

АВТОМАТИЧЕСКИЙ ПОИСК КОНЦЕПТОВ КОГНИТИВНОЙ КАРТЫ В ОБЛАСТИ СТРАТЕГИЧЕСКОГО УПРАВЛЕНИЯ

А. В. Заграновская

Санкт-Петербургский государственный экономический университет,
Санкт-Петербург, Россия

Когнитивная карта дает целостное представление о сложной хозяйственной ситуации. В статье предлагается выявлять концепты когнитивной карты не на основе экспертных суждений, как это принято, а на основе методов тематического моделирования – активно развивающегося направления статистического анализа текстов. Это позволит повысить доверие к выводам и прогнозам, сделанным на основе методов когнитивного моделирования. Отдельно для каждого года (с 2016 по 2020 г.) построена модель сходства выявленных тем в области стратегического управления на основе ключевых публикаций из ScienceDirect за указанный период, что дало возможность найти ключевые темы по годам с использованием меры центральности на основе собственного вектора. Автором выявлены семантическая структура научных публикаций и ключевые темы, волновавшие научное сообщество в области стратегического управления за период с 2016 по 2020 г., а также показана их динамика, что служит важным шагом в направлении автоматического построения когнитивной карты, являющейся инструментом анализа и прогнозирования состояния сложных систем.

Ключевые слова: тематическое моделирование, модель LDA, модель сходства тем, центральность на основе собственного вектора.

AUTOMATED SEARCH FOR CONCEPTS OF COGNITIVE MAP IN THE FIELD OF STRATEGIC MANAGEMENT

Anna V. Zagranovskaia

Saint Petersburg State Economic University,
Saint Petersburg, Russia

The cognitive map provides an integral idea of complicated economic situation. The article proposes to identify cognitive map concepts not on the basis of experts' reports, as it is usually done, but on the basis of thematic modeling methods – a fast developing line in statistic text analysis. It will give an opportunity to increase confidence in forecasts made on the basis of cognitive modeling methods. For each year (from 2016 to 2020) a model of similarity of subjects found in the sphere of strategic management was plotted on the basis of key publications in ScienceDirect for this period, which allowed to find key topics by years through using measures of centrality based on own vector. The author identified semantic structures of academic publications and key topics that excited scientific community in the field of strategic management from 2016 to 2020 and their dynamics, which is considered an important step towards automated plotting of the cognitive map being a tool of analyzing and forecasting the condition of complicated systems.

Keywords: thematic modeling, LDA model, model of subject similarity, centrality based on own vector.

Введение

Современные условия хозяйствования характеризуются высокой степенью неопределенности и сложности, что существенно ограничивает воз-

можности применения количественных методов для анализа и прогнозирования экономической ситуации. Когнитивные карты позволяют целостно взглянуть на ситуацию за счет качественно-количест-

венного и пространственно-временного описания рассматриваемой ситуации. Поэтому они активно применяются в условиях неопределенности [8].

Впервые когнитивные карты были предложены в 1970-х гг. известным американским социологом и политологом Робертом Аксельродом для моделирования принятия решений в социальных и политических системах [2].

Когнитивные карты отображаются с помощью знаковых ориентированных графов, состоящих из вершин и дуг. Вершины соответствуют концептам (наиболее значимым событиям, факторам), дуги отражают причинно-следственные связи между концептами. Когнитивные карты строят на основе мнений экспертов, что вызывает к ним недоверие из-за возможной субъективности суждений отдельных специалистов, сомнений в соблюдении процедуры проведения экспертизы. Поэтому задача повышения объективности в выявлении ключевых факторов сложившейся хозяйственной ситуации является актуальной, а с развитием методов машинного обучения становится осуществимой.

Инструментарий исследования

В статье предлагается выполнить ряд шагов к автоматическому построению когнитивной карты, а именно выявить концепты на основе методов тематического моделирования и построить модель сходства выявленных тем в области стратегического управления на основе ключевых публикаций из ScienceDirect. Проследим их динамику за 5 лет (с 2016 по 2020 г.), с тем чтобы понять, менялись ли количество и содержание тем по годам. ScienceDirect – один из крупнейших в мире онлайн-сборников опубликованных научных исследований. При этом для каждого рассматриваемого периода отбирались статьи, которые отвечают следующим требованиям:

- Key words: strategic management;
- Years: 2020 (2016, 2017, 2018, 2019);
- Article type: Research articles;

– Subject areas: Business, Management and Accounting;

– Access type: Open access & Open archive.

Благодаря имеющимся возможностям системы для каждого года был скачан архив из первых 100 статей, упорядоченных по убыванию значимости.

Обработка и тематическое моделирование статей проводились в среде программирования python. При этом использовался Google Colab – бесплатный облачный сервис на основе Jupyter Notebook.

Построение и анализ тематических моделей проводились с использованием следующих основных модулей: gensim (для тематического моделирования), spacy (для лемматизации), nltk (для удаления стоп-слов), pyLDAvis (для визуализации тем и соответствующих им ключевых слов). Модель сходства выявленных тем строилась на основе модуля для построения графов networkx.

Методология исследования

Тематическое моделирование – активно развивающееся направление статистического анализа текстов. Вероятностная тематическая модель позволяет выявить тематику коллекции текстовых документов, описывая каждую тему дискретным распределением на множестве терминов, каждый документ – дискретным распределением на множестве тем. Тематические модели используются для информационного поиска, классификации, категоризации, аннотирования, сегментации текстов [1]. В нашем случае проводилась сегментация текстов статей с целью выявления наиболее обсуждаемых тем в области стратегического управления.

Латентное размещение Дирихле LDA является доминирующим подходом в вероятностном тематическом моделировании. Именно этим методом мы и воспользовались. LDA предполагает, что слова каждого документа возникают из смеси тем, где каждая тема является многочленом из слов фиксированного словаря. Те-

мы являются общими для всех документов в коллекции, но пропорции тем варьируются стохастически в разных документах, так как они случайным образом берутся из распределения Дирихле [3].

Модель LDA относится к методам обучения без учителя, т. е. для ее построения нужно задать количество выделяемых тем, но проблема состоит в том, что мы хотим выявить скрытые темы, поэтому их количества мы не знаем. Приходится обращаться к критериям качества моделей и методом перебора находить удовлетворяющее количество тем в документе.

Экспертные оценки хорошо коррелируют с такой мерой качества тематической модели, как когерентность, которая может быть вычислена полностью автоматически без участия человека. Тема называется когерентной, если термины, наиболее частые в данной теме, неслучайно часто совместно встречаются рядом в документах коллекции [6]. Средняя когерентность тем считается хорошей мерой интерпретируемости тематической модели [7].

Тематическая модель ключевых публикаций из ScienceDirect за 2020 г.

По данным за 2020 г. нами был построен набор тематических моделей с количеством тем от 2 до 12. Для каждой модели вычислен показатель когерентности. Результаты расчетов приведены на рис. 1.

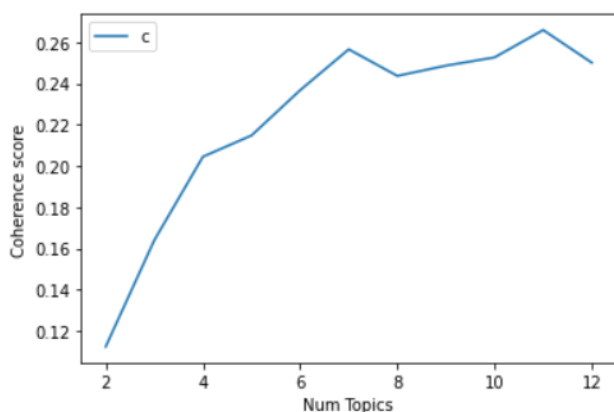


Рис. 1. Изменение когерентности в зависимости от числа тем в тематической модели за 2020 г.

Значение когерентности в зависимости от числа тем в тематической модели за 2020 г. представлено на рис. 2.

Num Topics = 2	has Coherence Value of 0.1123
Num Topics = 3	has Coherence Value of 0.1639
Num Topics = 4	has Coherence Value of 0.2045
Num Topics = 5	has Coherence Value of 0.2148
Num Topics = 6	has Coherence Value of 0.2367
Num Topics = 7	has Coherence Value of 0.2566
Num Topics = 8	has Coherence Value of 0.2437
Num Topics = 9	has Coherence Value of 0.2487
Num Topics = 10	has Coherence Value of 0.2527
Num Topics = 11	has Coherence Value of 0.2659
Num Topics = 12	has Coherence Value of 0.2501

Рис. 2. Значение когерентности в зависимости от числа тем в тематической модели за 2020 г.

Как видно из рис. 1 и 2, показатель когерентности растет до 7 тем, после чего спад и снова рост до 11 тем, после чего снова спад. Следовательно, стоит остановиться на 7 или 11 темах в зависимости от их интерпретируемости [9].

Приведем иллюстрацию тем и частоты встречаемости основных терминов в модели с 7 темами (рис. 3).

Из рис. 3 видно, что топ-5 наиболее часто встречающихся слов включает в себя следующий набор:

- use;
- business;
- strategy;
- social;
- change.

Видимо, в большинстве статей используются бизнес-стратегии для изменений в обществе. Здравому смыслу это не противоречит.

Обратим внимание на 1-ю тему (рис. 4).

Из рис. 4 видно, что в 1-й теме в основном речь идет о влиянии цифровых технологий на изменение общества и бизнес-процессов.

Таким способом анализировались все 7 тем.

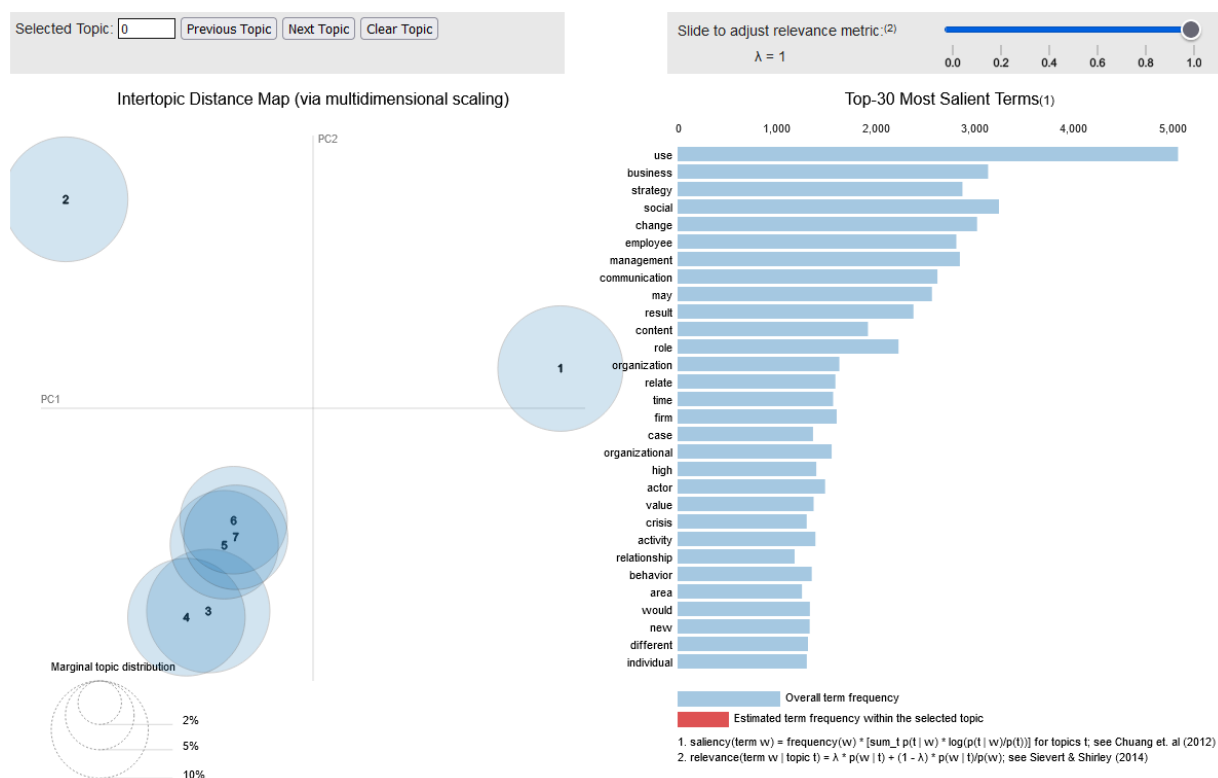


Рис. 3. Иллюстрация ключевых терминов в тематической модели из 7 тем за 2020 г.

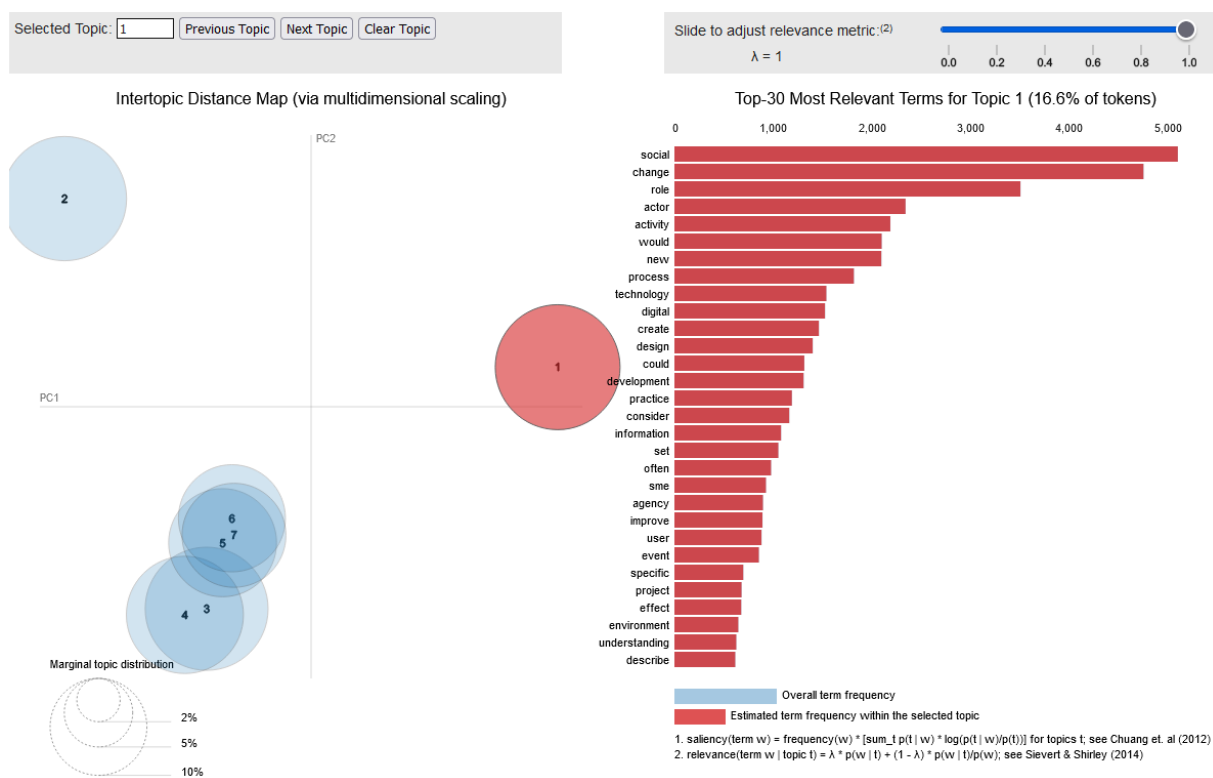


Рис. 4. Иллюстрация темы 1 в модели за 2020 г.

Ниже представлена тематическая модель LDA с 7 темами и 10 ключевыми терминами.

Тема 1. Влияние цифровых технологий на изменение общества и бизнес-процессов:

0.049*"social" + 0.046*"change" + 0.034*"role" + 0.022*"actor" + 0.021*"activity" + 0.020*"would" + 0.020*"new" + 0.017*"process" + 0.015*"technology" + 0.015*"digital".

Тема 2. Возможность индивидуализации услуг благодаря использованию новых технологий:

0.039*"may" + 0.036*"result" + 0.021*"behavior" + 0.020*"different" + 0.020*"individual" + 0.017*"level" + 0.017*"company" + 0.015*"platform" + 0.014*"future" + 0.013*"follow".

Тема 3. Нацеленность менеджмента на оптимальное использование ресурсов для создания ценности для потребителя:

0.079*"use" + 0.044*"management" + 0.025*"firm" + 0.024*"organizational" + 0.021*"value" + 0.019*"important" + 0.019*"show" + 0.018*"resource" + 0.018*"service" + 0.017*"provide".

Тема 4. Влияние кризиса на коммуникации сотрудников и поведение потребителей:

0.048*"employee" + 0.045*"communication" + 0.024*"high" + 0.022*"crisis" + 0.022*"area" +

0.021*"consumer" + 0.020*"increase" + 0.019*"manager" + 0.014*"price" + 0.014*"condition".

Тема 5. Сетевые организации и их влияние на потребление:

0.033*"organization" + 0.032*"relate" + 0.032*"time" + 0.020*"network" + 0.017*"city" + 0.017*"datum" + 0.016*"consumption" + 0.015*"system" + 0.015*"example" + 0.014*"affect".

Тема 6. Разработка бизнес-стратегии:

0.064*"business" + 0.059*"strategy" + 0.028*"case" + 0.024*"relationship" + 0.019*"indicate" + 0.017*"strategic" + 0.014*"urban" + 0.012*"finding" + 0.011*"identify" + 0.010*"turnover".

Тема 7. Необходимость финансовой поддержки отдельных сфер деятельности:

0.042*"content" + 0.024*"capital" + 0.023*"support" + 0.021*"make" + 0.021*"community" + 0.020*"people" + 0.019*"suggest" + 0.017*"policy" + 0.016*"need" + 0.014*"perceive".

В указанной модели каждый термин приведен со своим весом, что упрощает интерпретацию тем.

Визуальный анализ и аналитический вид модели позволили проинтерпретировать 7 выявленных тем. Проинтерпретировать 11 тем оказалось затруднительно. Кроме того, как видно из рис. 5, при такой структуре наблюдается много пересекающихся тем, что нежелательно.

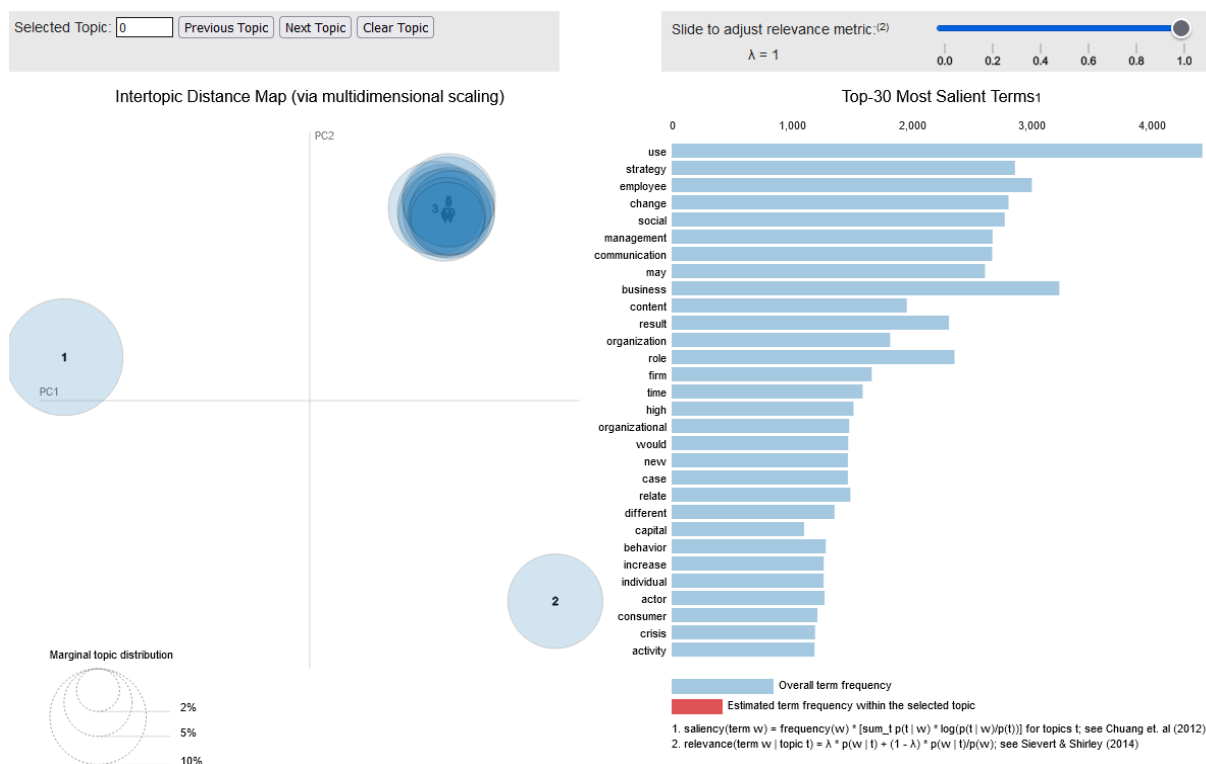


Рис. 5. Иллюстрация ключевых терминов в тематической модели из 11 тем за 2020 г.

В связи с этим было принято решение остановиться на 7 темах, которые можно воспринимать как концепты в когнитивной карте, отражающей ситуацию в области стратегического управления по данным 100 наиболее значимых публикаций из ScienceDirect за 2020 г.

**Тематические модели
ключевых публикаций
из ScienceDirect за 2016–2020 гг.**

По аналогии с 2020 г. были выявлены темы в каждом периоде с 2016 по 2019 г. В табл. 1 собраны выявленные темы по годам. Как видно из табл. 1, с течением времени появляются новые темы, например, цифровизация в 2020 г. Актуальность некоторых тем снижается. Однако неизмен-

ными являются темы, связанные с инвестициями, инновациями, результатами деятельности.

Названия тем, подобно факторному анализу, давались исходя из ключевых терминов, которые вносят основной вклад в содержание темы. Поэтому здесь может быть элемент субъективности. Чтобы его избежать, многие авторы не дают названий темам, а приводят 5 основных терминов, которые с этой темой связаны [3; 5]. Публикации, где даются названия темам, встречаются реже [11]. Считаем, что такой подход упрощает восприятие тем и их сравнение. Поэтому в статье сделана попытка отразить содержание темы в ее названии.

Т а б л и ц а 1

**Темы научных исследований в области стратегического управления
с 2016 по 2020 г.**

2016	2017	2018	2019	2020
Тема 1. Инвестиции в производство	Тема 1. Компания и рынок	Тема 1. Поддержка различных типов компаний	Тема 1. Использование информации в предпринимательстве	Тема 1. Влияние цифровых технологий на изменение общества и бизнес-процессов
Тема 2. Различный уровень сетевых ресурсов	Тема 2. Управление процессом изобретений	Тема 2. Потребность менеджмента в новых знаниях для развития фирмы	Тема 2. Высокая производительность фирм	Тема 2. Возможность индивидуализации услуг благодаря использованию новых технологий
Тема 3. Возможности развития системы	Тема 3. Результаты деятельности системы	Тема 3. Цели и результаты деятельности	Тема 3. Развитие бизнеса	Тема 3. Нацеленность менеджмента на оптимальное использование ресурсов для создания ценности для потребителя
Тема 4. Знания по предотвращению рисков компании	Тема 4. Использование стратегической политики страны	Тема 4. Разработка стратегии фирмы	Тема 4. Технологические инновации в предпринимательстве	Тема 4. Влияние кризиса на коммуникации сотрудников и поведение потребителей
Тема 5. Вклад инноваций в эффективность деятельности	Тема 5. Инновационные технологии и эффективность	Тема 5. Вклад персонала в развитие фирмы	Тема 5. Стратегия развития городской среды	Тема 5. Сетевые организации и их влияние на потребление
Тема 6. Активность в области технологического дизайна продукта	Тема 6. Развитие сотрудников	Тема 6. Альянсы и потребность в долгосрочных сценариях развития ситуации	Тема 6. Менеджмент организации, знание процессов, рынка	Тема 6. Разработка бизнес-стратегии
-	-	-	-	Тема 7. Необходимость финансовой поддержки отдельных сфер деятельности

Раскроем содержание тем за рассматриваемый период. Приведем недостающие тематические модели LDA с 2016 по 2019 г.

*Вид тематической модели LDA
за 2016 г. с 6 темами*

Тема 1. Инвестиции в производство:

0.029*"investment" + 0.021*"plant" + 0.017*"would" + 0.015*"et_al" + 0.014*"similar" + 0.012*"similarity" + 0.011*"net" + 0.011*"construct" + 0.011*"become" + 0.010*"social".

Тема 2. Различный уровень сетевых ресурсов:

0.027*"may" + 0.023*"different" + 0.022*"network" + 0.021*"resource" + 0.016*"level" + 0.016*"high" + 0.016*"datum" + 0.015*"value" + 0.015*"time" + 0.013*"factor".

Тема 3. Возможности развития системы:

0.037*"capability" + 0.036*"process" + 0.020*"actor" + 0.020*"service" + 0.018*"make" + 0.018*"develop" + 0.014*"system" + 0.013*"need" + 0.011*"sector" + 0.011*"information".

Тема 4. Знания по предотвращению рисков компании:

0.048*"knowledge" + 0.041*"use" + 0.034*"case" + 0.032*"acquire" + 0.026*"risk" + 0.023*"company" + 0.017*"asset" + 0.014*"employee" + 0.012*"customer" + 0.012*"follow".

Тема 5. Вклад инноваций в эффективность деятельности:

0.065*"target" + 0.048*"innovation" + 0.030*"commitment" + 0.028*"performance" + 0.024*"opportunity" + 0.019*"see" + 0.016*"involve" + 0.016*"behavior" + 0.013*"take" + 0.012*"evidence".

Тема 6. Активность в области технологического дизайна продукта:

0.059*"firm" + 0.035*"technology" + 0.025*"design" + 0.023*"product" + 0.022*"development" + 0.022*"activity" + 0.020*"business" + 0.020*"acquisition" + 0.018*"new" + 0.017*"relationship".

Вид тематической модели LDA

за 2017 г. с 6 темами

Тема 1. Компания и рынок:

0.028*"company" + 0.019*"market" + 0.018*"include" + 0.017*"export" + 0.015*"category" + 0.014*"information" + 0.014*"perspective" + 0.011*"improve" + 0.011*"feature" + 0.011*"public".

Тема 2. Управление процессом изобретений:

0.049*"management" + 0.024*"patent" + 0.022*"process" + 0.017*"resource" + 0.016*"strategy" + 0.015*"industry" + 0.014*"time" + 0.013*"relationship" + 0.013*"new" + 0.012*"gap".

Тема 3. Результаты деятельности системы:

0.030*"result" + 0.024*"group" + 0.024*"system" + 0.017*"gain" + 0.017*"paper" + 0.016*"show" + 0.015*"difference" + 0.015*"decline" + 0.014*"significant" + 0.014*"indicator".

Тема 4. Использование стратегической политики страны:

0.043*"use" + 0.033*"country" + 0.022*"strategic" + 0.019*"policy" + 0.018*"datum" + 0.014*"make" + 0.013*"follow" + 0.013*"phase" + 0.012*"network" + 0.011*"set".

Тема 5. Инновационные технологии и эффективность:

0.053*"innovation" + 0.036*"performance" + 0.028*"technology" + 0.021*"capability" + 0.016*"focus" + 0.014*"discipline" + 0.014*"indicate" + 0.012*"assess" + 0.011*"individual" + 0.010*"theme".

Тема 6. Развитие сотрудников:

0.029*"knowledge" + 0.019*"development" + 0.016*"value" + 0.016*"high" + 0.015*"employee" + 0.015*"develop" + 0.014*"social" + 0.014*"organizational" + 0.013*"firm" + 0.013*"business".

Вид тематической модели LDA

за 2018 г. с 6 темами

Тема 1. Поддержка различных типов компаний:

0.020*"type" + 0.019*"support" + 0.018*"different" + 0.014*"include" + 0.014*"way" + 0.014*"opportunity" + 0.012*"consider" + 0.012*"mean" + 0.011*"company" + 0.011*"category".

Тема 2. Потребность менеджмента в новых знаниях для развития фирмы:

0.039*"use" + 0.026*"paper" + 0.026*"new" + 0.022*"knowledge" + 0.022*"management" + 0.017*"develop" + 0.017*"solution" + 0.015*"problem" + 0.014*"framework" + 0.014*"author".

Тема 3. Цели и результаты деятельности:

0.035*"result" + 0.023*"activity" + 0.021*"objective" + 0.020*"method" + 0.018*"phase" + 0.016*"datum" + 0.015*"criterion" + 0.015*"technology" + 0.013*"information" + 0.013*"government".

Тема 4. Разработка стратегии фирмы:

0.035*"strategy" + 0.028*"firm" + 0.022*"process" + 0.021*"innovation" + 0.020*"performance" + 0.020*"partner" + 0.020*"may" + 0.012*"social" + 0.012*"development" + 0.012*"business".

Тема 5. Вклад персонала в развитие фирмы:

0.050*"decision" + 0.037*"employee" + 0.033*"contribution" + 0.031*"future" + 0.020*"make" + 0.016*"design" + 0.013*"case" + 0.012*"take" + 0.012*"technological" + 0.011*"different".

Тема 6. Альянсы и потребность в долгосрочных сценариях развития ситуации:

0.030*"alliance" + 0.026*"change" + 0.022*"need" + 0.022*"scenario" + 0.019*"time" + 0.015*"long" + 0.014*"infrastructure" + 0.014*"high" + 0.014*"area" + 0.013*"cost".

Вид тематической модели LDA

за 2019 г. с 6 темами

Тема 1. Использование информации в предпринимательстве:

0.033*"use" + 0.030*"entrepreneurial" +
0.025*"information" + 0.022*"intermediary" +
0.020*"article" + 0.019*"different" + 0.018*"datum" +
0.015*"industry" + 0.014*"suggest" + 0.012*"list".

Тема 2. Высокая производительность фирм:

0.032*"firm" + 0.028*"performance" + 0.024*"high" +
0.021*"may" + 0.021*"focus" + 0.019*"relationship" +
0.018*"value" + 0.018*"resource" + 0.015*"firm" +
0.015*"external".

Тема 3. Развитие бизнеса:

0.027*"business" + 0.026*"change" + 0.024*"result" +
0.024*"new" + 0.020*"opportunity" +
0.017*"development" + 0.016*"network" + 0.015*"policy" +
0.014*"make" + 0.012*"show".

Тема 4. Технологические инновации в предпринимательстве:

0.035*"technology" + 0.027*"innovation" + 0.026*"time" +
0.026*"entrepreneur" + 0.020*"system" + 0.019*"emerge" +
0.017*"support" + 0.017*"fund" + 0.016*"period" +
0.014*"author".

Тема 5. Стратегия развития городской среды:

0.035*"cluster" + 0.021*"strategy" + 0.021*"paper" +
0.020*"urban" + 0.018*"environment" +
0.016*"publication" + 0.014*"develop" +
0.012*"integration" + 0.012*"goal" + 0.011*"average".

Тема 6. Менеджмент организации, знание процессов, рынка:

0.053*"management" + 0.035*"knowledge" +
0.035*"process" + 0.030*"market" + 0.019*"level" +
0.017*"strategic" + 0.014*"action" + 0.013*"term" +
0.012*"field" + 0.012*"take".

Модель сходства тем в области стратегического управления за 2020 г.

Тематические модели представляют собой модели латентных переменных, которые опираются на корреляции между словами и скрытый в них смысл. Тематические модели могут извлекать интерпретируемую и полезную структуру без какого-либо явного понимания языка компьютером. Можно пойти дальше и построить модель сходства тем на основе метрики Хеллингера [4].

Так, для любых двух тем $x, y \in \{1, \dots, K\}$ сходство Хеллингера вычисляется следующим образом:

$$H_s(\beta_x, \beta_y) = 1 - \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{|W|} (\beta_{x_i} - \beta_{y_i})^2},$$

где β – матрица распределения слов по темам;

W – словарь набора документов.

Метрика Хеллингера дает результат в диапазоне $[0; 1]$ для двух распределений вероятностей, причем значения ближе к 0 означают меньшее расстояние, а значит, большее сходство.

Далее построим взвешенный граф по данным за 2020 г. Его можно трактовать как модель сходства тем в области стратегического управления, построенную на основе ключевых публикаций из ScienceDirect за 2020 г.

Веса ребер в графе можно определить как величину, обратную расстоянию Хеллингера [10].

Выделим сплошной линией ребра, у которых расстояние Хеллингера больше 0,5, или веса больше 2, а пунктирной линией – более удаленные темы (рис. 6).

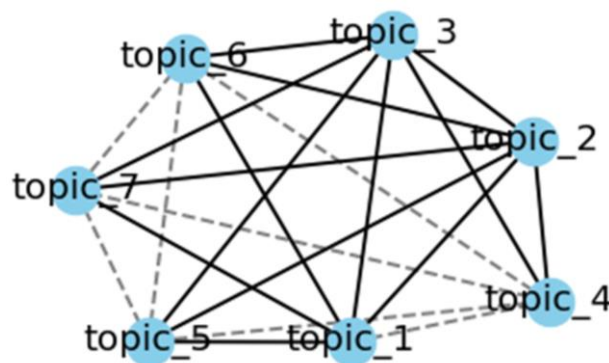


Рис. 6. Взвешенный граф с весами, рассчитанными на основе метрики Хеллингера, по данным за 2020 г.:

topic_1 – влияние цифровых технологий на изменение общества и бизнес-процессов; topic_2 – возможность индивидуализации услуг благодаря использованию новых технологий; topic_3 – нацеленность менеджмента на оптимальное использование ресурсов для создания ценности для потребителя; topic_4 – влияние кризиса на коммуникации сотрудников и поведение потребителей; topic_5 – сетевые организации и их влияние на потребление; topic_6 – разработка бизнес-стратегии; topic_7 – необходимость финансовой поддержки отдельных сфер деятельности

Выведем на экран наиболее тесно связанные темы, между которыми ребра на рис. 6 отмечены сплошной линией (рис. 7).

```
[('topic_1', 'topic_2', {'weight': 2.1616885058355844}),
 ('topic_1', 'topic_3', {'weight': 2.159167585437652}),
 ('topic_1', 'topic_5', {'weight': 2.051956704170308}),
 ('topic_1', 'topic_6', {'weight': 2.0716769761626046}),
 ('topic_1', 'topic_7', {'weight': 2.0498001542269693}),
 ('topic_2', 'topic_3', {'weight': 2.1744269101781404}),
 ('topic_2', 'topic_4', {'weight': 2.0100756305184238}),
 ('topic_2', 'topic_5', {'weight': 2.0650405139110024}),
 ('topic_2', 'topic_6', {'weight': 2.0851441405707476}),
 ('topic_2', 'topic_7', {'weight': 2.0628424925175866}),
 ('topic_3', 'topic_4', {'weight': 2.008048322256247}),
 ('topic_3', 'topic_5', {'weight': 2.0628424925175866}),
 ('topic_3', 'topic_6', {'weight': 2.0828813681835667}),
 ('topic_3', 'topic_7', {'weight': 2.060651474913109})]
```

Рис. 7. Наиболее тесно связанные темы, по данным за 2020 г.

Модели сходства тем в области стратегического управления за 2016–2019 гг.

По аналогии с 2020 г. были построены модели сходства тем за 2016–2019 гг. Приведем взвешенные графы по годам с выделением сплошной линией наиболее близких тем. Обратимся к данным за 2016 г. (рис. 8).

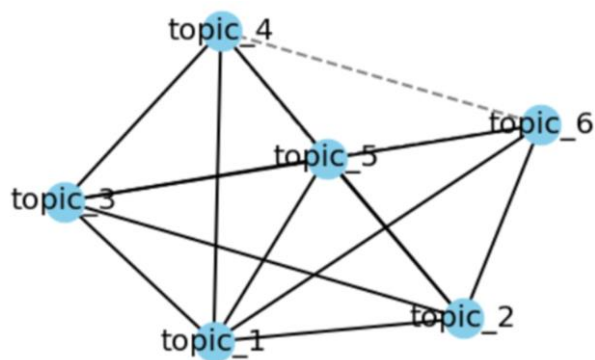


Рис. 8. Взвешенный граф с весами, рассчитанными на основе метрики Хеллингера, по данным за 2016 г.:

topic_1 – инвестиции в производство; topic_2 – различный уровень сетевых ресурсов; topic_3 – возможности развития системы; topic_4 – знания по предотвращению рисков компании; topic_5 – вклад инноваций в эффективность деятельности; topic_6 – активность в области технологического дизайна продукта.

Из рис. 8 видно, что в 2016 г. большое расстояние наблюдалось лишь между темами 4 и 6.

Обратимся к данным за 2017 г. (рис. 9).

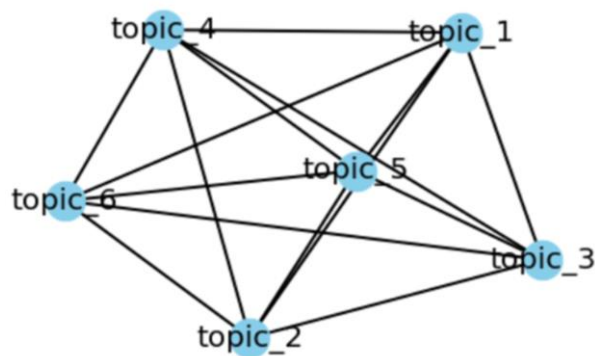


Рис. 9. Взвешенный граф с весами, рассчитанными на основе метрики

Хеллингера, по данным за 2017 г.:

topic_1 – компания и рынок; topic_2 – управление процессом изобретений; topic_3 – результаты деятельности системы; topic_4 – использование стратегической политики страны; topic_5 – инновационные технологии и эффективность; topic_6 – развитие сотрудников

Из рис. 9 видно, что сходство между всеми темами за 2017 г. достаточно высокое. Аналогичная ситуация наблюдается в 2018 и 2019 гг. Поэтому графы за эти периоды приводить не будем.

Таким образом, в период с 2016 по 2019 г. в основном наблюдается высокое сходство выявленных тем.

Наиболее значимые темы в области стратегического управления за 2020 г.

Найдем центральные (наиболее значимые) темы на основе собственного вектора x для построенного графа G . Центральность узла на основе собственного вектора вычисляется с учетом значимости его соседей. Центральность узла i на основе собственного вектора является i -м элементом вектора x из следующего уравнения:

$$A \cdot x = \lambda \cdot x,$$

где A – матрица смежности графа G с собственным значением λ .

В силу теоремы Перрона – Фробениуса существует единственное решение – вектор x , все элементы которого положительны, если λ является наибольшим собственным числом матрицы смежности A .

Результаты вычислений центральности на основе собственного вектора для построенной модели сходства тем за 2020 г. представлены на рис. 10.

```
{ 'topic_1': 0.3847517670699508,
  'topic_2': 0.38649593111706054,
  'topic_3': 0.3862045862610898,
  'topic_4': 0.3661312192789838,
  'topic_5': 0.3731481847636655,
  'topic_6': 0.3756421337205434,
  'topic_7': 0.37287308882265896 }
```

Рис. 10. Центральности тем на основе собственного вектора за 2020 г.

Как видно из рис. 10, центральными (наиболее значимыми) за 2020 г. являются первые 3 темы:

– тема 1 – влияние цифровых технологий на изменение общества и бизнес-процессов;

– тема 2 – возможность индивидуализации услуг благодаря использованию новых технологий;

– тема 3 – нацеленность менеджмента на оптимальное использование ресурсов для создания ценности для потребителя.

Наиболее значимые темы в области стратегического управления за 2016–2019 гг.

По аналогии с 2020 г. были найдены наиболее значимые темы за 2016–2019 гг.

Сведем воедино ключевые темы по годам (табл. 2).

Т а б л и ц а 2

Ключевые темы научных исследований в области стратегического управления за 2016–2020 гг.

Место	2016	2017	2018	2019	2020
1	Тема 1. Инвестиции в производство	Тема 1. Компания и рынок	Тема 1. Поддержка различных типов компаний	Тема 5. Стратегия развития городской среды	Тема 1. Влияние цифровых технологий на изменение общества и бизнес-процессов
2	Тема 2. Различный уровень сетевых ресурсов	Тема 6. Развитие сотрудников	Тема 6. Альянсы и потребность в долгосрочных сценариях развития ситуации	Тема 3. Развитие бизнеса	Тема 2. Возможность индивидуализации услуг благодаря использованию новых технологий
3	Тема 3. Возможности развития системы	Тема 3. Результаты деятельности системы	Тема 3. Цели и результаты деятельности	Тема 1. Использование информации в предпринимательстве	Тема 3. Нацеленность менеджмента на оптимальное использование ресурсов для создания ценности для потребителя

Из табл. 2 видно, что состав наиболее значимых тем меняется, видимо, под влиянием складывающейся геополитической и социально-экономической ситуации на территории авторов научных публикаций.

Заключение

Таким образом, построенная модель сходства найденных тем в области стратегического управления на основе ключевых публикаций из ScienceDirect за 2016–2020 гг. позволила выявить семантическую структуру научных публикаций, ключевые темы, волновавшие мировое научное сообщество в области стратегического

управления за рассматриваемый период, и их динамику. Было обнаружено, что состав тем из года в год меняется, появляются новые темы. Значимость тем то повышается, то понижается, но всегда актуальными в области стратегического управления являются темы, связанные с инвестициями, инновациями и результатами деятельности. В дальнейших исследованиях предстоит исследовать возможность автоматического установления причинно-следственных связей между концептами, чтобы иметь возможность моделировать ситуацию с помощью аппарата нечетких когнитивных карт.

Список литературы

1. Воронцов К. В. Аддитивная регуляризация тематических моделей // Математические методы распознавания образов. – 2013. – Т. 17. – № 1. – С. 88.
2. Axelrod R. Structure of Decision: The Cognitive Maps of Political Elites. – Princeton, NJ : Princeton University Press, 1976.
3. Blei D. M., Lafferty J. D. A Correlated Topic Model of Science // The Annals of Applied Statistics. – 2007. – Vol. 1 (1). – P. 17–35. – URL: <http://www.jstor.org/stable/4537420> (дата обращения: 07.02.2022).
4. Maiya A. S., Rolfe R. M. Topic Similarity Networks: Visual Analytics for Large Document Sets. – Institute for Defense Analyses, 2014. – URL: <http://www.jstor.org/stable/resrep23632> (дата обращения: 07.02.2022).
5. Moraes L. O., Pedreira C. E. Clustering Introductory Computer Science Exercises Using Topic Modeling Methods // IEEE Transactions on Learning Technologies. – 2021. – Vol. 14. – N 1. – P. 42–54.
6. Newman D., Lau J. H., Grieser K., Baldwin T. Automatic Evaluation of Topic Coherence // Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. – Stroudsburg, PA, USA : Association for Computational Linguistics, 2010. – P. 100–108.
7. Newman D., Noh Y., Talley E., Karimi S., Baldwin T. Evaluating Topic Models for Digital Libraries // Proceedings of the 10th annual Joint Conference on Digital libraries. – New York : ACM, 2010. – P. 215–224.
8. Papageorgiou E. Review study on Fuzzy Cognitive Maps and their Applications During the Last Decade // Paper presented at the IEEE International Conference on Fuzzy Systems. – Taipei, Taiwan, 2011.
9. Topic Modeling with Gensim (Python). – URL: <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/> (дата обращения: 07.02.2022).
10. Wang Y., Kockelman K. M., Wang X. (Cara). The Impact of Weight Matrices on Parameter Estimation and Inference: A Case Study of Binary Response Using Land-Use Data // Journal of Transport and Land Use. – 2013. – Vol. 6 (3). – P. 75–85. – URL: <http://www.jstor.org/stable/26202665> (дата обращения: 07.02.2022).
11. Ye Tian. A Multilayer Correlated Topic Model. – URL: <https://doi.org/10.48550/arXiv.2101.02028> (дата обращения: 07.02.2022).

References

1. Vorontsov K. V. Additivnaya regularizatsiya tematicheskikh modeley [Additive Regularization of Topic Models]. *Matematicheskie metody raspoznavaniya obrazov* [Mathematical Methods of Pattern Recognition], 2013, Vol. 17, No. 1, p. 88. (In Russ.).
2. Axelrod R. Structure of Decision: The Cognitive Maps of Political Elites. Princeton, NJ, Princeton University Press, 1976.
3. Blei D. M., Lafferty J. D. A Correlated Topic Model of Science. *The Annals of Applied Statistics*, 2007, Vol. 1 (1), pp. 17–35. Available at: <http://www.jstor.org/stable/4537420> (accessed 07.02.2022).
4. Maiya A. S., Rolfe R. M. Topic Similarity Networks: Visual Analytics for Large Document Sets. Institute for Defense Analyses, 2014. Available at: <http://www.jstor.org/stable/resrep23632> (accessed 07.02.2022).

5. Moraes L. O., Pedreira C. E. Clustering Introductory Computer Science Exercises Using Topic Modeling Methods. *IEEE Transactions on Learning Technologies*, 2021, Vol. 14, No. 1, pp. 42–54.
6. Newman D., Lau J. H., Grieser K., Baldwin T. Automatic Evaluation of Topic Coherence. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA, USA, Association for Computational Linguistics, 2010, pp. 100–108.
7. Newman D., Noh Y., Talley E., Karimi S., Baldwin T. Evaluating Topic Models for Digital Libraries. *Proceedings of the 10th annual Joint Conference on Digital libraries*. New York, ACM, 2010, pp. 215–224.
8. Papageorgiou E. Review study on Fuzzy Cognitive Maps and their Applications During the Last Decade. *Paper presented at the IEEE International Conference on Fuzzy Systems*. Taipei, Taiwan, 2011.
9. Topic Modeling with Gensim (Python). Available at: <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/> (accessed 07.02.2022).
10. Wang Y., Kockelman K. M., Wang X. (Cara). The Impact of Weight Matrices on Parameter Estimation and Inference: A Case Study of Binary Response Using Land-Use Data. *Journal of Transport and Land Use*, 2013, Vol. 6 (3), pp. 75–85. Available at: <http://www.jstor.org/stable/26202665> (accessed 07.02.2022).
11. Ye Tian. A Multilayer Correlated Topic Model. Available at: <https://doi.org/10.48550/arXiv.2101.02028> (accessed 07.02.2022).

Сведения об авторе

Анна Васильевна Заграновская

кандидат экономических наук,
доцент кафедры прикладной математики
и экономико-математических методов
СПбГЭУ.

Адрес: ФГБОУ ВО «Санкт-Петербургский
государственный экономический университет»,
191023, Санкт-Петербург, ул. Садовая, д. 21.
E-mail: zagranel@rambler.ru

Information about the author

Anna V. ZagranoVskaiA

PhD, Assistant Professor
of the Department for Applied Mathematics
and Economic and Mathematical Methods
of the University SPbSEU.

Address: Saint Petersburg State Economic
University, 21 Sadovaya Str., Saint Petersburg,
191023, Russian Federation.
E-mail: zagranel@rambler.ru