
МЕТОДИКА РЕГИОНАЛЬНЫХ ИССЛЕДОВАНИЙ

УДК 676

ИСПОЛЬЗОВАНИЕ ТЕКСТ-МАЙНИНГА В ЭКОНОМИКО-ГЕОГРАФИЧЕСКОМ ОТРАСЛЕВОМ АНАЛИЗЕ ЦЕЛЛЮЛОЗНО-БУМАЖНОЙ ПРОМЫШЛЕННОСТИ ЕВРОПЕЙСКОЙ РОССИИ

© 2021 г. И.Ф. Кузьминов*, П.А. Лобанова**

*Национальный исследовательский университет «Высшая школа экономики»,
Москва, Россия*

* e-mail: ikuzminov@hse.ru

** e-mail: plobanova@hse.ru

Показаны возможности анализа нетрадиционных источников данных для получения более полной и актуальной картины пространственного развития отраслей промышленности. Предложена методика исследования, включающая использование одного из инструментов анализа больших данных, а именно текст-майнинга (text mining), для прикладных экономико-географических исследований, актуальность которой определяется недостаточной полнотой и своевременностью официальных статистических данных, удешевлением соответствующих технологий обработки информации и избытком источников больших текстовых/документных данных открытого доступа. Рассмотрены место и роль целлюлозно-бумажной промышленности как ключевой отрасли отечественного лесопромышленного комплекса в экономическом и пространственном развитии современной России. Выявлены основные тенденции экономического и пространственного развития целлюлозно-бумажной промышленности Европейской России, что позволило выявить тенденции отраслевого развития и сформулировать рекомендации по стратегическим управленческим решениям в качестве ответа на создаваемые этими тенденциями вызовы. Показано, что отрасли требуется либерализация и стабилизация, в первую очередь путем мораториев на изменения в политике. Подчеркнута роль применения анализа больших данных, и в частности текст-майнинга, в экономико-географических исследованиях для формирования обоснованных и объективных выводов, которые смогут использоваться для принятия своевременных и взвешенных управленческих решений в лесопромышленном комплексе и целлюлозно-бумажной промышленности.

Ключевые слова: Big Data, текст-майнинг, целлюлозно-бумажная промышленность, Европейская Россия, предприятия, лесопромышленный комплекс, производство бумаги, целлюлозы и древесины.

DOI: 10.5922/1994-5280-2021-1-2.

Введение и постановка проблемы. Обоснованная оценка стратегических возможностей и рисков, связанных с развитием, сжатием, существованием или исчезновением тех или иных отраслей в структуре отечественной промышленности, играет важную роль в разработке и реализации эффективной экономической политики в России. Тем не менее, в отечественной географической науке существует проблема нехватки регулярных межатраслевых исследований, показыва-

ющих актуальное состояние и тенденции пространственного развития отдельных отраслей с опорой как на официальную статистическую информацию и материалы полевых исследований, так и на разного рода большие данные (в том числе текстовые массивы с геопривязкой информационных поводов, содержащихся в этих материалах). По мнению авторов, использование больших данных (Big Data), характеризующихся полнотой и постоянным обновлением,

в отличие от ограниченной в охвате анализируемого явления и запаздывающей на 1–2 года официальной статистики, может позволить сделать отраслевые экономико-географические исследования регулярными и приближенными к формату мониторинга, и тем самым полезными для практического принятия решений в госуправлении и корпоративном секторе.

В связи с важной ролью целлюлозно-бумажной промышленности (ЦБП) в лесопромышленном комплексе (ЛПК) России, ценностью данной отрасли для экспортной валютной выручки страны и социально-экономического развития ряда регионов, авторы считают актуальным проведение исследования, отражающего объективную ситуацию в отрасли, с применением текст-майнинга (text mining). Научная значимость предлагаемого исследования связана с полномасштабным применением современных технологий текст-майнинга наряду с традиционными методиками отраслевых экономико-географических исследований. Такая комбинация методов позволит получить значимые данные о тенденциях пространственного развития отрасли, которые без применения текст-майнинга не могли бы быть получены в принципе. Тем самым исследование вносит вклад в становление нового, пока только зарождающегося направления прикладных экономико-географических исследований, которое можно назвать геосемантическим отраслевым анализом.

Обзор ранее выполненных исследований по теме. Географические отраслевые исследования – давно существующий и активно развивающийся в России тип географических исследований [2; 4; 16]. Однако в научном сообществе нет единого подхода к пониманию понятия «отрасль». В рамках данного исследования авторы понимают под отраслью совокупность субъектов хозяйствования, осуществляющих сходный вид деятельности [14]. Поэтому в состав ЦБП авторы включают как целлюлозно-бумажные комбинаты, так и небольшие фабрики по производству картона и гофротары, так как основным критерием определения принадлежности того или иного предприятия к отрасли в рамках данного исследования выступает не высокотехнологичность или масштаб производства, а принадлежность конечной продук-

ции к номенклатуре целлюлозной, бумажной или картонной продукции. Такой подход позволяет продемонстрировать возможности проведения многостороннего и расширенного семантического анализа, включая анализ сложившихся связей ключевых тематик отрасли с опорой на отраслевые источники информации, в том числе новостные.

Согласно одному из наиболее подробных и часто используемых определений, отраслевая география промышленности исследует закономерности развития отдельных отраслей промышленности, их взаимодействие на всех территориальных уровнях с другими отраслями промышленности и экономики [15]. Отдельное направление отраслевых географических исследований – промышленное районирование [12; 18], под которым понимается способ выделения в пределах экономически освоенного пространства территорий, специализирующихся на промышленных видах деятельности [15].

Часть отраслевых географических исследований посвящена анализу ЛПК и ЦБП. Подобные исследования отличаются многообразием. Например, в них проводится анализ географии мирового ЛПК [17], пространственной организации и концентрации производств [8], структуры экспорта и сдвигов в расположении производственных объектов [25]. Активно развиваются междисциплинарные и межотраслевые исследования [1]. География глубоко укореняется в экономике, создавая направление новой экономической географии: при разработке экономических моделей активно учитываются пространственные факторы [9].

На стыке смежных отраслей начинает зарождаться и новое направление, которое можно назвать геосемантическим отраслевым анализом, основанным на автоматизированном анализе больших массивов текстов (текст-майнинге). Анализ больших данных и в частности текст-майнинг начали применяться в отраслевой аналитике не так давно, но количество подобных исследований как в зарубежном, так и в российском научном сообществе с каждым годом растет [7; 19; 22]. Тенденция не обошла стороной и экономико-географические исследования, в которых текст-майнинг используется, в том числе, для выявления устойчивых тематик, преобладающих в массиве проанализированных документов, которые могут определить

основные тенденции развития отраслей [6; 24; 26]. К числу таких исследований можно отнести и настоящее исследование.

Материалы и методика исследования. Материалами проведенного исследования текущего состояния и тенденций пространственного развития целлюлозно-бумажной промышленности Европейской России стали как традиционные источники данных (научная литература, отраслевая аналитика и новости в профессиональных СМИ, официальная федеральная и региональная статистика и др.), так и относительно новый для отечественной экономической географии источник данных – материалы текст-майнинга.

Текст-майнинг представляет собой метод интеллектуального анализа больших данных, направленный на автоматическое извлечение ценных сведений из больших, принципиально не поддающихся экспертному прочтению, коллекций текстовых документов разного рода, включая научные статьи, отраслевую периодику, аналитические отчеты, экспертные интервью в СМИ и др.

Наиболее практически полезными направлениями извлечения ценных сведений из больших массивов текста для отраслевого экономико-географического анализа являются: а) извлечение тематических тегов (например, наименований предприятий и продуктов) с их автоматической геопривязкой; б) извлечение предметных терминов и расчет частоты их встречаемости, в том числе в динамике и в географической привязке, в) извлечение из текстов экспертно-ценных утверждений о фактах и их систематизация (например, количественные оценки, утверждения о запуске или остановке производства), г) сентимент-анализ, или анализ тональности публикаций в контексте той или иной тематики (например, выявление преобладающего характера публикаций о том или ином предприятии).

Главная ценность применения текст-майнинга для экономико-географических исследований состоит в восполнении пробелов в данных там, где исследуемые явления не покрыты официальной статистикой, либо официальная статистика сильно запаздывает. Наличие индикаторов на основе альтернативных данных может быть полезно для формирования гипотез о текущем состоянии отрасли. Преимущество метода состоит в его

дешевизне по сравнению с анализом других видов больших данных и релевантности извлекаемых им данных для задач отраслевых экономико-географических исследований. Эффективность использования метода текст-майнинга для формирования содержательных экспертных выводов о тенденциях пространственного развития отрасли достигается при комбинации с использованием других типов данных и исследовательской интуиции для формирования не сугубо технических выводов о распределениях статистических переменных, а содержательных выводов, имеющих экспертный и управленческий смысл.

В качестве технического инструмента для использования текст-майнинга в настоящей статье была использована информационная система *iFORA*, созданная с участием авторов НИУ «Высшая школа экономики», информация о структуре источников данных и архитектуре которой доступна в публикациях и открытых выступлениях авторов [29]. Помимо этого, в исследовании был использован также общедоступный инструмент *Google Ngram / Google Trends*.

Ниже приведены основные результаты исследования, включая оценку состояния и перспектив развития отрасли с позиций возможных социально-экономических последствий для нестоличных регионов и городов-промышленных центров ЦБП, а также сельских территорий, зависимых от деятельности, связанной с поставкой сырья на отечественные целлюлозно-бумажные комбинаты (ЦБК).

Полученные результаты и их обсуждение. Углубленное изучение лесопромышленного комплекса России целесообразно начать с анализа освещаемой в отраслевых профессиональных медиа тематики ЛПК. На рисунке 1 представлена карта числа новостей, упоминающих возможное строительство новых ЦБК по регионам Европейской России (по данным Яндекс.Новости). Размер пунсонов на карте прямо пропорционален количеству новостей, упоминающих строительство новых ЦБК в регионе. Прямоугольниками отмечены некоторые зоны, на которые будет акцентироваться внимание при анализе карты.

На карте, представленной на рисунке 1, виден ряд закономерностей. Большинство

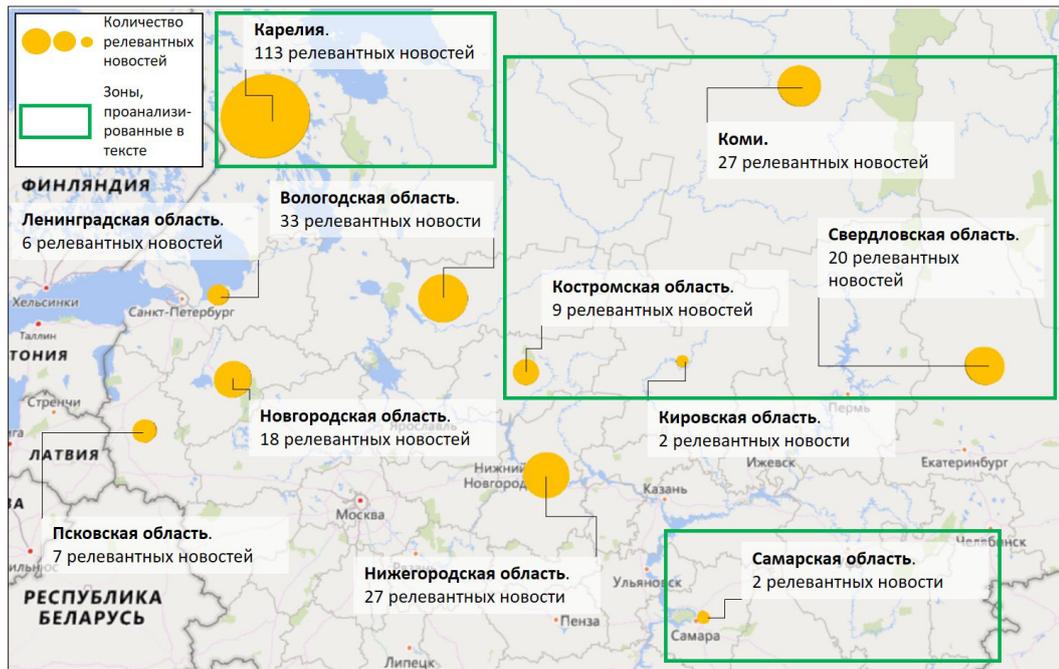


Рис. 1. Карта числа новостей, упоминающих возможное строительство нового ЦБК по регионам Европейской России.

Источник: данные Яндекс.Новости.

упоминаний приходится на Карелию как регион, обладающий одновременно комплексом факторов, благоприятных для развития лесопромышленных производств. Незначительное число новостей или их отсутствие по западным лесным регионам страны (Псковская, Смоленская, Брянская области) говорит о недостаточно сильном спросе на российскую целлюлозную продукцию со стороны европейских стран [28]. Наличие значительного числа новостей о возможном строительстве новых ЦБК в отношении неудачно расположенных с транспортно-экспортной точки зрения Костромской области, Республики Коми, Свердловской области свидетельствует об инерции мышления многих лиц, принимающих решения, продолжающих мыслить в советской логике предложения, а не спроса: наличия ресурсной базы, а не удобных каналов сбыта [32].

Анализ упоминаний в новостях возможного строительства нового ЦБК по регионам Европейской России можно дополнить, рассмотрев динамику изменения интенсивности появления новостей на эту тему (по данным системы интеллектуального анализа

больших данных *iFORA*) и динамику популярности запросов по теме «Целлюлозно-бумажное производство» (по данным *Google Trends*). Диаграмма, отражающая динамику этих показателей, представлена на рисунке 2.

Диаграмма демонстрирует наблюдаемое в последние годы постепенное, но стабильное угасание практически до нуля дискурса о возможном строительстве новых крупных ЦБК в РФ. Можно видеть корреляцию соответствующих информационных поводов с циклами экономического развития РФ. Быстрый рост значимости данной тематики стал наблюдаться с 2002 г., когда тенденции восстановительного промышленного роста в РФ, укрепленный рост валютной выручки за счет роста цен на углеводороды, проявились с полной силой. Мировой финансово-экономический кризис резко подорвал инвестиционные ожидания, которые сменились пессимизмом, снижением активности подготовки и пиара проектов по строительству новых ЦБК в РФ [5]. После легкого оживления в 2011 г. тема впоследствии стала сходиться на нет, демонстрируя практически полную рыночную бесперспективность

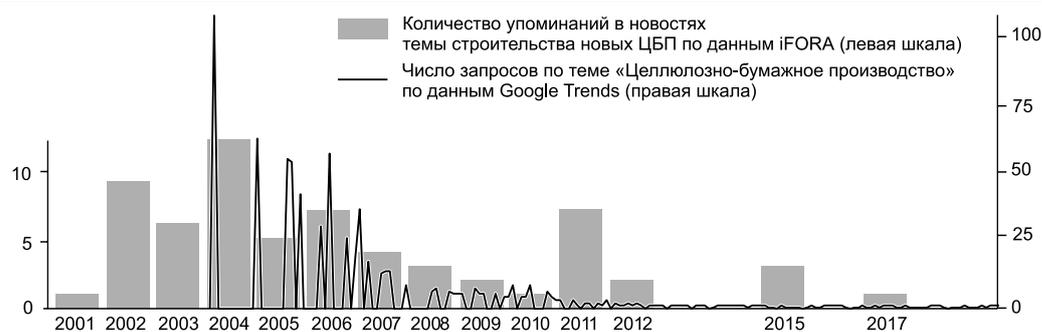


Рис. 2. Статистика упоминаний тематик новых ЦБК в России.

крупных гринфилд-проектов в ЦБК России, по крайней мере в условиях имеющихся законодательных и административных барьеров ведения бизнеса [11].

Результаты более углубленного текст-майнинга представлены в виде семантической карты и тренд-карты целлюлозно-бумажной промышленности России, построенных на основе интеллектуального анализа открытых отраслевых источников информации за 2011–2020 гг. (см. рис. 3).

Семантическая карта – это структурно-инфографический аналитический продукт, который является результатом интеллектуальной обработки миллионов текстов в базе данных системы *iFORA*. На семантической карте представлены наиболее значимые тематики, часто встречающиеся в текстах. Они обозначены на карте пунсонами с подписями. Координаты пунсонов определяются путем проекции на плоскость многомерных векторных представлений тематик (используются модели *word2vec*). Чем сильнее контекстная связь между тематиками, тем ближе друг к другу они расположены на семантической карте. Дополнительно тесно взаимосвязанные тематики группируются в устойчивые кластеры, показанные пунсонами одного цвета. Названия кластеров даются автоматически на основе определение наиболее значимой (часто упоминающийся) тематики кластера.

Размер пунсонов и их подписей отражает величину параметра векторной центральности тематик: чем в большей степени тематика близка по контексту употребления с остальными выявленными тематиками, тем крупнее соответствующий пунсон и его подпись.

Построенная семантическая карта показывает цифровой след ЦБК России. Тематики

на карте не сгруппированы в обособленные группы разного цвета, расположенные далеко друг от друга, напротив, разные цвета скорее смешаны, что может свидетельствовать о тесной смысловой связи выявленных тематик и взаимопроникновении технологий и наработок отдельных поднаправлений отрасли друг в друга. Среди центральных тематик, формирующих семантическую основу области, можно выделить производство бумаги и бумажную промышленность, древесную массу, волокнистую массу, производство целлюлозы, макулатуру, бумажные отходы, производство картона и др.

Анализ особенностей взаимосвязи ключевых тематик ЦБК России дополняется анализом их значимости и динамичности, результат которого представлен на тренд-карте на рисунке 4. Представленные на семантической карте тематики группируются на тренд-карте по четырем категориям (квадрантам). Правый верхний квадрант тренд-карты охватывает наиболее значимые и динамично развивающиеся тематики. Для них характерны высокая частота упоминаний и ее устойчивый рост в проанализированных источниках информации. В правом нижнем квадранте сконцентрированы тематики, которые отличаются относительно невысокой значимостью, но высокой динамичностью. В левом верхнем квадранте представлены стабильные тренды (с высокой, но почти не растущей значимостью), в левом нижнем – тематики, которые демонстрируют невысокую значимость и динамичность, но могут получить развитие в будущем. Как и на семантической карте, совокупности тесно связанных тематик, формирующие тематические области, отображаются на тренд-карте круговыми значками одного цвета. Размер

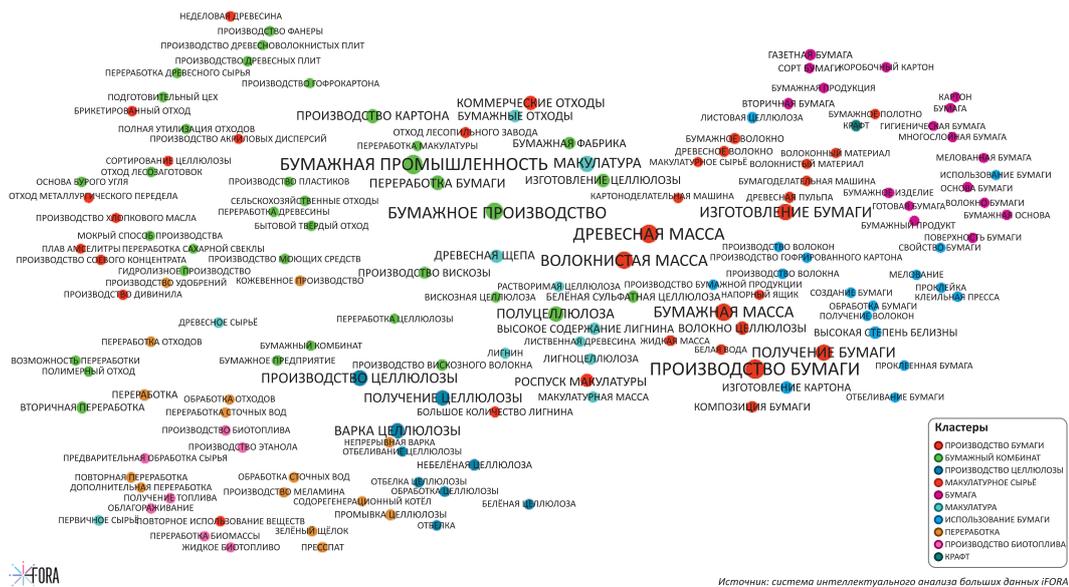


Рис. 3. Семантическая карта целлюлозно-бумажной промышленности России.

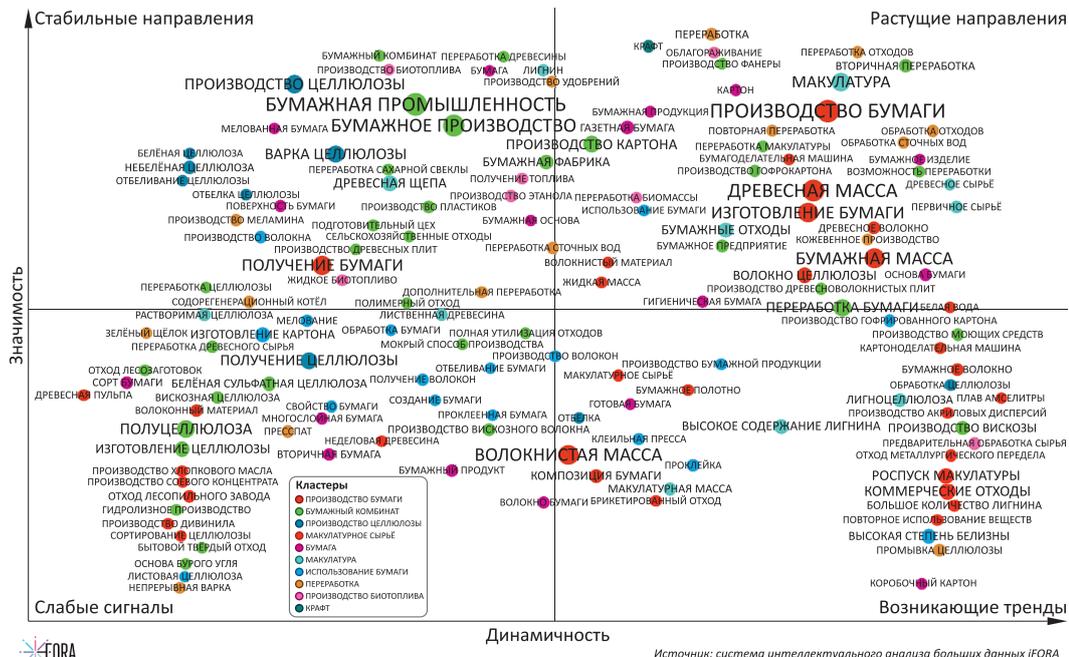


Рис. 4. Тренд-карта целлюлозно-бумажной промышленности России.

пунсонов и их подписей отражает величину векторной центральности.

К растущим направлениям относятся производство бумаги и фанеры, древесная масса, бумажная масса, бумажные отходы, макулатура, переработка бумаги, макулатуры и биомассы и др. В квадранте стабильных на-

правлений расположены: бумажная промышленность, производство целлюлозы, производство древесных плит, варка целлюлозы, получение топлива, сельскохозяйственные отходы и др. К возникающим трендам относятся такие тематики, как волокнистая масса, высокое содержание лигнина, роспуск

макулатуры, коммерческие отходы, производство вискозы, высокая степень белизны и др. Среди слабых сигналов выделяются: полуцеллюлоза, белёная сульфатная целлюлоза, вторичная бумага, непрерывная варка, полная утилизация отходов, неделовая древесина, проклеенная бумага и др. В целом наблюдается равномерное распределение тематик по квадрантам, что говорит об отсутствии как критического упадка, так и резкого роста отрасли.

Анализ освещаемой в профессиональных новостных источниках целлюлозно-бумажной промышленности России в рамках данной статьи будет дополнен анализом ряда статистических показателей и визуализаций, построенных на их основе, в том числе картограмм размещения предприятий ЦБК. Рисунок 5 демонстрирует географию прекративших свое существование ЦБК. Пунсонами с черной обводкой показан размер выручки предприятия. Пунсонами фиксированного размера с пунктирной черной обводкой отмечены прекратившие свое существование ЦБК, для которых отсутствуют данные по выручке. Размер всех пунсонов прямо пропорционален размеру выручки предприятия. Рядом со всеми пунсонами на выносках приведена краткая справка, содержащая название ЦБК, область, в которой он функционировал, дата ликвидации, точное значение выручки (при

наличии) и период, за который приводится выручка. В целях проведения объективной оценки показатели выручки приводятся по возможности за благополучный для ЦБК период за несколько лет до ликвидации. Здесь и далее авторы используют значения выручки за год по нескольким причинам. Во-первых, она отражает объем экономических операций, то есть во многом физический базис производства. Во-вторых, она показывает на низовом уровне ситуацию с ВВП, в отличие от прибыли, диктуемой мировой финансовой конъюнктурой. Наконец, корреляция выручки и частоты упоминаний за год достигает по разным отраслям и контекстам 0.9 и обычно не опускается ниже 0.75. При этом сильной корреляции прибыли с частотой упоминаний не наблюдается. Дополнительно, если для рисунка 5 выручка приводится скорее как справочный показатель, для более обоснованного анализа функционирования действующих ЦБК (см. рис. 6) показатель выручки дополняется пятью другими показателями, основанными на текст-майнинге отраслевых источников за десять лет.

Площадью прямоугольных диаграмм рядом с пунсонами отражен размер населенного пункта, в котором расположен ЦБК. Над прямоугольными диаграммами приводится название населенного пункта, под ними – размер населенного пункта.

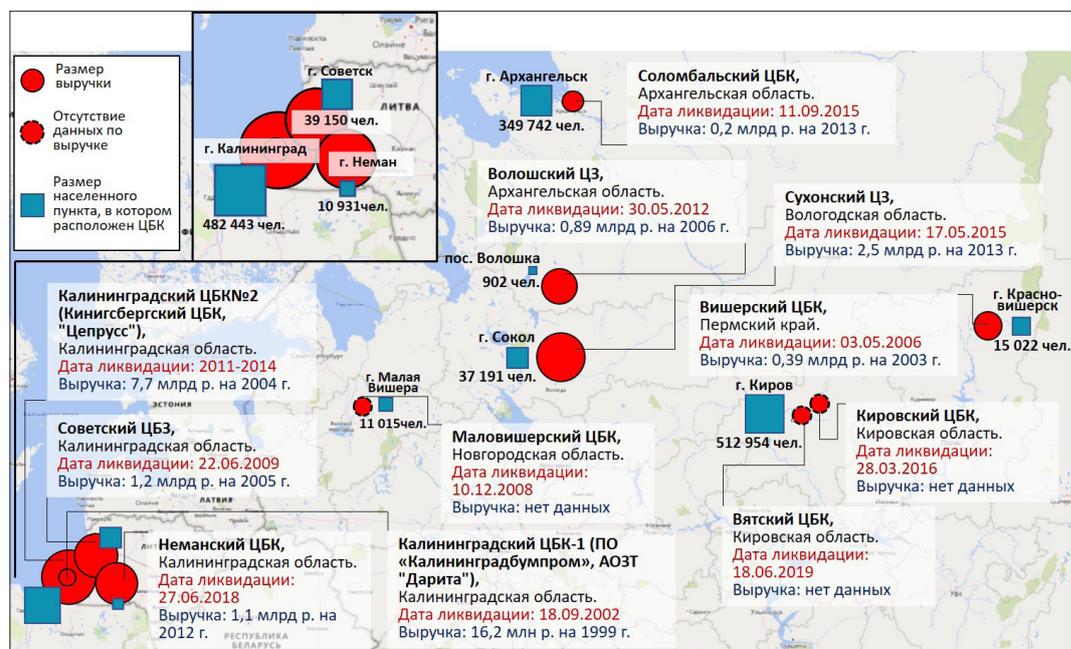


Рис. 5. Ликвидированные целлюлозно-бумажные предприятия Европейской части России (2002–2019 гг.).

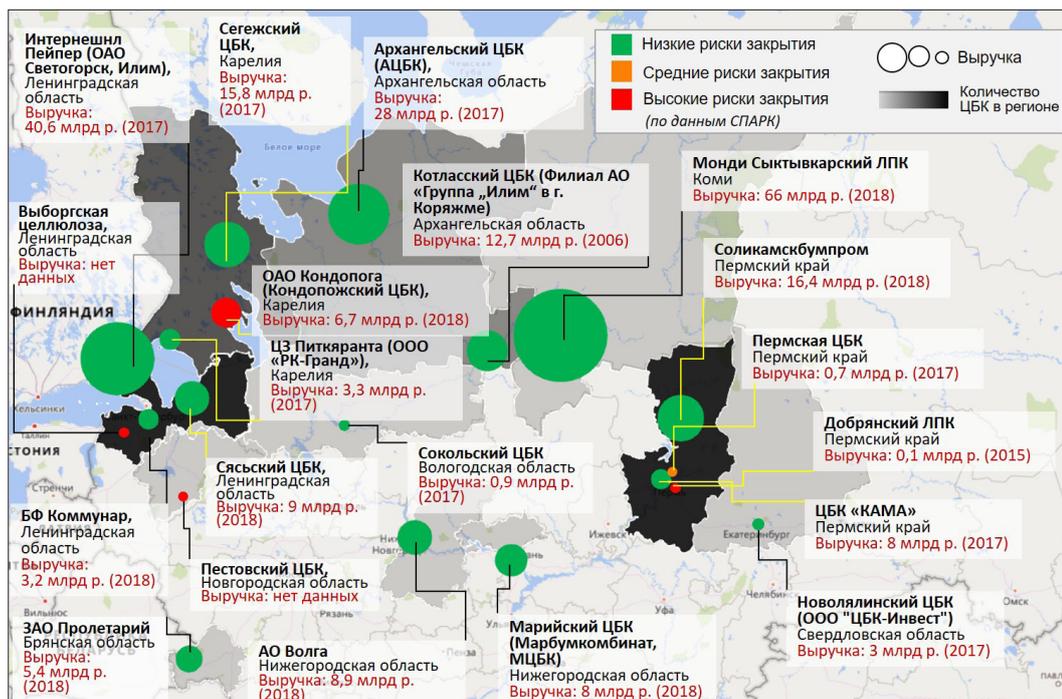


Рис. 6. Действующие целлюлозно-бумажные предприятия Европейской части России (данные по состоянию на 2019 г.).

Из карты, представленной на рисунке 5, видно, что процесс санации в отрасли носит долговременный и стабильный характер, с относительно равномерной ликвидацией неконкурентоспособных предприятий ЦБП в течение последних 15 лет. При этом процесс, по-видимому, еще не приблизился к окончанию. Последние закрытия относятся к 2018 г., в центральной части Европейской России еще сохранились относительно малые ЦБК, входящие в группу риска с точки зрения устойчивости бизнеса.

Ликвидация ЦБК в транспортно-географически удобно расположенной Калининградской области демонстрирует доминирование других факторов конкурентоспособности в России, включая административные и иные барьеры, криминализацию, политические взаимоотношения с ЕС, проводящим эффективную политику по защите внутреннего рынка и отечественных производителей, проблемы качества продукции и ее стоимости, определяющей высокими удельными издержками от неэффективного менеджмента [27; 30].

В дополнение к карте ликвидированных целлюлозно-бумажных предприятий Европейской части России приводится карта

действующих целлюлозно-бумажных предприятий (см. рис. 6). Размер пунсонов диаграмм на карте пропорционален выручке предприятия. На выносках приводится краткая справка, включающая название предприятия, регион, размер выручки и дату, на которую был получен размер выручки. Цветом пунсонов показаны риски закрытия предприятия (по оценкам системы СПАРК). Зеленый цвет означает низкие риски закрытия, оранжевый – средние риски, красный – высокие риски закрытия. Насыщенность заливки региона прямо пропорциональна количеству ЦБК в регионе.

Авторы предлагают методику исследования действующих целлюлозно-бумажных предприятий, основанную на совмещении анализа финансово-экономических показателей деятельности предприятий и в целом отрасли и методики текст-майнинга. Для этого была построена матрица действующих целлюлозно-бумажных предприятий Европейской части России на основе отраслевых источников информации за 2011–2020 гг. В строках матрицы расположены предприятия, а в столбцах – показатели, которые позволяют сделать исследование комплексным, а его результаты – обоснованными

и практичным: потенциал развития предприятия, обсуждаемость и поиск решений угроз развития, технологические внедрения и инновации, конкурентоспособность и устойчивость. Размер пунсона отражает интенсивность значимой совместных упоминаний в отраслевых источниках информации предприятий и совокупности ключевых слов, отражающих выбранные показатели. Комплект ключевых слов для каждого показателя был получен автоматически с помощью мультипликации показателя на отдельные компоненты и синонимы: слова, наиболее близкие по смыслу к показателю, согласно модели *word2vec*. Полученные списки были экспертно валидированы авторами статьи в целях исключения из анализа неизбежных для интеллектуального анализа данных «ложных срабатываний» – слов, которые система ошибочно посчитала релевантными. Для каждого показателя синим цветом показаны три наиболее интенсивно упоминаемые предприятия, а красным – три наименее интенсивно упоминаемые предприятия, за исключением тех, которые не набрали достаточное количество значимых упоминаний, чтобы быть представленными на матрице (пунсоны на пересечении отсутствуют).

Карта и матрица, представленные на рисунках 6 и 7, позволяют выделить четыре пространственно-типологические группы (пояса) предприятий в Европейской части

страны, с различными для каждой группы характеристиками размера, географического положения, потенциала развития, конкурентоспособности и рисков. При этом все четыре пояса имеют наглядное широтно-долготное (диагональное) простираение с юго-запада на северо-восток:

1) Первый пояс объединяет высоко конкурентоспособные, крупные и устойчивые ЦБК Ленинградской области, Республики Карелия и Архангельской области, входящие в том числе в состав крупных промышленных групп, таких как Илим. На данный пояс приходится большая часть целлюлозно-бумажного производства Европейской части России. По результатам текст-майнинга эти предприятия характеризуются высоким потенциалом развития, интенсивностью обсуждений и поиском решений угроз развития, высокой степенью технологических внедрений и инноваций, конкурентоспособностью и устойчивостью. К примеру, Архангельский ЦБК вошел в топ-3 предприятий по всем проанализированным показателям. Предприятие «Интернешнл Пейпер» (ОАО Светогорск, Илим, Ленинградская область) также вошло в топ-3 по всем показателям с немного меньшими значениями семантических показателей, чем Архангельский ЦБК. Другой представитель этой группы – Сегежский ЦБК (Республика Карелия) – характери-

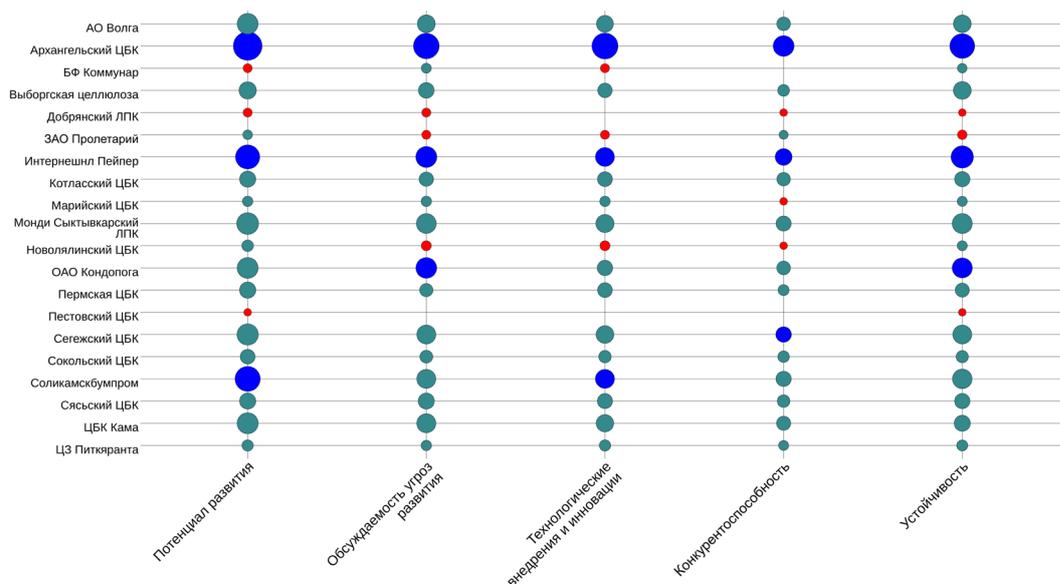


Рис. 7. Матрица действующих целлюлозно-бумажных предприятий Европейской части России.

зуются высокой конкурентоспособностью и семантической значимостью остальных показателей на уровне выше среднего (размер пунсонов – крупнее среднего).

2) Второй (вырождающийся) пояс объединяет малые, потенциально нестабильные предприятия в депрессивной по ряду отраслей лесной промышленности и АПК зоне удаленных от побережья нечерноземных регионов (Новгородская, Вологодская и соседние области). Малый размер, не вполне удачное расположение и малочисленность ЦБП, а также факты закрытия аналогичных ЦБП в предыдущие годы свидетельствуют о высоких рисках стабильности ЦБП в этом поясе, вероятном их закрытии в перспективе. Например, Пестовский ЦБК (Новгородская область) не только характеризуются высокими рисками закрытия по данным СПАРК, но и самой низкой значимостью по всем семантическим показателям. Предприятие вошло в топ-3 предприятий с наименьшим потенциалом развития и наименьшей устойчивостью и не набрало значимого числа упоминаний контексте остальных показателей в проанализированных источниках. Все семантические показатели для Сокольского ЦБК (Вологодская область) продемонстрировали значения ниже среднего (небольшой размер пунсонов).

Неясные перспективы имеющихся в Новгородской и Вологодской области ЦБП, отсутствие ЦБП в других регионах пояса позволяет говорить об этом поясе как перспективным в качестве лесосырьевой базы для ЦБК, лежащих от него к северо-западу и юго-востоку первого и третьего поясов. В первую очередь следует обратить внимание на сырьевые ресурсы Вологодской, Костромской, Кировской областей. Очевидно тяготение сырьевых зон к предприятиям (запад: Псковская, Новгородская, Смоленская, Тверская, Брянская области – тяготение к кластеру целлюлозно-бумажных предприятий Ленинградской области; центр: Ярославская, Владимирская, Вологодская, Костромская – к предприятиям одновременно Архангельской и Нижегородской областей; северо-восток: Кировская область, Республика Коми, север Нижегородской области – к кластеру предприятий Поволжья/Урала).

3) Третий пояс – кластер предприятий Поволжья и Урала, характеризующийся от-

носительно крупными размерами, стабильностью и «укорененностью» предприятий, наличием четких рыночных ниш, прежде всего на внутреннем рынке, включая рынок госзаказа. Например, Соликамскбумпром (Пермский край) вошел в лидеры по семантическим показателям потенциала развития и внедрения новых технологических решений и инноваций. Однако невыгодное континентальное положение, не вхождение в крупные промышленные группы, бренды, узкая специализация создают для данной группы определенные риски, которые могут реализоваться в случае отсутствия проактивных усилий со стороны менеджмента данных предприятий по стратегической переориентации. Это подтверждают и результаты текст-майнинга, согласно которому некоторые предприятия этого пояса показали наименьшие значения по ряду семантических показателей: например, Марийский ЦБК – по конкурентоспособности, Добрянский ЛПК – по всем четырем показателям, по которым было найдено значимое число упоминаний. Последствиями может быть в том числе проигрыш в борьбе за сырьевые ресурсы, когда истощенные сырьевые базы северо-западных предприятий заставят их все больше удлинять плечо доставки сырья. В плане перспектив переориентации обращает на себя внимание близость данных предприятий к Волжско-Камской транспортной артерии, обеспечивающей потенциально как подвоз сырья, так и имеющей выход в Каспийское море для возможного экспорта продукции в Иран, Турцию, страны Закавказья и Центральной Азии [10].

4) Четвертый (вырожденный) пояс включает разрозненные, малые предприятия, в том числе вне основной зоны размещения лесной промышленности, включая Новолялинский ЦБК и ряд нереализованных проектов в нелесных регионах. По результатам текст-майнинга Новолялинский ЦБК набрал наименьшие значения по трем показателям: интенсивность обсуждений угроз развития, технологические внедрения и инновации, конкурентоспособность. Остальные показатели для предприятия оказались на уровне значительно ниже среднего. Тем не менее, для этого пояса существуют перспективы развития целлюлозных производств на альтернативных, в то числе недревесных

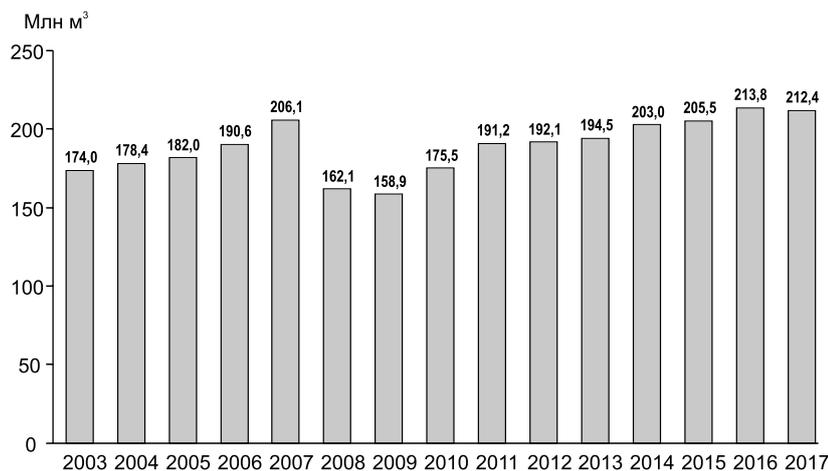


Рис. 8. Производство древесины в России в динамике (м³).

видах сырья, таких как сельскохозяйственные отходы, тростник, возможные в степных и полупустынных районах орошаемые плантации быстрорастущих видов, дающих большую биомассу (бамбук, тополь, недревесные целлюлозно-ориентированные плантации), с ориентацией как на ограниченный внутренний рынок, так и на экспорт в прилегающие страны Центральной Азии.

Завершим исследование анализом графиков, демонстрирующих динамику основных статистических показателей, характеризующих состояние целлюлозно-бумажной отрасли. На рисунке 5 представлен график, показывающий производство древесины в России в динамике с 2003 по 2017 г. (в м³).

График на рисунке 8 свидетельствует о существенной стабильности производства деловой и дровяной древесины в России в рыночных условиях, несмотря на многочисленные внешние шоки. Также видно отсутствие провалов по данному показателю по итогам мер по ограничению экспорта древесного сырья, принимавшихся Правительством РФ по Указу Президента в конце 2000-х гг. (мораторий на повышение экспортных пошлин на лес-кругляк был наложен в связи со значительным падением экспорта древесины на фоне финансового кризиса и действовал до 2011 г. [13]). Это связано в числе прочего с изобретением бизнесом теневых схем и обходных мер. В целом, искусственные ограничения на

экспорт сырья из России, особенно в отношении такого возобновляемого, представленного в избытке в России сырья, как лесные ресурсы, представляются нецелесообразными в долгосрочной перспективе для ведения бизнеса на разных уровнях [20].

Еще один заслуживающий внимания показатель – удельного тоннаж целлюлозы на один кубометр древесины национального ЛПК России, США и Канады в динамике с 2003 по 2017 г. (см. рис. 9).

Ресурсно-сырьевая специализация, низкий уровень передела очевиден из графика, представленного на рисунке 9, как и очевидна его стабильность, неизменяемость под действием даже самых сильных внешних шоков (мировой финансово-экономический кризис), нечувствительность к мелким шокам (попытки ограничить экспорт круглого леса из страны [13]; ограничения на заготовку древесины населением для собственных нужд; новая лесная стратегия [31]; смены руководства Рослесхоза, Минприроды и др.).

Сама по себе сырьевая специализация не является минусом и препятствием для высокого уровня жизни и высокого уровня развития страны в целом, что в течение многих десятилетий своим примером демонстрируют такие страны, как Канада, Австралия и Новая Зеландия [23].

Выводы. Проведенное исследование существующего положения и тенденций

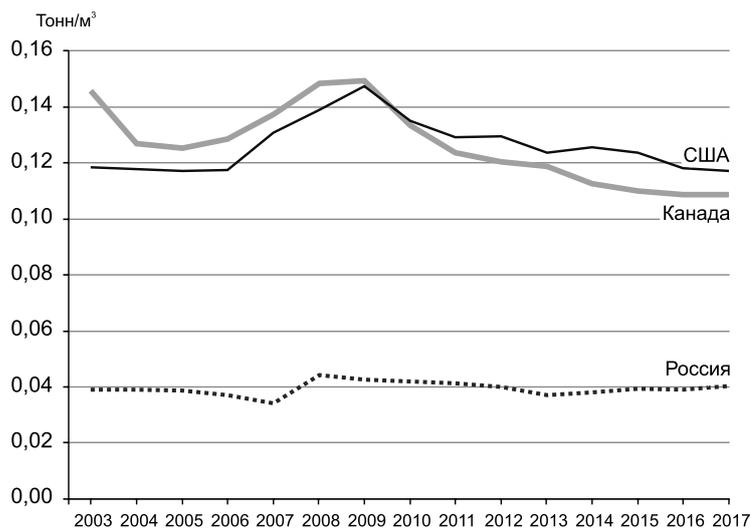


Рис. 9. Удельный тоннаж целлюлозы на один кубометр древесины ЛПК России, США и Канады.

развития целлюлозно-бумажной промышленности Европейской России с использованием инструментов текст-майнинга позволило получить следующие основные выводы:

Лесная промышленность России – это отрасль, достаточно укорененная в российском и мировом разделении труда, устойчивая, во многом за счет своей технологической и организационной примитивности, обеспечивающая гарантированный доход для занятого в нем, преимущественно сельского, населения. В отличие от высокотехнологичных, сложных по организации отраслей [3] лесная промышленность характеризуется устойчивостью к внешним шокам.

Развитие отрасли в среднесрочной перспективе, по мнению авторов, обеспечат максимальное упрощение регулирования, создание наиболее благоприятных условий для импорта всех без дискриминации средств производства и экспорта всех видов продукции, включая сырье и даже продажу леса на корню иностранным организациям, долгосрочную аренду леса любыми юридическими лицами, а также введение института собственности на земли лесного фонда для физических лиц.

В долгосрочной перспективе будущее отрасли характеризуется глубокой неопределенностью в связи с технологическим фактором: прогресс технологий и материалов,

изменение контуров мировой экономики, сдвиги ценностных установок под воздействием технологических факторов могут сделать целые классы лесной продукции, большую часть лесного сырья, невостребованными. В ожидании этих весьма вероятных «тектонических» изменений задачей государства должно быть получение максимально возможной финансовой и социальной выгоды от эксплуатации лесного фонда страны без ущерба для его долгосрочного экосистемного воспроизводственного потенциала там, где этот воспроизводственный потенциал связан с системными экологическими и социальными рисками.

Проведенный анализ соотношения различных видов продукции ЛПК и ЦБП страны, сопоставления этих статистических данных с другими фактами функционирования отрасли, позволяет сделать однозначный вывод о том, что ЛПК и ЦБП в своей сырьевой специализации стабильны, не поддаются трансформации и должны быть приняты и поддерживаемы лицами, принимающими решения, именно в такой конфигурации, для максимизации финансовых и социальных эффектов функционирования отрасли. В целом отрасли требуется либерализация и стабилизация (путем мораториев на изменения в политике), в первую очередь таможенной политики и отношений в части

лесопокрытых земель, а также экологического регулирования. Это позволит сделать неизбежное продолжение сжатия освоенного пространства сельских территорий и постепенное угасание традиционных экономических активностей, связанных с лесными ресурсами, социально менее болезненными и сделать отрасль более выгодной для государства в целом.

В условиях запаздывания и неполноты официальных статистических данных, а также нехватки доверия к ним, важным

является применение новых методов обработки больших данных в проведении отраслевых исследований, в первую очередь текст-майнинга, контент-анализа и сентимент-анализа документных коллекций. Выводы, полученные в результате таких исследований, будут обоснованными и объективными и смогут использоваться для принятия своевременных и взвешенных управленческих решений в лесопромышленном комплексе и целлюлозно-бумажной промышленности.

СПИСОК ЛИТЕРАТУРЫ

1. *Аверкиева К.В.* Симбиоз сельского и лесного хозяйства на староосвоенной периферии Нечерноземья: опыт Тарногского района Вологодской области // Крестьяноведение. 2017. Т. 2. № 4. С. 86–106.
2. *Александрова А.Ю.* География туризма: теоретические основания и пути развития // География и туризм. Сб. статей. Вып. 1. Пермь, 2005. С. 3–16.
3. *Бабикова А.В., Ханина А.В.* Государственное регулирование развития высокотехнологичных отраслей промышленности в условиях модернизации экономики // Проблемы экономики и менеджмента. 2015. № 1 (41). С. 28–34.
4. *Закруткин В.Е., Иваник В.М., Гибков Е.В.* Эколого-географический анализ рисков реструктуризации угольной промышленности в Восточном Донбассе // Изв. РАН. Сер. геогр. 2010. № 5. С. 94–102.
5. *Златин Д.А.* Лесной комплекс России в условиях мирового финансового кризиса // Лесной вестник. 2011. № 1. С. 101–102.
6. *Кузьминов И.Ф., Лобанова П.А., Логинова И.В., Бахтин П.Д.* Система научно-технологического прогнозирования и стратегического планирования в лесном комплексе Российской Федерации // Труды С.-Петербургского науч.-исслед. ин-та лесного хозяйства. 2020. № 1. С. 54–73.
7. *Кузьминов И.Ф., Логинова И.В., Лобанова П.А.* Перспективы использования технологий анализа больших данных для стратегической аналитики агропромышленного комплекса // Сахарная свекла. 2018. № 9. С. 2–7.
8. *Кузьминов И.Ф., Нефедова Т. Г.* Концентрация производства в агропромышленном и лесопромышленном комплексах и поляризация пространства России // Вестн. РАН. 2012. № 1. С. 76–90.
9. *Куричев Н.К.* Новая экономическая география: взгляд экономико-географа // Региональные исследования. 2011. № 4. С. 3–16.
10. *Маркелов К.А., Усманов Р.Х., Головин В.Г.* Черноморско-Каспийский регион: от геополитики к геоэкономике транспортных систем // Каспийский регион: политика, экономика, культура. 2019. № 1. С. 74–88.
11. *Носков В.А.* Доступность древесных ресурсов // Корпоративное управление и инновационное развитие экономики Севера. 2013. № 1. С. 195–211.
12. *Подгорнев П.В.* Структурные и географические особенности фармацевтической промышленности Западной Европы в начале XXI в // Региональные исследования. 2015. № 3. С. 57–64.
13. *Пойкер М.Б.* Торговая политика России в отношении древесины и изделий из нее // Росс. внешнеэкономический вестн. 2013. № 9. С. 41–78.
14. *Сокольский В.М.* Отрасль промышленности как объект экономико-географического исследования (на примере химической индустрии) // Изв. РАН. Сер. геогр. 2011. № 2. С. 25–39.
15. *Социально-экономическая география: понятия и термины. Словарь-справочник.* Отв. ред. А.П. Горкин. Смоленск: Ойкумена, 2013. 329 с.
16. *Тархов С.А.* География транспорта. // Социально-экономическая география в России / под общ. ред. П.Я. Бакланова и В.Е. Шувалова. Владивосток: Дальнаука, 2016. С. 61–69.
17. *Хохлов А.В.* Мировой лесопромышленный комплекс на современном этапе // География мирового развития. Вып. 1. М., 2009. С. 325–347.
18. *Шувалов В.Е.* Районирование в российской социально-экономической географии: современное состояние и направления развития // Региональные исследования. 2015. № 3. С. 19–29.
19. *Vakhtin P.D., Khabirova E., Kuzminov I., Thurner T.* The future of food production – a text-mining approach // Technology Analysis & Strategic Management. 2020. № 32 (5). P. 516–528.
20. *Blondeel M., Van de Graaf T.* Toward a global coal mining moratorium? A comparative analysis of coal mining policies in the USA, China, India and Australia // Climatic change. 2018. № 150 (1–2). P. 89–101.
21. *Doyle Jr J.J., Samphantharak K.* \$2.00 Gas! Studying the effects of a gas tax moratorium // Journal of public economics. 2008. № 92 (3–4). P. 869–884.

22. He W., Zha S., Li L. Social media competitive analysis and text mining: A case study in the pizza industry // *International journal of information management*. 2013. № 33 (3). P. 464–472.
23. Keay I. Immunity from the resource curse? The long run impact of commodity price volatility: evidence from Canada, 1900–2005 // *Cliometrica*. 2015. № 9 (3). P. 333–358.
24. Kirilenko A.P., Stepchenkova S. Tourism research from its inception to present day: Subject area, geography, and gender distributions // *PloS one*. 2018. № 13 (11). Article ID e0206820.
25. Kuzminov I. Geographical conditions for the development of the Russian timber industry in a market economy // *Regional Research of Russia*. 2013. № 3 (1). P. 1–11.
26. Merald E. et al. Changing ideas in forestry: A comparison of concepts in Swedish and American forestry journals during the early twentieth and twenty-first centuries // *Ambio*. 2016. № 45 (2). P. 74–86.
27. В Гватемале арестован владелец Неманского целлюлозно-бумажного комбината. Лесная индустрия. [Электр. ресурс]. URL: https://www.lesprom.com/ru/news/B_Гватемале_арестован_владелец_Неманского_целлюлозно-бумажного_комбината_65790/ (дата обращения 07.08.2020).
28. В Европе продолжается падение цен на целлюлозу. Леспроминфо. Журнал профессионалов ЛПК. [Электр. ресурс]. URL: <https://lesprominform.ru/news.html?id=11338> (дата обращения 07.08.2020).
29. Кузьминов Илья Филиппович. Научная деятельность. [Электр. ресурс]. URL: https://www.hse.ru/staff/ikuzminov#_tab2 (дата обращения 07.08.2020).
30. Неманский ЦБК не может выпускать продукцию с низкой добавленной стоимостью. Лесная индустрия. [Электр. ресурс]. URL: https://www.lesindustry.ru/issues/Li_n2/Nemanskiy_TSBK_ne_mozhet_vipuskat_produktsiyu_s_nizkoy_dobavlennoy_stoimostyu_767/ (дата обращения 07.08.2020).
31. Основы государственной политики в области использования, охраны, защиты и воспроизводства лесов в Российской Федерации на период до 2030 года. [Электр. ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_152506/ (дата обращения 07.08.2020).
32. Снова о строительстве ЦБК в Троицко-Печорском районе. Союз предприятий печатной индустрии (ГИПП). [Электр. ресурс]. URL: <https://gipp.ru/news/poligrafiya-rynok-bumagi/snova-ostroitelstve-tsbk-v-troitsko-pechorskom-rayone/> (дата обращения 07.08.2020).

Статья поступила в редакцию 11 августа 2020 г.

Статья принята к публикации 29 марта 2021 г.

Сведения об авторах

Кузьминов Илья Филиппович – кандидат географических наук, ведущий научный сотрудник, директор центра стратегической аналитики и больших данных Института статистических исследований и экономики знаний Национального исследовательского университета «Высшая школа экономики», г. Москва.

Лобанова Полина Александровна – эксперт отдела информационно-аналитических систем центра стратегической аналитики и больших данных Института статистических исследований и экономики знаний Национального исследовательского университета «Высшая школа экономики», г. Москва.

Для цитирования:

Кузьминов И.Ф., Лобанова П.А. Использование текст-майнинга в экономико-географическом отраслевом анализе целлюлозно-бумажной промышленности Европейской России // *Региональные исследования*. 2021. № 1. С. 18–33.

DOI: 10.5922/1994-5280-2021-1-2.

Text mining for economic geographical sectoral analysis of the pulp and paper industry in European Russia

I.F. Kuzminov*, P.A. Lobanova**

National Research University Higher School of Economics, Moscow, Russia

* e-mail: ikuzminov@hse.ru

** e-mail: plobanova@hse.ru

The authors show the need and some existing opportunities for analysis of non-traditional data sources to obtain a complete and more relevant picture of industries spatial development. The research methodology includes the use of text mining for economic and geographical studies. The relevance of the research is determined by insufficient completeness of official statistical data, cheapening

of relevant information processing technologies and abundance of large text data sources in open access. The article discusses the role of the pulp and paper industry (as a key part of the timber industry) in economic and spatial development of modern Russia. The authors identify main trends in the economic and spatial development of the pulp and paper industry of European Russia, draw the conclusions on the expected industry trends and give recommendations for strategic management decisions to respond to industry challenges. The authors claim that the industry needs liberalization and stabilization, primarily through moratoriums on policy changes. The role of the use of big data, and in particular of text mining in economic and geographical research for reasonable and objective conclusions formation that can be used to make timely and balanced management decisions in the timber industry and the pulp and paper industry, is emphasized.

Keywords: Big Data, text-mining, pulp and paper industry, European Russia, pulp and paper enterprises, timber industry, pulp, paper and wood production.

REFERENCES

1. Averkieva K.V. Symbiosis of agriculture and forestry on the early-developed periphery of the Non-Black Earth Region: The case of the Tarnogsky district of the Vologda Region. *Krest'yanovedenie*, 2017, vol. 2, no. 4., pp. 86-106 (In Russ.).
2. Aleksandrova A.Yu. Geography of tourism: theoretical foundations and ways of development. *Geografiya i turizm*, 2005, pp. 3-16. (In Russ.).
3. Babikova A.V., Hanina A.V. State regulation of the development of high-tech industries in the modernization of the economy. *Problemy ekonomiki i menedzhmenta*, 2015, no. 1 (41), pp. 28-34. (In Russ.).
4. Zakrutkin V.E., Ivanik V.M., Gibkov E.V. Ecology-Geographical Analyses of Coal Industry Risks Restructure in Eastern Donbass. *Izvestiya Rossijskoj akademii nauk. Seriya geograficheskaya*, 2010, no. 5, pp. 94-102. (In Russ.).
5. Zlatin D.A. Forestry complex in Russia in the context of the global financial crisis. *Lesnoj vestnik/ Forestry bulletin*, 2011, no. 1, pp. 101-102. (In Russ.).
6. Kuzminov I.F., Lobanova P.A., Loginova I.V., Bahtin P.D. The system of scientific and technological forecasting and strategic planning in the forest complex of the Russian Federation. *Trudy Sankt-Peterburgskogo nauchno-issledovatel'skogo instituta lesnogo hozyajstva*, 2020, no.1, pp. 54-73. (In Russ.).
7. Kuzminov I.F., Loginova I.V., Lobanova P.A. Prospects for using big data technologies for strategic analysis of agro-industrial complex. *Saharnaya svekla*, 2018, no. 9, pp. 2-7. (In Russ.).
8. Kuzminov I.F., Nefedova T.G. Concentration of production in the agro-industrial complex and timber industry and polarization in Russia. *Vestnik Rossijskoj akademii nauk*, 2012, no. 1, pp. 76-90. (In Russ.).
9. Kurichev N.K. New economic geography: the view of an economic geographer. *Regional'nye issledovaniya*, 2011, no. 4, pp. 3-16. (In Russ.).
10. Markelov K.A., Usmanov R.H., Golovin V.G. Black Sea-Caspian Region: From Geopolitics to Geoeconomics of Transport Systems. *Kaspijskij region: politika, ekonomika, kul'tura*, 2019, no. 1, pp. 74-88. (In Russ.).
11. Noskov V.A. Availability of wood resources. *Korporativnoe upravlenie i innovacionnoe razvitie ekonomiki Severa*, 2013, no. 1, pp. 195-211. (In Russ.).
12. Podgornev P.V. Structural and geographical features of the pharmaceutical industry in Western Europe at the beginning of the XXI century. *Regional'nye issledovaniya*, 2015, no. 3, pp. 57-64. (In Russ.).
13. Pojker M.B. Trade policy of Russia in relation to wood and wood products. *Rossijskij vnesheekonomicheskij vestnik*, 2013, vol. 2013, no. 9, pp. 41-78. (In Russ.).
14. Sokol'skij V.M. Industry as an object of economic and geographical research (on the example of the chemical industry). *Izvestiya Rossijskoj akademii nauk. Seriya geograficheskaya*, 2011, no. 2, pp. 25-39. (In Russ.).
15. Social'no-ekonomicheskaya geografiya: ponyatiya i terminy. Slovar'-spravochnik [Human geography: concepts and terms. Encyclopedic dictionary]. A.P. Gorkin, ed. Smolensk: Ojkumena, 2013. 329 p. (In Russ.).
16. Tarhov S. A. 1.4.5. Transport geography. In: *Social'no-ekonomicheskaya geografiya v Rossii [Socio-economic geography in Russia]*. Vladivostok: Dalnauka, 2016, pp. 61-69. (In Russ.).
17. Hohlov A. V. Mirovoj. The world timber industry at the present stage. *Geografiya mirovogo razvitiya*, 2009, pp. 325-347. (In Russ.).
18. Shuvalov V. E. Regionalization in Russian socio-economic geography: development directions. *Regional'nye issledovaniya*, 2015, no. 3, pp. 19-29. (In Russ.).
19. Bakhtin P.D., Khabirova E., Kuzminov I., Thurner T. The future of food production - a text-mining approach. *Technology Analysis & Strategic Management*. 2020, vol. 32, no. 5, pp. 516-528.
20. Blondeel M., Van de Graaf T. Toward a global coal mining moratorium? A comparative analysis of coal mining policies in the USA, China, India and Australia. *Climatic change*, 2018, vol. 150, no. 1-2, pp. 89-101.
21. Doyle Jr J.J., Samphantharak K. \$2.00 Gas! Studying the effects of a gas tax moratorium. *Journal of public economics*, 2008, vol. 92, no. 3-4, pp. 869-884.
22. He W., Zha S., Li L. Social media competitive analysis and text mining: A case study in the pizza industry. *International journal of information management*, 2013, vol. 33, no. 3, pp. 464-472.
23. Keay I. Immunity from the resource curse? The long run impact of commodity price volatility: evidence from Canada, 1900-2005. *Cliometrica*, 2015, vol. 9, no. 3, pp. 333-358.

24. Kirilenko A.P., Stepchenkova S. Tourism research from its inception to present day: Subject area, geography, and gender distributions. *PloS one*, 2018, vol. 13, no. 11, Article ID e0206820.
25. Kuzminov I. Geographical conditions for the development of the Russian timber industry in a market economy. *Regional Research of Russia*. 2013, no. 1, pp. 1-11.
26. Mårald E. et al. Changing ideas in forestry: A comparison of concepts in Swedish and American forestry journals during the early twentieth and twenty-first centuries. *Ambio*, 2016, vol. 45, no. 2, pp. 74-86.
27. *V Gvatemale arestovan vladelec Nemanskogo cellyulozno-bumazhnogo kombinata* [Owner of Neman Pulp and Paper Mill arrested in Guatemala]. *Lesnaya industriya*. URL: https://www.lesprom.com/ru/news/V_Gvatemale_arestovan_vladelec_Nemanskogo_cellyulozno-bumazhnogo_kombinata_65790/ [Accessed 07.08.2020] (In Russ.).
28. *V Evrope prodolzhaetsya padenie cen na cellyulozu* [Pulp prices continue to fall in Europe]. *Lesprominfo. Zhurnal professionalov LPK*. URL: <https://lesprominform.ru/news.html?id=11338> [Accessed 07.08.2020] (In Russ.).
29. *Kuz'minov Il'ya Filippovich. Nauchnaya deyatel'nost'* [Kuzminov Ilya Filippovich. Scientific research.]. URL: https://www.hse.ru/staff/ikuzminov#__tab2 [Accessed 07.08.2020] (In Russ.).
30. *Nemanskij CBK ne mozhet vypuskat' produkciyu s nizkoj dobavlennoj stoimost'yu* [Neman Pulp and Paper Mill cannot produce products with low added value]. *Lesnaya industriya*. URL: https://www.lesindustry.ru/issues/Li_n2/Nemanskiy_TSBK_ne_mozhet_vipuskat_produktsiyu_s_nizkoj_dobavlennoj_stoimostyu_767/ [Accessed 07.08.2020] (In Russ.).
31. *Osnovy gosudarstvennoj politiki v oblasti ispol'zovaniya, ohrany, zashchity i vosproizvodstva lesov v Rossijskoj Federacii na period do 2030 goda* [Fundamentals of state policy in the field of use, protection, and reproduction of forests in the Russian Federation for the period up to 2030]. URL: http://www.consultant.ru/document/cons_doc_LAW_152506/ [Accessed 07.08.2020] (In Russ.).
32. *Snova o stroitel'stve CBK v Troicko-Pechorskome rajone* [Again about the construction of a pulp and paper mill in the Troitsko-Pechora region]. *Soyuz predpriyatij pechatnoj industrii (GIPP)*. URL: <https://gipp.ru/news/poligrafiya-rynok-bumagi/snova-o-stroitelstve-tsbk-v-troitsko-pechorskome-rayone/> [Accessed 07.08.2020] (In Russ.).

Received 11.08.2020

Accepted 29.03.2021