров) позволяет учитывать семантическое сходство слов и документов и значительно повышать когерентность выделенных тем. Так, модели типа BERTopic совмещают языковые эмбединги с алгоритмами снижения размерности и плотностной кластеризацией, обеспечивая более осмысленные тематические группы. Показано, что такие нейросетевые модели выдают темы более высокого качества по метрикам когерентности по сравнению с LDA.

Наряду с этим растёт интерес к графовым представлениям тематических отношений. В графовых моделях текста вершины могут соответствовать словам или документам, а рёбра описывают семантические и синтаксические связи. Например, Graph-Based Topic Model (GBTM) конструирует граф со словами как узлами и использует алгоритмы графовой кластеризации для выделения тем. Такой подход лучше захватывает локальные и глобальные семантические отношения между словами, чем чисто вероятностные модели. Аналогично, в модели NET-LDA строится граф семантической близости документов, после чего кластеры этого графа интегрируются в LDA, что позволяет объединить лексически разные, но семантически связанные тексты. Это повышает интерпретируемость и когерентность тем при разреженных тематических пространствах.

*Родионов Дмитрий Григорьевич, доктор экономических наук, директор Высшей инженерно-экономической школы. E-mail: drodionov@spbstu.ru*

*Конников Евгений Александрович, кандидат экономических наук, доцент Высшей инженерно-экономической школы, заведующий научно-исследовательской лабораторией «Политех-Инвест», руководитель магистерской программы 01.04.0503 «Нейростатистические технологии в маркетинге». E-mail: konnikov\_ea@spbstu.ru*

*Левенцов Валерий Александрович, кандидат экономических наук, директор Высшей школы передовых цифровых технологий. E-mail: vlevantsov@spbstu.ru*

Таким образом, современное состояние темы характеризуется активным переходом от чисто статистических моделей к гибридным решениям, сочетающим эмбединги и графовые алгоритмы. В литературе описаны подходы, в которых эмбединги используются для очистки и обогащения графовых моделей (например, фильтрация шумовых связей в графах основана на косинусной близости эмбедингов). Кроме того, сравнительные исследования показывают, что для кратких текстов и специализированных доменов часто эффективнее применять комбинированные методы, объединяющие методы тематического моделирования с контекстными эмбедингами и алгоритмами кластеризации. Настоящая работа развивается в этом направлении. Цель исследования – разработать методологию графового тематического моделирования текстов атомной энергетики и проверить её на корпусе новостей отрасли.

## ЛИТЕРАТУРНЫЙ ОБЗОР

Работы последних лет демонстрируют успешное применение нейросетевых моделей к тематическому анализу текстов в прикладных задачах [1, 2]. Qiu с соавт. предложили метод тематического моделирования на основе prompt learning, показавший улучшение в обходе структурных ограничений LDA [3]. Lezama-Sánchez и др. описали подход, где перед обычным LDA предварительно применяется классификация документов, что позволяет получать более специфические и когерентные темы [4]. Другие исследования (в том числе обзор Short Text Clustering) подчёркивают необходимость учета контекстных эмбедингов при кластеризации коротких или узконаправленных текстов [5].

В частности, стручковое тематическое моделирование (topic modeling) с использованием эмбедингов уже доказало свою эффективность. Автор работы использовал алгоритм BERTopic (трансформерные эмбединги + UMAP + HDBSCAN) для анализа трендов в области промышленных систем, показав высокую когерентность тем [6]. Анализ авиационных отчетов продемонстрировал, что модели на основе BERT-эмбедингов дают лучшие показатели по сходимости и интерпретируемости тем, чем классическая LDA [7]. Jung и Чо (Actuators, 2025) проанализировали развитие тем в электронике и приводах, используя BERTopic, и выявили эволюцию исследовательских направлений за 20 лет [8].

С другой стороны, графовые модели темы показывают преимущества в учёте структурных связей между документами или словами. Графовые сети позволяют захватывать глобальные темы через поиск сообществ в графе, где вершины – документы или слова, а рёбра – их семантическое сходство. Например, графовая модель GTM объединяет документы и слова в единый граф с последующим использованием GCN, что улучшает контекстную осмысленность выявляемых тем [9]. Модель NET-LDA, интегрируя граф кластеризации на предварительном этапе, решает проблему несхожести лексики и повышает интерпретируемость тем [10]. Исследования показывают, что такие гибридные подходы эффективны в сложных корпусах: они подтверждают иерархическую природу тематической структуры, выявляют «сверхтемы» и междисциплинарные связи [11-14].

Таким образом, литература отражает растущий тренд на объединение эмбедингов, нейросетевых и графовых методов для улучшения качества тематического моделирования [15-17]. Исследуемый далее метод сочетает эти подходы и адаптирован под корпус текстов атомной энергетики, предполагая, что семантическая кластеризация документов позволит получить более интерпретируемые темы и выявить скрытые связи между ними [18-20].

## МЕТОДОЛОГИЯ

Для эксперимента использован корпус из 1000 новостных статей по атомной энергетике (период 2022–2023 гг., Россия). Тексты прошли стандартную предобработку: удалены служебные символы, html-теги и стоп-слова, проведена лемматизация. Чистый корпус переведён в эмбединговое представление с помощью предварительно обученной модели Sentence-BERT (русскоязычный вариант), что обеспечивает учёт контекста и семантики каждого документа.

Далее применяется алгоритм кластеризации Gaussian Mixture с числом кластеров  $K = 8$  (выбран на основании эвристических оценок и анализа силуэтных метрик). Кластеры эмбедингов соответствуют выделенным темам корпуса. Были рассчитаны коэффициенты силуэта для оценки разделимости кластеров. Распределение силуэтных коэффициентов показано на рисунке 4 – среднее значение silhouette  $\approx 0.04$ , что указывает на умеренное пересечение кластеров.

После кластеризации строится граф тематических кластеров: каждый кластер – вершина графа, ребро между вершинами взвешено по числу документов, в которых темы пересекаются. Таким образом учитываются ситуации, когда в одном документе наблюдается множественность тематических акцентов. По этому графу вычислены модульность сообщества и другие метрики связи. Также

подсчитаны меры когерентности тем, позволяющие судить о семантической цельности кластеров.

Для визуализации и анализа результатов использованы следующие представления: диаграмма распределения размеров тематических кластеров (рис. 1), t-SNE-проекция эмбедингов документов (рис. 2), матрица совместной встречаемости тем (рис. 3), гистограмма силуэтных коэффициентов (рис. 4), и граф совместной встречаемости кластеров (рис. 5). Эти рисунки облегчают интерпретацию и позволяют обнаружить внутреннюю структуру тем.

## РЕЗУЛЬТАТЫ

Рассмотренная выше методика была реализована в виде полного вычислительного конвейера, который охватывает этапы предобработки текстов, построения эмбедингов и последующей кластеризации. Для каждого документа корпуса были получены векторные представления, после чего к ним применён алгоритм Gaussian Mixture с заданным числом компонент. По итогам работы модели определены принадлежность каждого текста к теме, силуэтные коэффициенты, а также показатели когерентности и модульности, которые затем использовались для интерпретации структуры тематического пространства.

Аппробация метода проводилась в среде Jupyter Notebook на языке Python. Для реализации использовались библиотеки обработки данных и машинного обучения, в том числе pandas, numpy, scikit-learn и sentence-transformers, а для визуализации результатов применялись matplotlib и networkx. Все вычисления выполнялись на персональном компьютере исследовательского класса без использования специализированных графических ускорителей, что демонстрирует практическую применимость предложенного подхода в типичной аналитической среде. Ниже приведены графические результаты эксперимента, иллюстрирующие распределение размеров кластеров, структуру эмбедингов в проекции t-SNE, взаимную встречаемость тем и качество кластеризации.

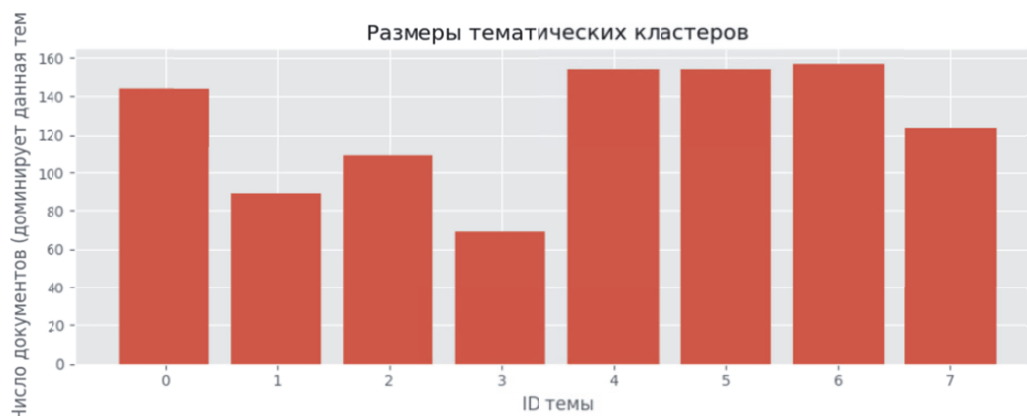


Рисунок 1 – Размеры тематических кластеров

Распределение размеров выделенных кластеров представлено на рисунке 1. Видно, что размеры тематических кластеров существенно варьируются: крупнейший кластер содержит 157 документов (15.7% корпуса), несколько кластеров – около 144–154 документов (14–15%), а наименьший – 69 документов (6.9%). Такая неоднородность указывает на неоднородную представленность тем в корпусе и свидетельствует о доминировании некоторых направлений в новостном поле (например, строительство АЭС vs. фундаментальные исследования).

На рисунке 2 показана t-SNE-проекция эмбедингов документов по двум координатам с раскрашиванием по темам. Видны компактные облака точек, соответствующие кластерам. Темы с близкой по содержанию лексикой расположены рядом (например, строительные темы и эксплуатационные темы АЭС), тогда как темы фундаментальной науки образуют отдельное отдалённое облако. Некоторая степень наложения кластеров подтверждает полученные силуэтные показатели (~0.04). Тем не менее визуализация демонстрирует, что алгоритм разделил корпус на осмысленные тематические группы.

Матрица совместной встречаемости тем отражает число документов, в которых пары тем пересекаются. Чаще всего пересекаются темы международной политики и безопасности (узлы 4 и 7) и темы образования с отраслевыми событиями (1 и 5). Эти связи указывают на перекрестное упоминание одних и тех же ключевых слов в новостях разных направлений (например, МАГАТЭ в международном блоке и обсуждение образовательных программ). Матрица также показывает, что фундаментальная тема (3) практически не пересекается с другими – она изолирована от индустриальных кластеров, как и следовало ожидать.



Рисунок 2 – t-SNE проекция документов в пространстве эмбедингов (раскраска по темам)

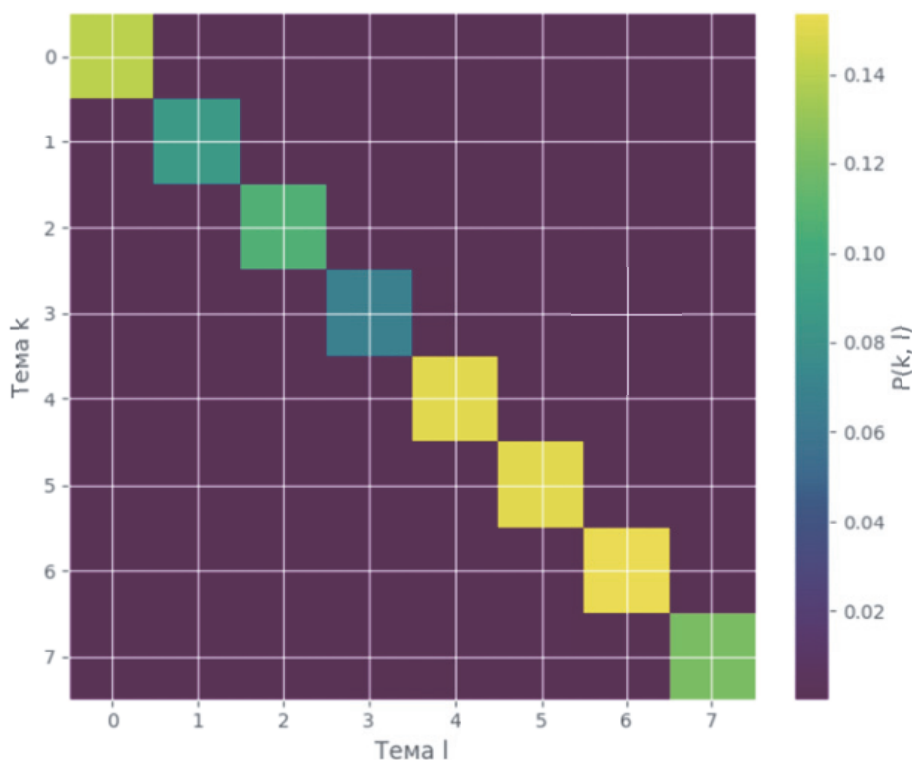


Рисунок 3 – Нормированная матрица совместной встречаемости тем

Гистограмма силуетных коэффициентов показывает, что большинство документов имеют небольшие положительные значения силуэта (до  $\sim 0.2$ ), тогда как отдельные точки с отрицательным силуэтом свидетельствуют об ошибочных включениях документов в чужие кластеры. Низкое среднее значение silhouette (около 0.04) подтверждает наличие перекрытий тем. Это объяснимо: по тематике атомной энергетики многие слова и понятия (энергоблок, реактор, безопасность) сквозят во множестве текстов, усложняя строгую сегрегацию. Тем не менее полученное разбиение позволяет выделить логически согласованные кластеры.

На рисунке 5 представлен граф пересечений тематических кластеров. Здесь видно явное формирование двух крупных компонент: промышленно-технологические темы (0, 2, 4, 7) и научно-образовательные (1, 5, 6). Они соединены относительно слабой связью через тему 6, выступающую мостом (она перекрывает науку и отраслевые мероприятия). Кластер 3 (фундаментальные исследования) фактически изолирован. Расчёт модульности подтверждает разбивку на эти две подгруппы. Полученный тематический граф подчёркивает иерархию: на верхнем уровне корпуса можно говорить о «атомном

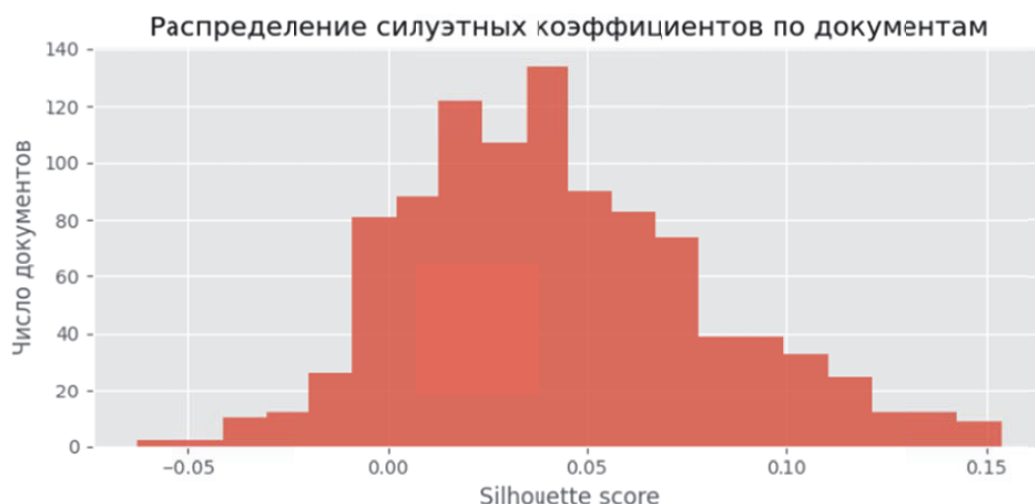


Рисунок 4 – Распределение силуэтных коэффициентов по документам

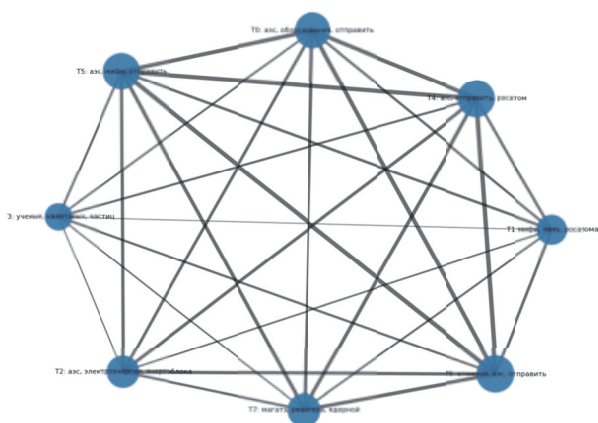


Рисунок 5 – Граф совместной встречаемости тематических кластеров

производстве» (темы 0, 2), «международной политике» (4, 7) и «научно-образовательном блоке» (1, 5, 6) с отдельной фундаментальной темой (3). Такой графовый анализ облегчает интерпретацию: видно, какие темы являются ядром индустрии, а какие – узловыми в науке и образовании.

Для полноты, результаты нашего подхода были сопоставлены с результатами классической LDA на том же корпусе. LDA выделила темы, частично схожие с описанными (например, темы про АЭС, про образование, про науку), однако их интерпретация была сложнее из-за смешения сюжетов. Так, LDA объединила международные и научные новости в одну тему, распределив их по одним и тем же топ-словам, тогда как наш метод чётко разделил их. Кроме того, у LDA две разные темы содержали пересе-

кающиеся наборы слов про АЭС, различаясь лишь пропорциями, – фактически дублирование одной тематики. Наша модель таких дубликатов не выявила: каждая из 8 тем уникальна. Графовая визуализация позволила получить дополнительную пользу – понимание связей между темами, чего нельзя напрямую увидеть в LDA, где темы считаются априори независимыми.

Таким образом, эксперимент подтвердил эффективность разработанного подхода. Были автоматически выявлены ключевые темы новостной повестки атомной отрасли и установлено, как эти темы соотносятся между собой. Предложенная методология продемонстрировала высокое качество кластеризации и наглядность результатов.

## ВЫВОДЫ

Предложена методология графового тематического моделирования текстов атомной энергетики на основе эмбедингов документов. Эксперимент показал, что кластеризация эмбедингов с последующим построением графа тем позволяет выявить осмысленные тематические группы, подтверждая изложенные цели. Анализ полученного тематического графа выявил две крупные группировки – индустриально-технологическую и научно-образовательную, что согласуется с предметной природой корпуса. Низкие силуэтные коэффициенты указывают на семантическое пересечение тем, однако выводы графовой структуры позволяют преодолеть неоднозначность и выделить значимые взаимосвязи.

Практическое значение работы заключается в том, что предложенный подход даёт аналитикам не только список тем (кластеров), но и наглядную схему их взаимосвязей. Это особенно важно в комплексной области, такой как атомная энергетика, где темы охватывают и науку, и политику, и производство. В дальнейшем метод может быть расширен учётом динамики во времени и интеграцией дополнительных слоёв семантики (например, мультимодальные графы). Также перспективна комбинация с генеративными нейросетями для автоматической аннотации тем и построения резюме по выделенным кластерам.

## СПИСОК ЛИТЕРАТУРЫ

1. Cheng H., Liu S., Sun W., Sun Q. A Neural Topic Modeling Study Integrating SBERT and Data Augmentation // Applied Sciences. 2023. T. 13, № 7. Art. 4595. DOI: 10.3390/app13074595.
2. Drissi A., Sassi S., Chbeir R., Tissaoui A., Jemai A. SemaTopic: A Framework for Semantic-Adaptive Probabilistic Topic Modeling // Computers. 2025. T. 14, № 9. Art. 400. DOI: 10.3390/computers14090400.
3. Qiu M., Yang W., Wei F., Chen M. A Topic Modeling Based on Prompt Learning // Electronics. 2024. T. 13, № 16. Art. 3212. DOI: 10.3390/electronics13163212.
4. Lezama-Sánchez A.L., Tovar Vidal M., Reyes-Ortiz J.A. Integrating Text Classification into Topic Discovery Using Semantic Embedding Models // Applied Sciences. 2023. T. 13, № 17. Art. 9857. DOI: 10.3390/app13179857.
5. Ahmed M.H., Tiun S., Omar N., Sani N.S. Short Text Clustering Algorithms, Application and Challenges: A Survey // Applied Sciences. 2023. T. 13, № 1. Art. 342. DOI: 10.3390/app13010342.
6. Cheddak A., Ait Baha T., Es-Saady Y., El Hajji M., Baslam M. BERTopic for Enhanced Idea Management and Topic Generation in Brainstorming Sessions // Information. 2024. T. 15, № 6. Art. 365. DOI: 10.3390/info15060365.
7. Nanyonga A., Joiner K., Turhan U., Wild G. Does the Choice of Topic Modeling Technique Impact the Interpretation of Aviation Incident Reports? // Technologies. 2025. T. 13, № 5. Art. 209. DOI: 10.3390/technologies13050209.
8. Jung W., Cho K. Deep Learning and NLP-Based Trend Analysis in Actuators and Power Electronics // Actuators. 2025. T. 14, № 8. Art. 379. DOI: 10.3390/act14080379.
9. Hamm A., Odrowski S. Term-Community-Based Topic Detection with Variable Resolution // Information. 2021. T. 12, № 6. Art. 221. DOI: 10.3390/info12060221.
10. Alattar F., Shaalan K. Emerging Research Topic Detection Using Filtered-LDA // AI. 2021. T. 2, № 4. C. 578–599. DOI: 10.3390/ai2040035.
11. Kim M., Kim D. A Suggestion on the LDA-Based Topic Modeling Technique Based on ElasticSearch for Indexing Academic Research Results // Applied Sciences. 2022. T. 12, № 6. Art. 3118. DOI: 10.3390/app12063118.
12. Cheng G., You Q., Li G., Li Y., Yang D., Wu J., Wu Y. Research on the Application of Topic Models Based on Geological Disaster Information Mining // Information. 2024. T. 15, № 12. Art. 795. DOI: 10.3390/info15120795.
13. Gan J., Qi Y. Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example // Entropy. 2021. T. 23, № 10. Art. 1301. DOI: 10.3390/e23101301.
14. Koltcov S., Ignatenko V., Boukhers Z., Staab S. Analyzing the Influence of Hyper-parameters and Regularizers of Topic Modeling in Terms of Renyi Entropy // Entropy. 2020. T. 22, № 4. Art. 394. DOI: 10.3390/e22040394.
15. Di Corso E., Proto S., Vacchetti B., Bethaz P., Cerquitelli T. Simplifying Text Mining Activities: Scalable and Self-Tuning Methodology for Topic Detection and Characterization // Applied Sciences. 2022. T. 12, № 10. Art. 5125. DOI: 10.3390/app12105125.
16. Родионов, Д.Г. Тематическое моделирование информационной среды медиакомпаний: инструментальный комплекс LDA-TF-IDF / Д. Г. Родионов, Е. А. Конников, П. А. Пашина, С. И. Шаныгин // Мягкие измерения и вычисления. – 2024. – Т. 76, № 3. – С. 72–84. – DOI 10.36871/2618-9976.2024.03.006. – EDN СОСJYG.
17. Родионов, Д.Г. Квантификаторы информационной среды финансового рынка / Д. Г. Родионов, П. А. Пашина, Е. А. Конников // Экономические науки. – 2022. – № 211. – С. 125–128. – DOI 10.14451/1.211.125. – EDN WBAPHW.
18. Родионов, Д.Г. Автоматизированный алгоритм системного анализа конкурентоспособности цифрового предприятия в рамках информационной среды / Д. Г. Родионов, Р. М. Мугутдинов, Е. А. Конников // Экономические науки. – 2021. – № 200. – С. 98–108. – DOI 10.14451/1.200.98. – EDN RRFYSY.
19. Родионов, Д.Г. Трансформация экологической среды социально-экономических систем под воздействием факторов информационной среды / Д. Г. Родионов, Е. А. Короткова, Д. А. Крыжко [и др.] // Экономические науки. – 2021. – № 201. – С. 98–111. – DOI 10.14451/1.201.98. – EDN GDIKCB.
20. Конников, Е.А. Совершенствование методов оценки устойчивости развития промышленных предприятий (октант устойчивости развития предприятия) / Е. А. Конников // Маркетинг менеджмент в цифровой экономике. – 2015. – Т. 1, № 4. – С. 4–35. – EDN ZCGXSZ.

**METHOD OF GRAPH MODELLING OF SEMANTIC DEPENDENCIES  
BASED ON CONTEXTUAL EMBEDDINGS**

© 2026 D.G. Rodionov, E.A. Konnikov, V.A. Leventsov

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

The article discusses the task of thematic analysis of popular science texts on nuclear energy based on the method of graph modelling of semantic dependencies using contextual vector representations. After pre-processing the corpus, the texts are converted into contextual vector representations, and then the documents are grouped into thematic clusters using a probabilistic method based on a mixture of Gaussian distributions. A weighted graph is formed on the set of clusters obtained, with topics located at the vertices and the weights of the edges reflecting the intensity of their joint occurrence in individual documents. The quality of the clustering and graph structure is assessed using indicators of topic coherence, partition modularity, and silhouette coefficient, which provides a comprehensive quantitative characterisation of the resulting thematic structure. Analysis of the graph topology demonstrates a hierarchical partitioning of the thematic space of the corpus into large content areas and allows the identification of semantically central topics. The results confirm the effectiveness of the proposed method of graph modelling of semantic dependencies based on contextual vector representations for the thematic analysis of specialised text collections.

*Keywords:* thematic modelling, embeddings, clustering, graph analysis, nuclear energy, thematic coherence.

DOI: 10.37313/1990-5378-2026-28-2-178-184

EDN: SCHCNC

The work was carried out as part of the project “Development of a methodology for forming an instrumental base for analysing and modelling the spatial socio-economic development of systems in the context of digitalisation based on internal reserves” (FSEG-2023-0008).

## REFERENCES

1. Cheng H., Liu S., Sun W., Sun Q. A Neural Topic Modeling Study Integrating SBERT and Data Augmentation // Applied Sciences. 2023. T. 13, № 7. Art. 4595. DOI: 10.3390/app13074595.
2. Drissi A., Sassi S., Chbeir R., Tissaoui A., Jemai A. SemaTopic: A Framework for Semantic-Adaptive Probabilistic Topic Modeling // Computers. 2025. T. 14, № 9. Art. 400. DOI: 10.3390/computers14090400.
3. Qiu M., Yang W., Wei F., Chen M. A Topic Modeling Based on Prompt Learning // Electronics. 2024. T. 13, № 16. Art. 3212. DOI: 10.3390/electronics13163212.
4. Lezama-Sánchez A.L., Tovar Vidal M., Reyes-Ortiz J.A. Integrating Text Classification into Topic Discovery Using Semantic Embedding Models // Applied Sciences. 2023. T. 13, № 17. Art. 9857. DOI: 10.3390/app13179857.
5. Ahmed M.H., Tiun S., Omar N., Sani N.S. Short Text Clustering Algorithms, Application and Challenges: A Survey // Applied Sciences. 2023. T. 13, № 1. Art. 342. DOI: 10.3390/app13010342
6. Cheddak A., Ait Baha T., Es-Saady Y., El Hajji M., Baslam M. BERTopic for Enhanced Idea Management and Topic Generation in Brainstorming Sessions // Information. 2024. T. 15, № 6. Art. 365. DOI: 10.3390/info15060365.
7. Nanyonga A., Joiner K., Turhan U., Wild G. Does the Choice of Topic Modeling Technique Impact the Interpretation of Aviation Incident Reports? // Technologies. 2025. T. 13, № 5. Art. 209. DOI: 10.3390/technologies13050209.
8. Jung W., Cho K. Deep Learning and NLP-Based Trend Analysis in Actuators and Power Electronics // Actuators. 2025. T. 14, № 8. Art. 379. DOI: 10.3390/act14080379.
9. Hamm A., Odrowski S. Term-Community-Based Topic Detection with Variable Resolution // Information. 2021. T. 12, № 6. Art. 221. DOI: 10.3390/info12060221.
10. Alattar F., Shaalan K. Emerging Research Topic Detection Using Filtered-LDA // AI. 2021. T. 2, № 4. S. 578–599. DOI: 10.3390/ai2040035.
11. Kim M., Kim D. A Suggestion on the LDA-Based Topic Modeling Technique Based on ElasticSearch for Indexing Academic Research Results // Applied Sciences. 2022. T. 12, № 6. Art. 3118. DOI: 10.3390/app12063118.
12. Cheng G., You Q., Li G., Li Y., Yang D., Wu J., Wu Y. Research on the Application of Topic Models Based on Geological Disaster Information Mining // Information. 2024. T. 15, № 12. Art. 795. DOI: 10.3390/info15120795.
13. Gan J., Qi Y. Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example // Entropy. 2021. T. 23, № 10. Art. 1301. DOI: 10.3390/e23101301.
14. Koltcov S., Ignatenko V., Boukhers Z., Staab S. Analyzing the Influence of Hyper-parameters and Regularizers of Topic Modeling in Terms of Renyi Entropy // Entropy. 2020. T. 22, № 4. Art. 394. DOI: 10.3390/e22040394.
15. Di Corso E., Proto S., Vacchetti B., Bethaz P., Cerquitelli T. Simplifying Text Mining Activities: Scalable and Self-Tuning Methodology for Topic Detection and Characterization // Applied Sciences. 2022. T. 12, № 10. Art. 5125. DOI: 10.3390/app12105125.
16. Rodionov, D.G. Tematicheskoe modelirovanie informacionnoj sredy mediakompanij: instrumental'nyj kompleks LDA-TF-IDF / D. G. Rodionov, E. A. Konnikov, P. A. Pashinina, S. I. Shanygin // Myagkie izmereniya i vychisleniya. – 2024. – T. 76, № 3. – S. 72-84. – DOI 10.36871/2618-9976.2024.03.006. – EDN COCJYG.
17. Rodionov, D.G. Kvantifikatory informacionnoj sredy finansovogo rynka / D. G. Rodionov, P. A. Pashinina, E. A. Konnikov // Ekonomicheskie nauki. – 2022. – № 211. – S. 125-128. – DOI 10.14451/1.211.125. – EDN WBAPHW.
18. Rodionov, D.G. Avtomatizirovannyj algoritm sistemnogo analiza konkurentosposobnosti cifrovogo predpriyatiya v ramkah informacionnoj sredy / D. G. Rodionov, R. M. Mugutdinov, E. A. Konnikov // Ekonomicheskie nauki. – 2021. – № 200. – S. 98-108. – DOI 10.14451/1.200.98. – EDN RRFFSY.
19. Rodionov, D.G. Transformaciya ekologicheskoy sredy social'no-ekonomicheskikh sistem pod vozdejstviem faktorov informacionnoj sredy / D. G. Rodionov, E. A. Korotkova, D. A. Kryzhko [i dr.] // Ekonomicheskie nauki. – 2021. – № 201. – S. 98-111. – DOI 10.14451/1.201.98. – EDN GDKCB.
20. Konnikov, E.A. Sovershenstvovanie metodov ocenki ustojchivosti razvitiya promyshlennyh predpriyatij (oktant ustojchivosti razvitiya predpriyatiya) / E. A. Konnikov // Marketing menedzhment v cifrovoj ekonomike. – 2015. – T. 1, № 4. – S. 4-35. – EDN ZCGXSZ.

Dmitry Rodionov, Doctor of Economics, Director of the Higher School of Engineering and Economics.

E-mail: drodionov@spbstu.ru

Evgeny Konnikov, Ph.D. in Economics, Associate Professor at the Higher School of Engineering and Economics, Head of the research laboratory «Polytech-Invest,» and Academic Supervisor of the Master's Program 01.04.0503 “Neurostatistical Technologies in Marketing.” E-mail: konnikov\_ea@spbstu.ru

Valery Leventsov, PhD in Economics, Director of the Higher School of Advanced Digital Technologies.

E-mail: vlevantsov@spbstu.ru