

**АНАЛИЗ БОЛЬШИХ ДАННЫХ
ДЛЯ ОБОСНОВАНИЯ РЕШЕНИЙ ПО НАУЧНО-ТЕХНОЛОГИЧЕСКОМУ
РАЗВИТИЮ В ЭНЕРГЕТИКЕ**

Михеев Алексей Валерьевич

к.т.н., ведущий специалист

Институт систем энергетики им. Л.А. Мелентьева СО РАН,

664130 г. Иркутск, ул. Лермонтова 130,

e-mail: mikheev@isem.irk.ru

Аннотация. В статье рассматриваются возможности применения методов анализа больших данных для принятия решений по инновационному развитию в энергетике. Выполнен библиометрический обзор научных исследований по использованию анализа больших данных для задач в сфере энергетики на основе публикаций международной базы Scopus за 2010-2020 гг. Приведены содержательные задачи мониторинга, прогнозирования и оценки перспективности технологических решений в энергетике на основе семантического анализа больших данных.

Ключевые слова: анализ больших данных, tech mining, семантический анализ текстов, научно-технологическое прогнозирование, инновационное развитие, инновационные индикаторы.

Цитирование: Михеев А.В. Анализ больших данных для обоснования решений по научно-технологическому развитию в энергетике//Информационные и математические технологии в науке и управлении. 2020. № 4 (20). С. 158-167. DOI: 10.38028/ESI.2020.20.4.014

Введение. Одними из самых сложных задач в области принятия решений в энергетике являются стратегические задачи управления развитием на уровне компании, региона или страны через внедрение передовых технологий путем модернизации действующей и создания новой энергетической инфраструктуры. Инновации в энергетике приводят к увеличению эффективности и качества генерации, конверсии, транспорта, хранения и распределения электрической и тепловой энергии, к снижению негативного воздействия на окружающую среду и в целом обеспечивают возможности экономического роста. При этом важно найти и реализовать такие пути трансформации уже существующей технологической и организационной структуры энергетики, которые будут наилучшими с точки зрения критериев технической, экономической, экологической и социально-политической эффективности. Принимаемые решения по развитию энергетических систем носят комплексный характер, отличаются высокой степенью сложности, долгосрочностью планирования и реализации, многокритериальностью выбора, неопределенностью и зависимостью от многочисленных внешних факторов [2]. Поэтому для обеспечения максимальной информированности при обосновании научно-технических рекомендаций, предложений, программ, а также при планировании и внедрении перспективных технологических инноваций в сфере энергетики крайне важно использование всех возможных и доступных данных, включая слабоструктурированные и неструктурированные,

объем которых в мире постоянно растет. Особый интерес для систем поддержки принятия решений представляет применение технологий, лежащих в русле концепции анализа больших данных: Big / Large Data Analysis – BDA или LDA.

Термин «большие данные» уже прочно вошел в оборот исследователей. Технологии анализа больших данных широко используется в качестве информационно-аналитического инструментария для решения задач во многих областях науки и техники, в первую очередь в медицине, социологии, компьютерных и технических науках [6].

Само понятие «большие данные» обычно определяется через 6 ключевых характеристик или 6-V измерений [5]. Большие данные должны обладать:

- 1) значительным объемом (*Volume*), когда величина объема характеризует полноту и репрезентативность данных для целей анализа;
- 2) разнообразием форм и источников данных (*Variety*), что подразумевает вовлечение для анализа не только структурированных, но слабоструктурированных и неструктурированных данных, которые по оценкам составляют до 95% всех полезных данных;
- 3) высокой скоростью (*Velocity*) накопления данных и регулярностью их обработки.

Некоторые исследователи [7] добавляют к этим характеристикам еще такие:

- 4) вариативность (*Variability*) или комплексность данных по времени или иной метрике, выражаемая в непостоянстве информационных потоков и гетерогенности их источников;
- 5) достоверность (*Veracity*): несмотря на ненадежность и неопределенность данных из отдельных источников, только анализ данных по совокупности позволяет получать близкие к достоверным результатам;
- 6) ценность (*Value*): часто данные, полученные в исходном виде или фрагментарно, сами по себе имеют низкую ценность, но высокую ценность можно получить, анализируя большие объемы таких данных.

Используемые методы анализа больших данных – это методы широкого спектра от классических методов математической статистики до методов глубокого обучения и нейронных сетей [8]. Подробное их рассмотрение находится за пределами данной статьи.

О популярности и востребованности аналитики на основе больших данных свидетельствует взрывной рост статей с ключевым выражением «big data» и его аналогами в международных индексах цитирования, начиная со второй декады XXI века (см. рис. 1). Именно к этому моменту в арсенале исследователей и аналитиков появились технологические возможности анализировать большие массивы накопленных и постоянно пополняемых данных, которые обладают максимально возможной информационной полнотой и доступностью. Как показано на диаграмме, с небольшим запозданием концепция больших данных начинает активно проникать в область наук о принятии решений (Decision Science) на рубеже 2014-2015 годов. Примерно в это же время начинается рост количества исследований с использованием анализа больших данных в предметной области энергетика (Energy). Однако, значимое число научных публикаций по методологии принятия решений в энергетике на основе больших данных начинает прослеживаться с достаточно недавнего времени с 2017-2018 гг. (см. область Decision Science + Energy на рис. 1), и этих пересекающихся публикаций относительно немного в общем потоке научных работ, связанных с BDA. Таким образом, текущее состояние, тренды, потенциальные пути

использования и общая полезность концепции больших данных для обоснования решений по развитию энергетических систем актуальны и требуют рассмотрения.

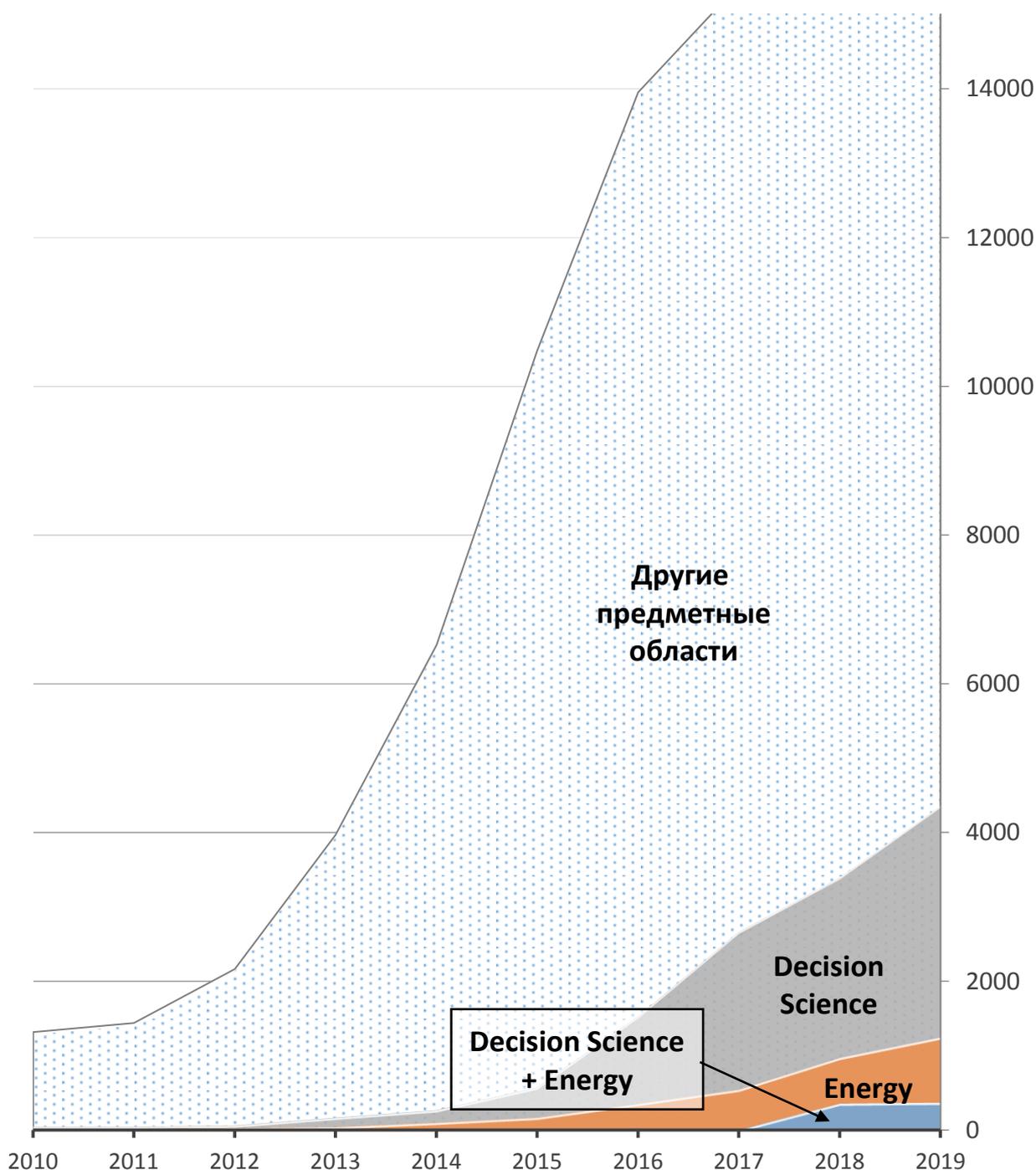


Рис. 1. Динамика количества публикаций по объединенному запросу из поисковых фраз «big data», «large data» по всем и по отдельным предметным областям наук - принятие решений (Decision Science), энергетика (Energy) и их пересечение - за 2010-2020 годы в международном индексе цитирования Scopus.

Краткий обзор применения анализа больших данных для задач энергетики.

В энергетике существует большая история работы со структурированными большими данными, особенно в электроэнергетике, где давно и успешно внедряются, постоянно совершенствуются автоматизированные системы контроля и учёта энергоресурсов (АСКУЭ). Через глубокую интеграцию электроэнергетических систем с продвинутыми

информационно-телекоммуникационными и интернет-технологиями управления и оснащение современными средствами измерений (устройства Smart Meters, PMU, FACTS и др.) происходит трансформация в интеллектуальные энергосистемы в русле концепции Smart Grid [11]. Энергокомпании используют свои накопленные большие данные для повышения эффективности функционирования систем и управления режимами, например, для прогнозирования и создания предиктивных моделей энергопотребления с использованием регрессии и обученных нейросетевых алгоритмов, мониторинга оборудования электростанций и сетей, и т.п. В перспективе планируется переход к «цифровым двойникам» энергетических систем и объектов, естественным образом погруженных в контекст больших данных [1, 7].

Поиск научных публикаций Big Data в международной базе Scopus по поисковому запросу TITLE-ABS-KEY ("big data" OR "large data" OR "tech mining") AND LIMIT-TO (SUBJAREA , "ENER") за период с 2010 по 2021 годы на начало декабря находит 4108 публикаций (количество публикаций будет увеличиваться по мере поступления новых статей, начиная с 2020 года). С помощью открытого программного средства CiteSpace [4], построив карту цитирований и применив процедуру кластеризации на основе анализа неструктурированных текстов заголовков и абстрактов статей, выделены 10 основных тематических кластеров исследований, как показано на рис. 2. Публикации в кластерах можно обобщить в несколько направлений:

1. Энергетические системы в целом с выделением электроэнергетических систем в русле концепции Smart Grid, систем распределенной генерации и решения проблем безопасности таких систем.
2. Развитие устойчивой экологичной энергетической инфраструктуры и энергоэффективных урбанистических форм будущего для «умных» городов (Smart City) и «умного» управления зданиями.
3. Разработку методов и технологий управления «умными» угольными шахтами.
4. Разработка систем мониторинга и контроля за состоянием электрических батарей как важного компонента гибридных электростанций на основе возобновляемых источников энергии.
5. Общая проблематика применения анализа больших данных для обеспечения устойчивого развития, оптимального управления цепочками продуктовых поставок, решения экологических проблем, разработки методов и алгоритмов, междисциплинарного анализа текстовых источников и т.п.

В целом большинство выделенных содержательных тем BDA лежит в области повышения эффективности работы энергетических систем на базе развивающихся технологий как информационных, так и энергетических, через потоки данных измерений и структурированной научно-технической информации о работе энергетических систем. Причем, порой сложно определить, к какому классу относятся решаемые с помощью методов анализа больших данных задачи – к классу задач управления функционированием энергетических систем или задач управления их развитием.

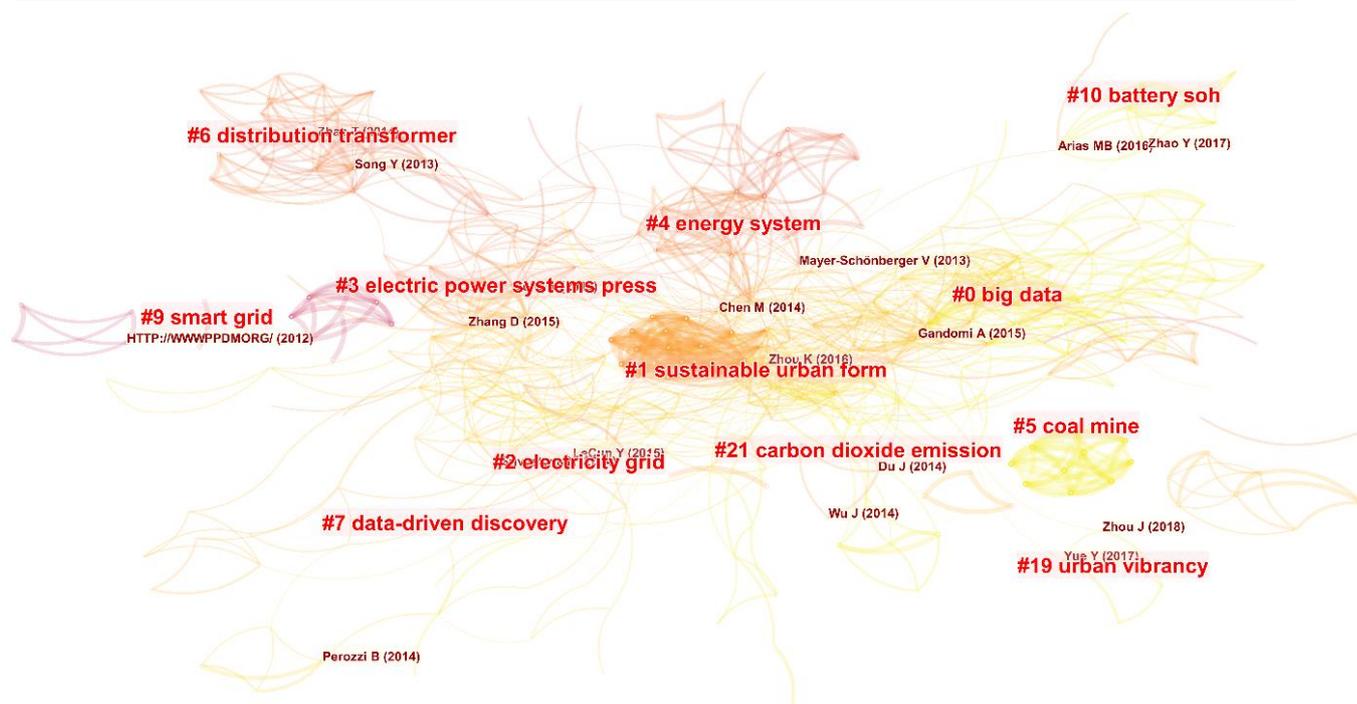


Рис. 2. Тематические кластеры, выделенные на сетевой карте совместного цитирования научных публикаций по Big Data в области энергетики в базе данных Scopus за 2010-2020 годы.

Анализ больших данных для обоснования решений по научно-технологическому развитию в энергетике. Практики использования анализа больших массивов данных для обоснования научно-технологического развития появились задолго до широкого распространения этой концепции и даже до эпохи интернета, начиная с традиционного контентного и библиометрического анализа. Одной из форм ВДА безусловно является Tech Mining (ТМ) - анализ текстовых документов из электронных ресурсов научно-технологической направленности и извлечение знаний из них в целях лучшего понимания контекста и трендов развития науки и технологий [10]. В дополнение к стандартным способам (статистические формы, финансовая статистика и т.п.) Tech Mining измеряет разнообразные статистические и семантические параметры прямых результатов исследований и разработок - научных статей, патентов, отчетов, новостных колонок СМИ и других документов. Кратко говоря, Tech Mining отвечает на вопросы «Что?», «Где?» и «Когда?». Подход ТМ не ограничивается статистическим анализом метаданных документов и связанной с ними текстовой информации, но часто включает различные методы и технологии семантического анализа полнотекстовых документов с использованием методов обработки естественного языка (Natural Language Processing) [9].

С другой стороны, в России сложилась комплексная методология системных исследований по обоснованию технологического развития энергетики [3], включающая в себя методы системного анализа, математического моделирования, технико-экономического анализа, анализа рисков и другие применительно к объекту исследования – энергетическим технологиям. Методология позволяет получать воспроизводимые, объясняемые, отражающие реальное положение дел результаты, хотя достаточно сложна в использовании

и требует высокой квалификации ее пользователей. Естественно, вывод результатов любого анализа должен быть подержан качественной экспертизой.

Привнесение концепции анализа больших данных может значительно расширить возможности приложений с использованием данной методологии, а в некоторых случаях упростить ее этапы. В таблице 1 приведен перечень обобщенных задач системных исследований по обоснованию в энергетике, которые можно рассмотреть в русле применения концепции больших данных.

Таблица 1. Задачи по обоснованию инновационного развития энергетики, сформированы на основе [3].

Этапы	Основные задачи по обоснованию инновационного развития энергетики
1	Научно-технологический мониторинг в энергетике, выявление научно-технологических тенденций, определение перспективных технологий, глобальных мировых трендов технологического развития
2	Прогнозирование технико-экономических показателей перспективных энергетических технологий, которое связано с определением зрелости технологии, а также скорости обучения. Для энергетических технологий – это как правило, удельная стоимость на единицу установленной мощности и на выработанную единицу энергии, например, USD / кВт, USD / кВт ч.
3	Системная оценка конкурирующих энергетических технологий, сравнительный анализ и выбор приоритетных технологий для развития энергетики, в т.ч. выбор технологий для разработки для последующей реализации собственных проектов НИОКР
4	Оценка системных эффектов (технических, экономических, экологических, социальных) от ожидаемых инноваций: эффекты, влияющие на энергетическую систему в целом, способны кардинально изменить выводы о целесообразности внедрения той или иной технологии.
5	Стратегическое научно-технологическое планирование, разработка дорожных карт развития технологий, формирование политики в сфере инновационного развития
6	Обеспечение выхода технологий на внутренние или внешние рынки, оценка перспектив сотрудничества и трансфера при собственной разработке или импорте технологий, оценка конкурентной среды и т.п.

На рисунке 3 представлена диаграмма взаимосвязей задач из таблицы 1 с содержательными задачами анализа больших данных. К каждой задаче BDA в качестве примеров даны ключевые фразы, кратко описывающие возможные представления полезной информации для последующей оценки экспертами. Диаграмма является предварительным эскизом для функциональной онтологии задач анализа больших данных по обоснованию научно-технологического развития в сфере энергетики.

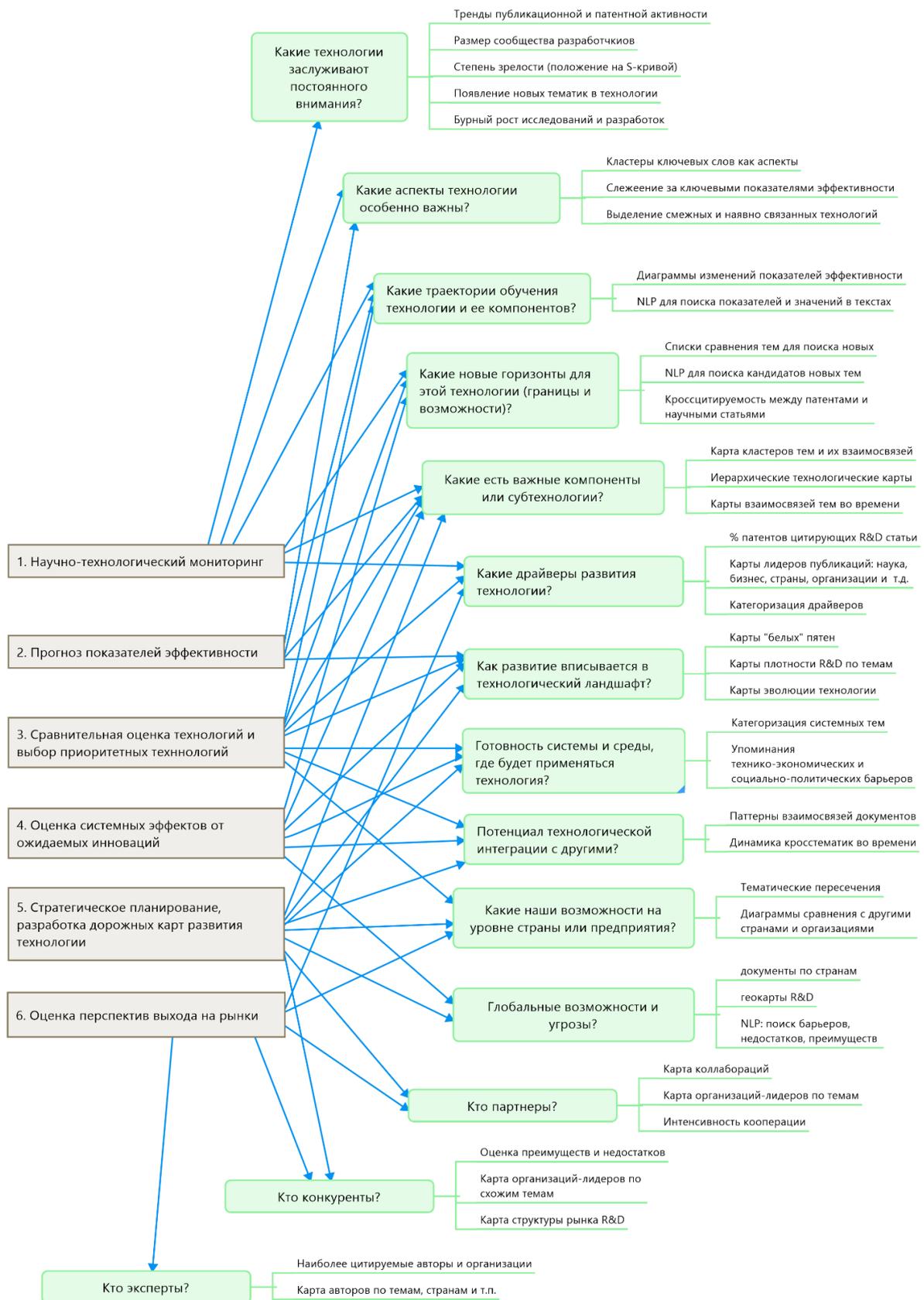


Рис. 3. Диаграмма взаимосвязей задач по обоснованию инновационного развития (серый цвет) с задачами анализа больших данных (зеленый цвет) и примерами результатов.

Заключение. Анализ больших данных имеет значительные перспективы использования для задач мониторинга и прогнозирования научно-технологического развития в энергетике. Результаты библиометрического анализа научных публикаций области больших данных показывают существенный рост интереса исследователей к этому направлению в приложениях в энергетике. Выполнено картирование между обобщенными задачами по обоснованию научно-технологического развития в энергетике и задачами семантического анализа больших данных.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ), научный проект № 20-07-00994.

СПИСОК ЛИТЕРАТУРЫ

1. Массель Л.В., Массель А.Г., Копайгородский А.Н. Эволюция технологий исследований энергетики и применения их результатов: от математических моделей и компьютерных программ к цифровым двойникам и цифровым образам // Информационные и математические технологии в науке и управлении. № 4 (16). 2019. С. 5-19. DOI: 10.25729/2413-0133-2019-4-01
2. Обоснование развития электроэнергетических систем: Методология, модели, методы, их использование / отв. ред. Н.И. Воропай. Новосибирск: "Наука", 2015. 448 с.
3. Филиппов С.П., Дильман М.Д. Системные исследования приоритетов научно-технологического развития энергетики: методологические аспекты // В кн. Системные исследования в энергетике: методология и результаты. Под. ред. Макарова А. А., Воропая Н.И. 2017. С. 63-86.
4. Chen C., Song M. Visualizing a field of research: A methodology of systematic scientometric reviews. PLoS ONE, 14(10). 2019. DOI.org/10.1371/journal.pone.0223994
5. Chen H., R.H.L. Chiang, V.C. Storey Business Intelligence and Analytics: From Big Data To Big Impact. MIS Quarterly. 36(4), 2012. Pp. 1165–1188. <http://www.jstor.org/stable/41703503>
6. Chen M., Mao S., Liu Y. Big data: A survey. *Mobile Networks and Applications*, 19(2), 2014. P. 171–209. DOI.org/10.1007/s11036-013-0489-0
7. Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*. 35(2). 2015. Pp. 137–144. DOI.org/10.1016/j.ijinfomgt.2014.10.007
8. Handbook of Big Data Analytics. Eds. Härdle W.K., Lu H. H.-S., Shen X. 2020. 538 pp. DOI.org/10.1007/978-3-319-18284-1
9. Kopyagorodsky A.N. Natural Language Processing for Forecasting Innovative Development of the Energy Infrastructure. E3S Web of Conferences 209. 03015. 2020. DOI.ORG/10.1051/E3SCONF/202020903015
10. Porter A. L., Cunningham S. W. Tech mining: exploiting new technologies for competitive advantage. 2005. 384 p.
11. Voropai N., Efimov D., Kurbatsky V., Tomin N. Smart control in the Russian electric power system // Proc. Int.Conf. of Smart Greens 2012, Porto, Portugal, 2012. Portugal. 19-20 April 2012. P.133-136. ISBN (print): 9789898565.

**BIG DATA ANALYSIS
FOR INNOVATION DEVELOPMENT DECISION MAKING
IN ENERGY SECTOR**

Alexey V. Mikheev

Ph.D., Leading specialist

Melentiev Energy Systems Institute

Siberian Branch of the Russian Academy of Sciences

130, Lermontov Str., 664033, Irkutsk, Russia, e-mail: mikheev@isem.irk.ru

Annotation. The article discusses the feasibility and possible applications of big data analysis for making decisions on innovative development in the energy sector. A bibliometric review of scientific research on the use of big data analysis for problems in the energy sector was carried out based on publications of Scopus database for 2010-2020. The substantive tasks of monitoring, forecasting and evaluating the prospects of technological solutions in the energy sector based on semantic analysis of big data are presented.

Keywords: big data analysis, tech mining, semantic analysis of texts, scientific and technological forecasting, innovative development.

References

1. Massel L.V., Massel A.G., Kopaygorodsky A.N. The evolution of energy research and application of their results: from mathematical models and computer programs to digital twins and digital images // *Informacionnye i matematicheskie tekhnologii v nauke i upravlenii* [Information and Mathematical Technologies in Science and Management]. 4(16). 2019. Pp. 5-19. DOI: 10.25729/2413-0133-2019-4-01 (in Russian)
2. Obosnovanie razvitija jelektrojenergeticheskikh sistem: Metodologija, modeli, metody, ih ispol'zovanie [Feasibility of energy power systems development: Methodology, models, methods and its use] / Eds. Voropai N.I. Novosibirsk: "Nauka". 2015. 448 p. (in Russian).
3. Filippov S.P., Dilman M.D. Sistemnye issledovaniya prioritetov nauchno-tekhnologicheskogo razvitiya energetiki: metodologicheskie aspekty [Systemic studies of the priorities of scientific and technological development of energy: methodological aspects] // In the book. Sistemnye issledovaniya v energetike: metodologiya i rezul'taty [Systems research in energy: methodology and results]. Eds. Makarov, A.A., Voropai. N.I. 2017. Pp. 63-86. (in Russian)
4. Chen C., Song M. Visualizing a field of research: A methodology of systematic scientometric reviews. *PLoS ONE*, 14(10). 2019. DOI.org/10.1371/journal.pone.0223994
5. Chen H., R.H.L. Chiang, V.C. Storey Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*. 36(4). 2012. Pp. 1165–1188. <http://www.jstor.org/stable/41703503>
6. Chen M., Mao S., Liu Y. Big data: A survey. *Mobile Networks and Applications*, 19(2), 2014. P. 171–209. DOI.org/10.1007/s11036-013-0489-0
7. Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 2015. Pp. 137–144. DOI.org/10.1016/j.ijinfomgt.2014.10.007

8. Handbook of Big Data Analytics. Eds. Härdle W.K., Lu H. H.-S., Shen X. 2020. 538 pp. DOI.org/10.1007/978-3-319-18284-1
9. Kopyaygorodsky A.N. Natural Language Processing for Forecasting Innovative Development of the Energy Infrastructure. E3S Web of Conferences 209. 03015. 2020. DOI.ORG/10.1051/E3SCONF/202020903015
10. Porter A. L., Cunningham S. W. Tech mining: exploiting new technologies for competitive advantage. 2005. 384 p.
11. Voropai N., Efimov D., Kurbatsky V., Tomin N. Smart control in the Russian electric power system // Proc. Int.Conf. of Smart Greens 2012, Porto, Portugal, 2012. Portugal. 19-20 April 2012. P.133-136. ISBN (print): 9789898565.