

DOI: 10.17323/2587-814X.2024.2.35.47

Тематическое моделирование и лингвистический анализ текстовых сообщений в социальной сети для информационно-аналитической поддержки логистического бизнеса

Л.А. Борисова ^a 

E-mail: la.borisova@hse.ru

Ю.И. Костюкевич ^b 

E-mail: y.kostyukevich@skoltech.ru

^a Высшая школа бизнеса, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

^b Сколковский институт науки и технологий, Москва, Россия

Аннотация

В современной экономике успех бизнеса во многом определяется способностью компании анализировать предпочтения потребителей, отношение потребителей к продукции компании, а также возможность быстро реагировать на изменяющиеся предпочтения, либо же на негативные тренды. Social listening или социальное прослушивание является технологией анализа разговоров, текстовых сообщений и любого рода упоминаний компании, ее продукции или бренда. В настоящее время осуществлять социальное прослушивание в интересах российских компаний наиболее эффективно путем мониторинга социальных сетей («ВКонтакте» и др.) как крупнейших источников текстовых сообщений миллионов пользователей. Целью настоящей работы является анализ практик использования технологии социального прослушивания, а также общих подходов к использованию социальных сетей отечественными и зарубежными компаниями. На основе разработанного авторами специализированного программного обеспечения был проведен анализ более 50 тыс. новостных сообщений, опубликованных в 2021–2024 гг. различными по уровню и специализации компаниями. Используя

лингвистический анализ корпуса текстовых сообщений для различных компаний и отраслей экономики, были определены наиболее часто встречающиеся слова, проведено тематическое моделирование, а также изучена динамика новостных сообщений и ее связь с внешними факторами.

Ключевые слова: логистика, социальное прослушивание, лингвистический анализ, бизнес, бренд, мониторинг

Цитирование: Борисова Л.А., Костюкевич Ю.И. Тематическое моделирование и лингвистический анализ текстовых сообщений в социальной сети для информационно-аналитической поддержки логистического бизнеса // Бизнес-информатика. 2024. Т. 18. № 2. С. 35–47. DOI: 10.17323/2587-814X.2024.2.35.47

Введение

Четвертая технологическая революция или Индустрия 4.0 [1–3] требует от компаний для сохранения конкурентоспособности существенно пересматривать свои бизнес-процессы, создавать новые продукты, гибко взаимодействовать с потребителями, анализировать и прогнозировать спрос и предпочтения потребителей, а также отношение потребителей к своему бренду или стратегии развития компании. Недавно была определена концепция социального прослушивания (в англоязычной литературе – social listening), как активный процесс обращения внимания, наблюдения, интерпретации и реагирования на различные стимулы через опосредованные, электронные и социальные каналы, (3)и многие компании активно используют метод социального прослушивания для оценки восприятия компании, ее продукции и бренда широким потребителем [4–7]. Так, Помпутиус [8] исследовал применение социального прослушивания организациями здравоохранения, университетами и правительственными организациями.

В работе Спитале и др. авторами [9] исследовалась переписка пользователей в чатах мессенджера Телеграмм для анализа причин неприятия отдельными группами граждан введения в Италии специального пропускного режима во время пандемии COVID-19. В результате были выявлены общие убеждения, свойственные таким гражданам. Пикон и др. [10] использовали подход социального прослушивания на основе анализа социальной сети Reddit для изучения мнений пользователей с симптомами COVID-19 относительно течения болезни, используемой терапии и пр. Тем самым ставилась задача разработки новых методов получения дополнительной необходимой информации, недоступной для получения традиционными спо-

собами. Хабиб и др. [11], анализируя сообщения в социальной сети Твиттер, исследовали, как пандемия COVID-19 изменила предпочтения людей в выборе транспортного средства.

Джанг и др. [12] использовали концепцию социального прослушивания для определения минутной вовлеченности потребителей при просмотре фильмов. Авторы анализировали текстовые комментарии, оставляемые зрителями в процессе просмотра фильма для анализа предпочтений потребителя. В результате была сформулирована общая концепция такого исследования – moment-to-moment synchronicity (MTMS).

В работе Петровой и Трунина [13] проводилось исследование всего корпуса пресс-релизов Банка России с целью определения тональности заявлений о денежно-кредитной политике, определения сигналов о смягчении или ужесточении монетарной политики, а также о влиянии пресс-релизов на ключевые показатели денежного рынка. Аналогичные работы проводились также и для анализа заявлений ФРС США на экономические показатели, Банка Англии и других [14].

Крупные публичные B2C компании, такие как Coca-Cola, Nike и др. активно используют социальное прослушивание для анализа воспринимаемого образа компании, а также для изучения отношения потребителей к деятельности и инициативам компании [15–19]. Толлинен и др. [20] исследовали использование мониторинга социальных сетей для оценки отношения потребителей к компаниям, работающим по принципу B2B.

Особую роль социальное прослушивание играет в логистическом секторе, где позволяет повышать эффективность выстраивания цепочек поставок и их дизайна, аккумулировать оценки потребителей о качестве предоставляемого сервиса, снижать

неопределенность планируемого спроса, решать проблему оптимизации поиска посредников и поставщиков (Галаскевич [21]). В работе Гал-Тзура и др. [22] была описана методология автоматического сбора и анализа новостных сообщений, публикуемых в социальных сетях, имеющих отношение к транспортному сектору для исследования настроения и предпочтений потребителей. В работе Гюнера и др. [23] было проанализировано более 2 миллионов сообщений, опубликованных пользователями в социальных сетях за период с 2011 г. по 2021 г. для оценки качества сервиса на Турецких железных дорогах с позиции потребителя. Авторы уделили особое внимание динамическим изменениям предпочтений потребителей. Джинг и др. [24] проанализировали более 40 тысяч сообщений в основных социальных сетях Китая (Sina Weibo и TikTok), чтобы изучить различия в общественном восприятии транспортных средств различного уровня автономности до и после аварий с их участием.

Баттачаржа и др. [25] анализировали взаимодействие в социальной сети Twitter компаний, занимающихся онлайн-торговлей, с потребителями. Было проанализировано более 200 тысяч коммуникаций, относящихся к вопросам логистики. Результатом явились рекомендации компаниям по базовым принципам выстраивания взаимодействия с потребителями в социальной сети, а также оперативного удовлетворения их приоритетных запросов. Потенциальную роль социальной сети Twitter в практической деятельности логистической компании исследовал также и Чае [26]. Анализ более чем 22 тысяч сообщений с меткой #supplychain позволил выявить ключевые темы, ассоциирующиеся с деятельностью логистической компании. Ахмади и др. [27] проанализировали более 37 миллионов сообщений в сети Twitter и разработали подходы, позволяющие компаниям аккумулировать отзывы потребителей о своем продукте, проводить анализ настроений (тональности) и использовать эту информацию для разработки наиболее эффективной стратегии возвратной логистики. Исследование проводилось на примере телефонов компании Apple. Позже те же авторы [28] расширили свое исследование на изучение возвратной логистики персональных компьютеров.

В предыдущей работе авторов [29, 30] проводилось исследование применения в российской логистической практике социальной сети «ВКонтакте». Анализ более 30 тыс. новостных сообщений, опу-

бликованных в 2014–2019 гг. различными по уровню и специализации участниками логистического рынка, позволил сделать выводы о предпочтениях в принципиальных подходах к использованию социальных сетей различными компаниями, а также изучить адаптацию транспортной отрасли России к пандемии COVID-19.

В настоящей работе продолжено исследование социальной сети «ВКонтакте» с фокусом на новостные сообщения, опубликованных в период 2021–2024 гг., а также, в отличие от нашего прошлого исследования, применяются получившие развитие в последнее время компьютерные методы текстового анализа на основе библиотек NLPT и Gensim для языка программирования Python. В работе проведено автоматическое тематическое моделирование, основанное на Латентном размещении Дирихле (Latent Dirichlet allocation, LDA) с целью определения какие темы присутствуют в наборах выгруженных сообщений, и какие слова характеризуют каждую тему.

1. Программная система для анализа текстовых сообщений в социальной сети «ВКонтакте»

1.1. Использование API социальной сети «ВКонтакте»

Общая архитектура разработанной системы представлена на рисунке 1. Социальная сеть «ВКонтакте» предоставляет пользователю специальные программные инструменты для разработки собственных приложений (application programming interface, API). Прежде всего, пользователю для

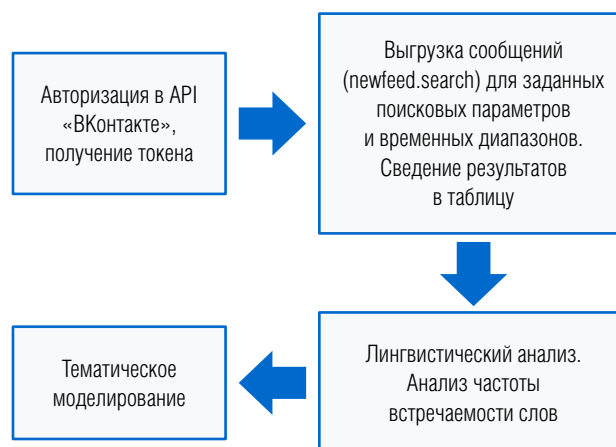


Рис. 1. Архитектура разработанной системы.

получения доступа к интерфейсу необходимо создать внутри сети «ВКонтакте» приложение и получить его идентификатор (MyAppID). Далее необходимо выполнить запрос: https://oauth.vk.com/authorize?client_id=MyAppID&displaypage&redirect_uri=https://oauth.vk.com/blank.html&response_type=token&v=5.199

После чего система предоставит ключ доступа (токен) для использования интерфейса. Для автоматизированной выгрузки и анализа новостных сообщений нами был разработан программный код на языке Python. Для поиска новостных сообщений была использована функция `newfeed.search`, предоставляемая API сети «ВКонтакте». Функция `newfeed.search` принимает следующие аргументы: токен, поисковый запрос в виде строки, временной диапазон для которого осуществлять поиск. Функция позволяет выгружать до 200 сообщений за один поисковый запрос.

1.2. Лингвистический анализ выгруженных текстов

Для заданного хештега нами были выгружены по 200 новостных сообщений для каждой календарной недели, начиная с 2021 г. (более ранние сообщения сеть «ВКонтакте» выгрузить не позволяет). Для каждого новостного сообщения были автоматически определены: текст сообщения, дата публикации, количество просмотров, идентификатор сообщества, его опубликовавшего, и все хэштеги. Результаты были собраны в таблицу.

Текст сообщения обработан следующим образом: все символы переведены в нижний регистр, удалены знаки пунктуации, переноса, табуляции и пр. Далее с использованием библиотеки NLTK (Natural Language Toolkit) текст был преобразован в последовательность отдельных слов или других элементов текста, которые имеют смысловое значение (так называемых токенов). Затем токены были лемматизированы (приведены в словарную форму) пакетом `rumorphy2`, после чего удалены стандартные, часто встречающиеся и не несущие значимой смысловой информации слова (так называемые «стоп-слова»). Мы использовали список из 561 стоп-слова. Обработанные таким образом тексты подвергались дальнейшему анализу, а именно определялось частота встречаемости отдельных слов или хеш-тегов, а также количество публикаций, сделанных каждым сообществом.

1.3. Тематическое моделирование

Для тематического моделирования использовались библиотеки *spacy* (предварительная подготовка текста и лемматизация), *gensim* (вычисления), *pyLDAvis* (визуализация). Оптимальное количество тем выбиралось экспериментально.

2. Текстовый анализ сообщений социальной сети «ВКонтакте»

По данным компании MediaScore по состоянию на 2023 г. в России наиболее популярными социальными сетями являются «ВКонтакте», Telegram, TikTok, «Одноклассники», Instagram и Facebook (две последние принадлежат компании META, признана экстремистской и запрещена в РФ). Сводные данные представлены на *рисунке 2*.

Исходя из данных, представленных на *рисунке 2*, можно заметить, что практически во всех возрастных группах за исключением группы 65+ средний россиянин проводит онлайн более нескольких часов, из них 18% времени тратится на социальные сети. Самой популярной социальной сетью на текущий момент в России является «ВКонтакте» с ежедневным охватом более 43% населения. Таким образом, выбор социальной сети «ВКонтакте» как объекта исследования является обоснованным.

Многочисленные российские компании, в том числе и крупные, давно и активно развивают свои сообщества в интернете, в том числе в социальной сети «ВКонтакте» и мессенджере Telegram. Количество подписчиков в официальных сообществах может превышать несколько миллионов (*таблица 1*).

Нами была проведена выгрузка новостных сообщений, опубликованных в социальной сети «ВКонтакте» за период с 2021 г. по 2024 г. Выгружено по 200 сообщений для каждой календарной недели. Тексты данных сообщений были обработаны и проанализированы.

На *рисунке 3* приведены результаты частотного анализа встречаемости слов в выгруженных текстах по запросам #ВШЭ, #РЖД, #ЖуковВилл и #Fesco. Можно видеть, что образ НИУ ВШЭ в социальной сети «ВКонтакте», ожидаемо, ассоциируется со словами «бизнес», «образование», «проект», «исследование». Образ компании ОАО РЖД ассоциируется со словами «поезд», «станция» и пр., однако частое присутствие слов «обучение», «зарработный», «вакансия» указывает на то, что компания активно использует социальную сеть «ВКонтакте» для поиска и най-

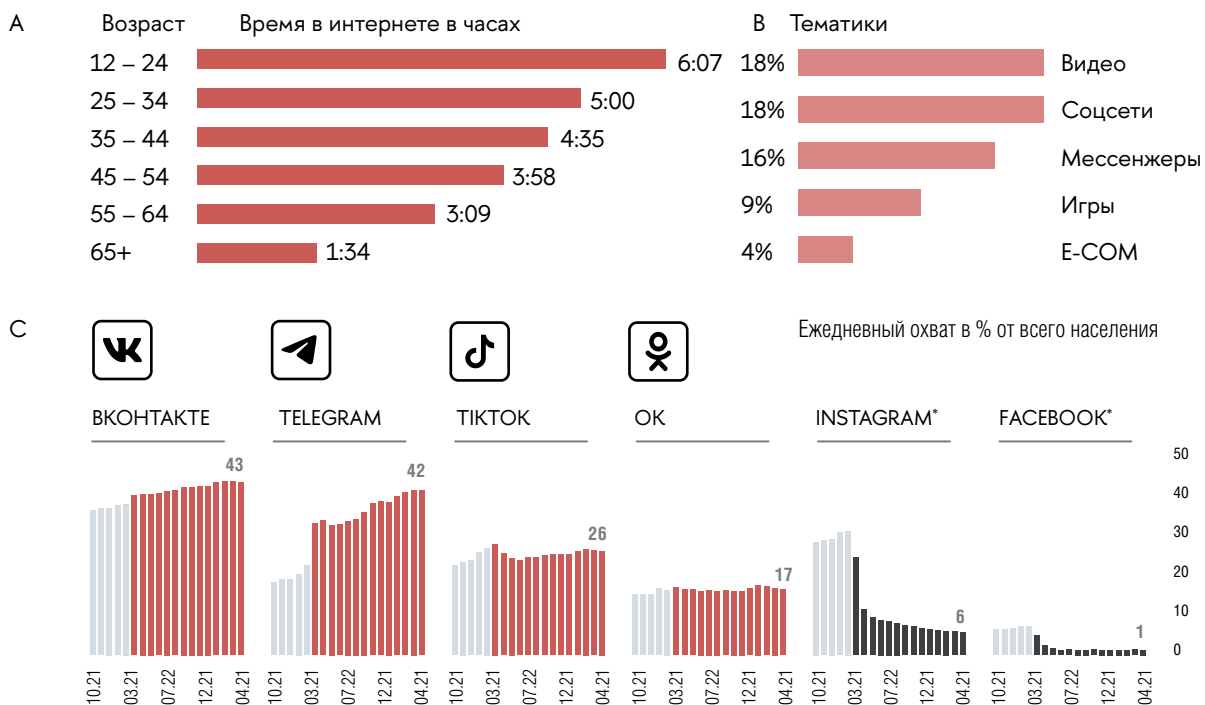


Рис. 2. (А) – среднее время, проводимое пользователем в интернете для разных возрастных групп; (В) – распределение тематик, на которые пользователь тратит свое время; (С) – динамика среднесуточного охвата (в процентах от населения) для наиболее популярных социальных сетей.

Таблица 1.

Официальные сообщества крупных компаний России в интернете

Компания	Годовой оборот, млрд руб. в год	Количество сотрудников, тыс.	Официальная группа в «ВКонтакте» и количество подписчиков, тыс.	Официальный канал в Telegram и количество подписчиков, тыс.
ПАО «Роснефть»	8760	330	rosneft.ru; 114	rosneftofficial; 18,5
ПАО «Сбербанк»		288	sber; 3200	sberbank; 651
АО АвтоВАЗ	160	35	lada; 236	lada_rf; 9
ПАО Газпром	8000	468	gazprom; 158	gazprom; 35
ОАО РЖД	2500	711	rzd_official; 437	telerzd; 40
Яндекс	522	25	yandex; 302	yandex; 151
ПАО Северсталь	728	50	severstal; 54	severstal; 8

ма нового персонала. Для запроса #ВкусВилл часто встречающиеся слова «45000», «месяц», «консультант», «продавец», «вакансия», что тоже говорит о поиске работников. Однако, стоит обратить внимание на группу слов «промокод», «скидка», «заказ», что говорит о том, что эти сообщения имеют целью ре-

кламу и привлечение новых потребителей. По запросу #Fesco наиболее часто употребляемые термины: «Владивосток», «порт», «сервис», «морской», «контейнер», «Китай», «перевозка», «логистика», что также соответствует образу этой компании, как крупного интермодального логистического оператора.

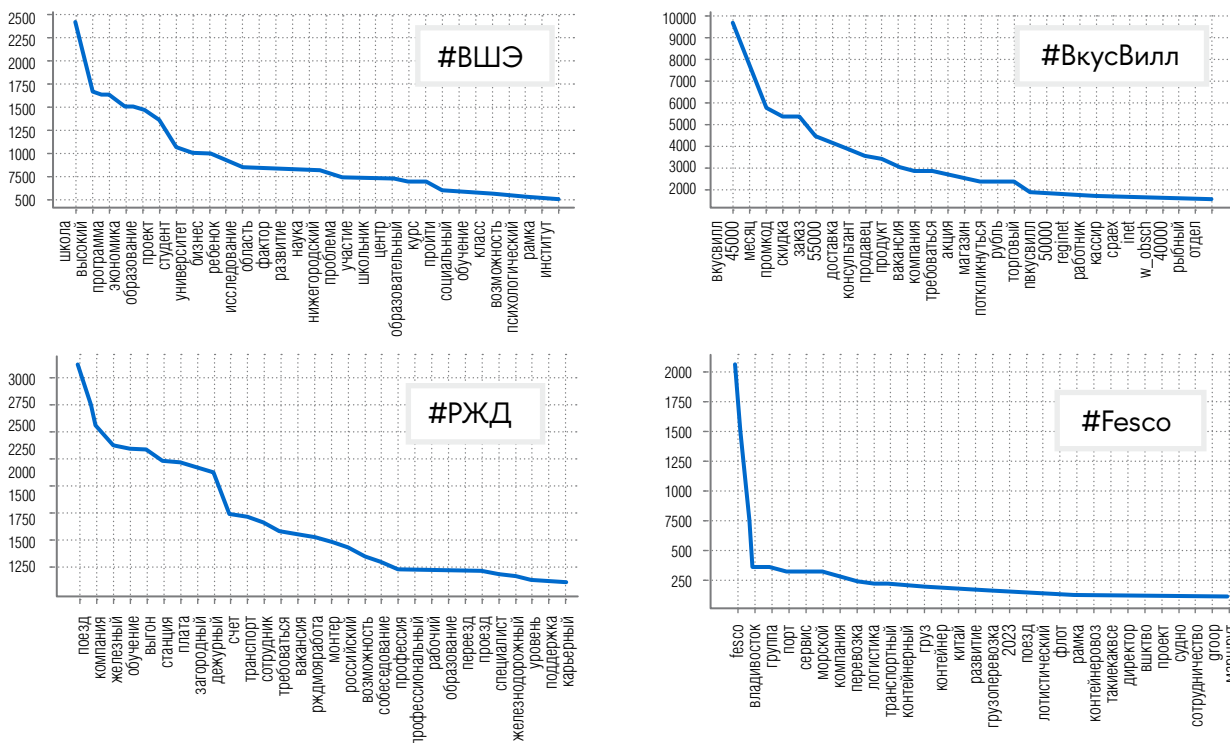


Рис. 3. Анализ частоты встречаемости слов для некоторых компаний, осуществляющих деятельность в различных отраслях экономики.

Для визуального анализа удобно использовать представление методом «облако слов». Пример такой визуализации для текстов, выгруженных по запросу «#Логистика» приведен на рисунке 4.

Можно видеть, что наиболее часто встречающиеся слова в текстах, ассоциированных с темой логистики, это – «вакансия», «заказ», «курьер», «во-

дитель», «заявка» и пр. Можно сделать вывод, что компании и физические лица используют социальные сети в области логистики для размещения вакансий, поиска и предложения услуг доставки, преимущественно автомобильным транспортом.

Также нами проведено тематическое моделирование, основанное на Латентном размещении Ди-

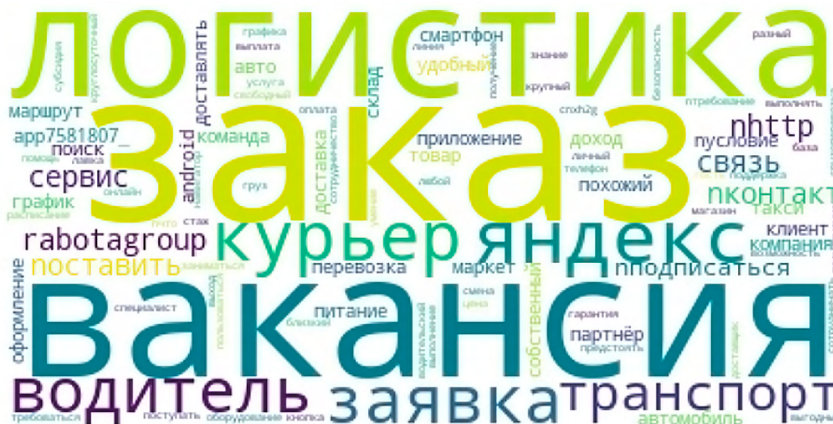


Рис. 4. Визуализация частоты встречаемости слов методом «облако слов» для текстов, выгруженных по запросу «#Логистика».

рихле (Latent Dirichlet allocation, LDA). Тематическое моделирование позволяет автоматически определить, какие темы присутствуют в наборе текстов, и какие слова характеризуют каждую тему. Это весьма существенно для эффективного описания и интерпретации больших объемов текстовой информации. При проведении тематического анализа важным пунктом является выбор количества тем. К сожалению, данный параметр определяется экспериментально путем обучения LDA модели для различного количества тем и выбора того количества, при котором является максимальным значение когерентности модели.

Для корпуса текстов, выгруженного по запросу #Транспорт, нами были обучены LDA модели для значений тем от 3 до 15. При этом максимальное значение когерентности соответствовало 11 темам. Результаты визуализации представлены на рисунке 5. Можно видеть, что каждая тема описывается своим набором слов. При этом, тема 1 соответствует тематике развития городской инфраструктуры, (преимуще-

ственно, в г. Москва), тема 3 – объявлениям о работе курьером, тема 5 – услугам доставки, тема 6 – автобусным перевозкам пассажиров, тема 8 – складским услугам, тема 10 – дорожно-транспортным происшествиям. В результате можно видеть, какие ассоциированные с транспортом темы являются наиболее популярными среди пользователей сети «ВКонтакте».

Социальное прослушивание также позволяет исследовать реагирование компаний, потребителей или даже целых отраслей экономики на внешние вызовы. Действительно, в случае вмешательства некоторого значительного внешнего фактора содержание, тональность новостных сообщений, а также их количество будет заметно меняться. Данный эффект был изучен на примере логистической отрасли. На рисунке 6 представлена помесечная статистика числа новостных сообщений, в которых одновременно присутствовали хэштеги #Транспорт и #Логистика, а также наиболее активные сообщества, отправлявшие сообщения. Для сообществ указано

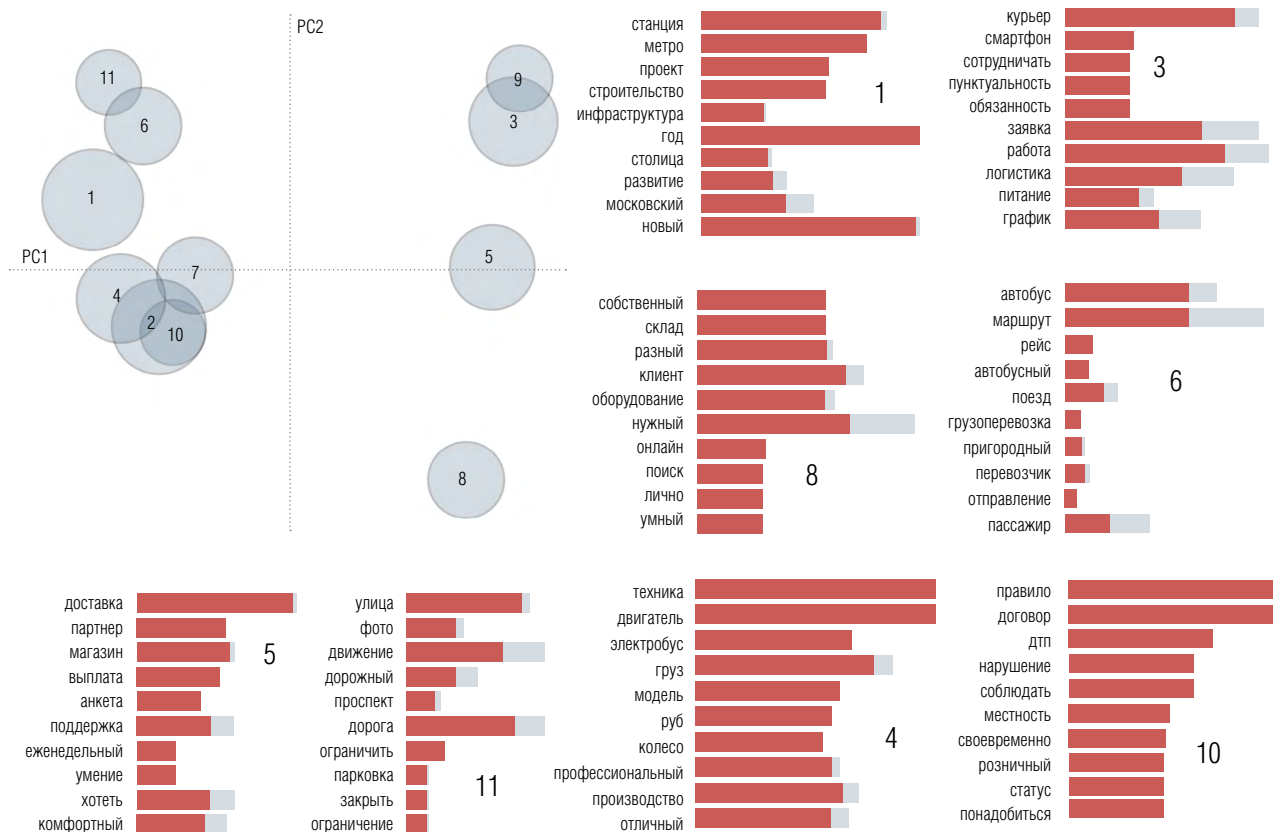
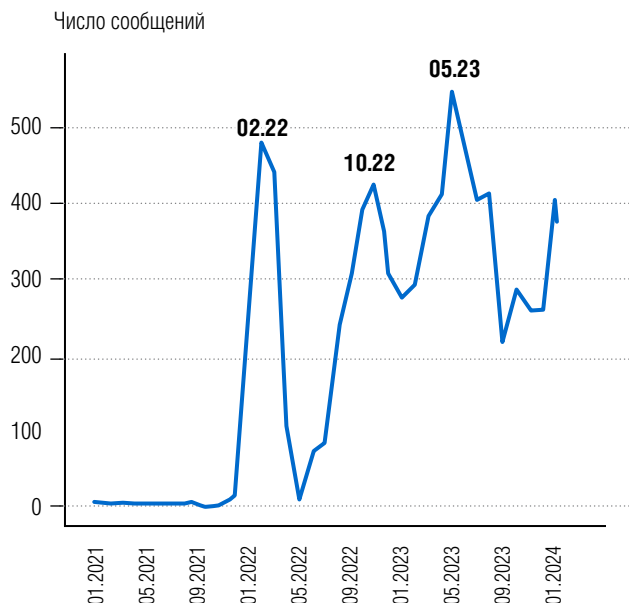


Рис. 5. Тематический анализ корпуса текстов, выгруженного по запросу #Транспорт. Карта расстояний между темами и десятью наиболее характерными словами, соответствующими выбранным темам.

Месячная статистика числа сообщений



Наиболее активные сообщества

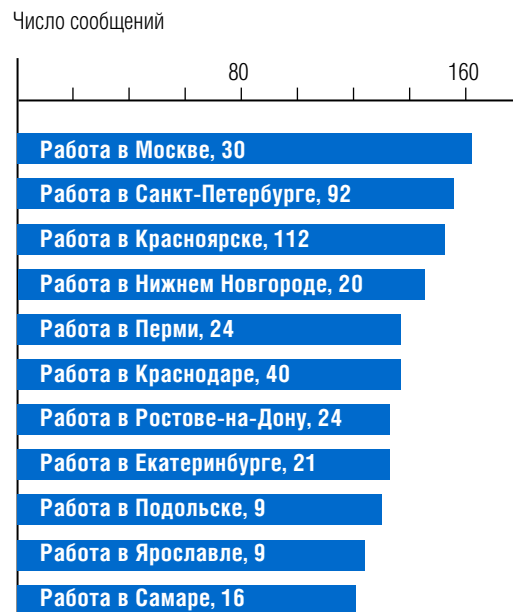


Рис. 6. Месячная статистика числа новостных сообщений, в которых одновременно присутствовали хэштеги #Транспорт и #Логистика, а также наиболее активные сообщества, отправлявшие сообщения.

Для сообществ указано число подписчиков.

Данные сообщества являются участниками сети сообществ Rabota.Group.

число подписчиков. Анализ показал, что подавляющее число новостных сообщений с данными хэштегами являлись объявлениями о предложении вакансии курьеров. Большинство сообщений размещалось в сети сообществ сети Rabota.Group.

На графике числа новостных сообщений в месяц можно видеть характерные пики, соответствующий февралю 2022 г., октябрю 2022 г., маю 2023 г. Данные пики, вероятно, возможно интерпретировать как реакцию отрасли на внешние макроэкономические и социальные явления.

Заключение

Социальное прослушивание (social listening) на основе автоматизированного мониторинга социальных сетей стало важным инструментом, позволяющим компаниям получать оперативную информацию для анализа отношения потребителей к своему бренду, продукции, уровню сервиса и т. п., тем самым облегчая решение одной из наиболее актуальных и сложных задач в бизнесе — моделированию спроса. Действительно, в современном мире

практически каждый пользователь имеет учетную запись в одной или нескольких социальных сетях и тратит значительную часть своего времени на чтение новостных сообщений или на создание собственного контента. Компании, государственные и некоммерческие организации также имеют свои сообщества в социальных сетях и регулярно публикуют новостные сообщения, имеющие целью информирование пользователей о новых событиях, нововведениях, товарах или услугах. Также социальные сети используются как площадки для рекрутинга, поиска поставщиков, продвижения товаров или услуг. Мировая практика последних лет свидетельствует об активном использовании крупным бизнесом, в т. ч. логистическим (например, в дистрибуции и ритейле, в секторе продовольственных товаров), данного инструмента, обычно предоставляемого специализированными высокобюджетными digital-агентствами, что труднодоступно малому бизнесу. На макроэкономическом уровне социальное прослушивание позволяет изучать реакцию общества и отраслей экономики на внешние факторы.

В работе представлены результаты исследования эффективности социального прослушивания на базе автоматизированного поиска и лингвистического анализа всех публикуемых новостных сообщений в социальной сети «ВКонтакте». Показана возможность быстрого и незатратного определения слов и ключевых обсуждаемых тематик, ассоциированных с конкретным брендом, компанией или отраслью экономики. Этот инструмент особенно актуален для малого бизнеса, как предпочтительная альтернатива требующим дополнительных инвестиций подходам к поиску решений о выходе на определенные рынки на основе более надежного прогнозирования спроса. Кроме того, рост потребности в таких эффективных инструментах изучения и планирования спроса на базе открытой информации в социальных сетях активизируется в логистическом секторе, при этом ожидаемо, что тенденции последних лет в отношении коренной трансформации цепей поставок будут только усиливать этот тренд.

Стоит отметить, однако, что ретроспективный анализ новостных сообщений осложняется тем фактом, что в социальных сетях опубликованные новостные сообщения могут быть по желанию пользователя удалены или изменены. Например, многочисленные западные компании при сворачивании бизнеса в России также закрыли и свои официальные сообщества в социальных сетях, что повлекло за собой удаление и всех опубликованных в них новостных сообщений. В результате данные сообщения в настоящее время не доступны для анализа.

Прогнозируя дальнейшее развитие этого процесса, можно ожидать такого дополнительного эффекта активного внедрения социального прослушивания, как реализация возможности потребителей напрямую, за счет активного высказывания своих предпочтений и пожеланий, стимулировать компании к разработке новых продуктов и услуг или другой корректировке их деятельности. ■

Литература

1. Barreto L., Amaral A., Pereira T. Industry 4.0 implications in logistics: an overview // *Procedia Manufacturing*. 2017. Vol. 13. P. 1245–1252. <https://doi.org/10.1016/j.promfg.2017.09.045>
2. Speranza M.G. Trends in transportation and logistics // *European Journal of Operational Research*. 2018. Vol. 264. No. 3. P. 830–836. <https://doi.org/10.1016/j.ejor.2016.08.032>
3. Winkelhaus S., Grosse E.H. Logistics 4.0: a systematic review towards a new logistics system // *International Journal of Production Research*. 2020. Vol. 58. No. 1. P. 18–43. <https://doi.org/10.1080/00207543.2019.1612964>
4. Westermann A., Forthmann J. Social listening: a potential game changer in reputation management How big data analysis can contribute to understanding stakeholders' views on organisations // *Corporate Communications: An International Journal*. 2021. Vol. 26. No. 1. P. 2–22. <https://doi.org/10.1108/CCIJ-01-2020-0028>
5. Cano-Marin E., Mora-Cantalops M., Sánchez-Alonso S. Twitter as a predictive system: a systematic literature review // *Journal of Business Research*. 2023. Vol. 157. Article 113561. <https://doi.org/10.1016/j.jbusres.2022.113561>
6. Yang J., Xiu P., Sun L., Ying L., Muthu B. Social media data analytics for business decision making system to competitive analysis // *Information Processing & Management*. 2022. Vol. 59. No. 1. Article 102751. <https://doi.org/10.1016/j.ipm.2021.102751>
7. Saura J.R., Ribeiro-Soriano D., Palacios-Marqués D. Evaluating security and privacy issues of social networks based information systems in Industry 4.0 // *Enterprise Information Systems*. 2022. Vol. 16. Nos. 10–11. P. 1694–1710. <https://doi.org/10.1080/17517575.2021.1913765>
8. Pomputius A. Can you hear me now? Social listening as a strategy for understanding user needs // *Medical Reference Services Quarterly*. 2019. Vol. 38. No. 2. P. 181–186. <https://doi.org/10.1080/02763869.2019.1588042>
9. Spitale G., Biller-Andorno N., Germani F. Concerns around opposition to the green pass in Italy: social listening analysis by using a mixed methods approach // *Journal of Medical Internet Research*. 2022. Vol. 24. No. 2. Article e34385. <https://doi.org/10.2196/34385>
10. Picone M., Inoue S., DeFelice C., Naujokas M. F., Sinrod J., Cruz V. A., Stapleton J., Sinrod E., Diebel S.E., Wassman E.R. Social listening as a rapid approach to collecting and analyzing COVID-19 symptoms and disease natural histories reported by large numbers of individuals // *Population Health Management*. 2020. Vol. 23. No. 5. P. 350–360. <https://doi.org/10.1089/pop.2020.0189>
11. Habib M.A., Anik M.A.H. Impacts of COVID-19 on transport modes and mobility behavior: Analysis of public discourse in twitter // *Transportation Research Record*. 2023. Vol. 2677. No. 4. P. 65–78. <https://doi.org/10.1177/03611981211029926>
12. Zhang Q., Wang W., Chen Y. Frontiers: In-consumption social listening with moment-to-moment unstructured data: The case of movie appreciation and live comments // *Marketing Science*. 2020. Vol. 39. No. 2. P. 285–295. <https://doi.org/10.1287/mksc.2019.1215>
13. Петрова Д.А., Трунин П.В. Анализ влияния пресс-релизов ЦБ РФ на показатели денежного рынка // *Бизнес-информатика*. 2021. Т. 15. № 3. С. 24–34. <https://doi.org/10.17323/2587-814X.2021.3.24.34>

14. Hansen S., McMahon M. Shocking language: Understanding the macroeconomic effects of central bank communication // *Journal of International Economics*. 2016. Vol. 99. P. 114–133. <https://doi.org/10.1016/j.jinteco.2015.12.008>
15. Liu Y., Lopez R.A. The impact of social media conversations on consumer brand choices // *Marketing Letters*. 2016. Vol. 27. P. 1–13. <https://doi.org/10.1007/s11002-014-9321-2>
16. Austin L.L., Gaither B.M. Examining public response to corporate social initiative types: A quantitative content analysis of Coca-Cola's social media // *Social Marketing Quarterly*. 2016. Vol. 22. No. 4. P. 290–306. <https://doi.org/10.1177/1524500416642441>
17. Reid E., Duffy K. A netnographic sensibility: Developing the netnographic/social listening boundaries // *Journal of Marketing Management*. 2018. Vol. 34. Nos. 3–4. P. 263–286. <https://doi.org/10.1080/0267257X.2018.1450282>
18. De Luca F., Iaia L., Mehmood A., Vrontis D. Can social media improve stakeholder engagement and communication of Sustainable Development Goals? A cross-country analysis // *Technological Forecasting and Social Change*. 2022. Vol. 177. Article 121525. <https://doi.org/10.1016/j.techfore.2022.121525>
19. Vitellaro F., Satta G., Parola F., Buratti N. Social media and CSR communication in European ports: the case of Twitter at the port of Rotterdam // *Maritime Business Review*. 2022. Vol. 7. No. 1. P. 24–48. <https://doi.org/10.1108/MABR-03-2021-0020>
20. Töllinen A., Järvinen J., Karjaluoto H. Social media monitoring in the industrial business to business sector // *World Journal of Social Sciences*. 2012. Vol. 2. No. 4. P. 65–76.
21. Galaskiewicz J. Studying supply chains from a social network perspective // *Journal of Supply Chain Management*. 2011. Vol. 47. No. 1. P. 4–8. <https://doi.org/10.1111/j.1745-493X.2010.03209.x>
22. Gal-Tzur A., Grant-Muller S. M., Kuffik T., Minkov E., Nocera S., Shoor I. The potential of social media in delivering transport policy goals // *Transport Policy*. 2014. Vol. 32. No. P. 115–123. <https://doi.org/10.1016/j.tranpol.2014.01.007>
23. Güner S., Taşkın K., Cebeci H. İ., Aydemir E. Service quality in rail systems: listen to the voice of social media // 26 August 2022, PREPRINT (Version 1) available at Research Square. <https://doi.org/10.21203/rs.3.rs-1980183/v1>
24. Jing P., Cai Y., Wang B., Wang B., Huang J., Jiang C., Yang C. Listen to social media users: Mining Chinese public perception of automated vehicles after crashes // *Transportation Research Part F: Traffic Psychology and Behaviour*. 2023. Vol. 93. P. 248–265. <https://doi.org/10.1016/j.trf.2023.01.018>
25. Bhattacharjya J., Ellison A., Tripathi S. An exploration of logistics-related customer service provision on Twitter: The case of e-retailers // *International Journal of Physical Distribution & Logistics Management*. 2016. Vol. 46. Nos. 6/7. P. 659–680. <https://doi.org/10.1108/IJPDLM-01-2015-0007>
26. Chae B.K. Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research // *International Journal of Production Economics*. 2015. Vol. 165. P. 247–259. <https://doi.org/10.1016/j.ijpe.2014.12.037>
27. Ahmadi S., Shokouhyar S., Shahidzadeh M.H., Papageorgiou E.I. The bright side of consumers' opinions of improving reverse logistics decisions: a social media analytic framework // *International Journal of Logistics Research and Applications*. 2022. Vol. 25. No. 6. P. 977–1010. <https://doi.org/10.1080/13675567.2020.1846693>
28. Ahmadi S., Shokouhyar S., Amerioun M., Tabrizi N.S. A social media analytics-based approach to customer-centric reverse logistics management of electronic devices: A case study on notebooks // *Journal of Retailing and Consumer Services*. 2024. Vol. 76. Article 103540. <https://doi.org/10.1016/j.jretconser.2023.103540>
29. Борисова Л.А., Костюкевич Ю.И. Цифровизация логистики: какова роль социальных сетей? // *Логистика и управление цепями поставок*. 2020. № 3. С. 44–50.
30. Борисова Л.А. Логистика – евразийский мост // *Мат-лы XVI Международ. науч.-практ. конф. (28 апреля – 01 мая 2021 г., г. Красноярск, г. Енисейск)*. Красноярск: Красноярский государственный университет. 2021. С. 15–18.

Об авторах

Борисова Людмила Андреевна

к.э.н.;

доцент, департамент операционного менеджмента и логистики, Высшая школа бизнеса, Национальный исследовательский университет «Высшая школа экономики», Россия, 119049, Москва, улица Шаболовка, д. 26–28;

E-mail: la.borisova@hse.ru

ORCID: 0000-0003-0691-4519

Костюкевич Юрий Иродионович

д.х.н.;

доцент, Сколковский институт науки и технологий, Россия, 121205, Москва, Территория Инновационного центра «Сколково», Большой бульвар, д. 30, стр. 1;

E-mail: y.kostyukevich@skoltech.ru

ORCID: 0000-0002-1955-9336

Thematic modeling and linguistic analysis of text messages from a social network for information and analytical support of logistics business

Ludmila A. Borisova^a

E-mail: la.borisova@hse.ru

Yury I. Kostyukevich^b

E-mail: y.kostyukevich@skoltech.ru

^a Graduate School of Business, HSE University, Moscow, Russia

^b Skolkovo Institute of Science and Technology, Moscow, Russia

Abstract

In the modern economy, the success of a business is largely determined by the company's ability to analyze consumer preferences, consumer attitudes towards the company's products, as well as the ability to quickly respond to changing preferences or negative trends. Social listening is a technology for analyzing conversations, text messages and any kind of mention of a company, its products or brand. Currently, it is most effective to carry out social listening by monitoring social networks (VKontakte, etc.), which are the largest sources of text messages from millions of users. The purpose of this work is to analyze the practices of using social listening technology, as well as common approaches to the use of social networks by domestic and foreign companies. Based on the specialized software developed by the authors, an analysis of more than 50 000 news reports published in 2021–2024 was carried out on companies of different levels and specialization. Using linguistic analysis of the corpus of text messages for various companies and sectors of the economy, the most common words were identified, thematic modeling was carried out, and the dynamics of news reports and their relationship with external factors were studied.

Keywords: logistics, social listening, linguistic analysis, business, brand, monitoring

Citation: Borisova L.A., Kostyukevich Y.I. (2024) Thematic modeling and linguistic analysis of text messages from a social network for information and analytical support of logistics business. *Business Informatics*, vol. 18, no. 2, pp. 35–47. DOI: 10.17323/2587-814X.2024.2.35.47

References

1. Barreto L., Amaral A., Pereira T. (2017) Industry 4.0 implications in logistics: an overview. *Procedia Manufacturing*, vol. 13, pp. 1245–1252. <https://doi.org/10.1016/j.promfg.2017.09.045>
2. Speranza M.G. (2018) Trends in transportation and logistics. *European Journal of Operational Research*, vol. 264, no. 3, pp. 830–836. <https://doi.org/10.1016/j.ejor.2016.08.032>

3. Winkelhaus S., Grosse E.H. (2020) Logistics 4.0: A systematic review towards a new logistics system. *International Journal of Production Research*, vol. 58, no. 1, pp. 18–43. <https://doi.org/10.1080/00207543.2019.1612964>
4. Westermann A., Forthmann J. (2021) Social listening: a potential game changer in reputation management How big data analysis can contribute to understanding stakeholders' views on organisations. *Corporate Communications: An International Journal*, vol. 26, no. 1, pp. 2–22. <https://doi.org/10.1108/CCIJ-01-2020-0028>
5. Cano-Marin E., Mora-Cantalops M., Sánchez-Alonso S. (2023) Twitter as a predictive system: A systematic literature review. *Journal of Business Research*, vol. 157, article 113561. <https://doi.org/10.1016/j.jbusres.2022.113561>
6. Yang J., Xiu P., Sun L., Ying L., Muthu B. (2022) Social media data analytics for business decision making system to competitive analysis. *Information Processing & Management*, vol. 59, no. 1, article 102751. <https://doi.org/10.1016/j.ipm.2021.102751>
7. Saura J.R., Ribeiro-Soriano D., Palacios-Marqués D. (2022) Evaluating security and privacy issues of social networks based information systems in Industry 4.0. *Enterprise Information Systems*, vol. 16, nos. 10–11, pp. 1694–1710. <https://doi.org/10.1080/17517575.2021.1913765>
8. Pomputius A. (2019) Can you hear me now? Social listening as a strategy for understanding user needs. *Medical Reference Services Quarterly*, vol. 38, no. 2, pp. 181–186. <https://doi.org/10.1080/02763869.2019.1588042>
9. Spitale G., Biller-Andorno N., Germani F. (2022) Concerns around opposition to the green pass in Italy: social listening analysis by using a mixed methods approach. *Journal of Medical Internet Research*, vol. 24, no. 2, article e34385. <https://doi.org/10.2196/34385>
10. Picone M., Inoue S., DeFelice C., Naujokas M. F., Sinrod J., Cruz V. A., Stapleton J., Sinrod E., Diebel S.E., Wassman E.R. (2020) Social listening as a rapid approach to collecting and analyzing COVID-19 symptoms and disease natural histories reported by large numbers of individuals. *Population Health Management*, vol. 23, no. 5, pp. 350–360. <https://doi.org/10.1089/pop.2020.0189>
11. Habib M.A., Anik M.A.H. (2023) Impacts of COVID-19 on transport modes and mobility behavior: Analysis of public discourse in twitter. *Transportation Research Record*, vol. 2677, no. 4, pp. 65–78. <https://doi.org/10.1177/03611981211029926>
12. Zhang Q., Wang W., Chen Y. (2020) Frontiers: In-consumption social listening with moment-to-moment unstructured data: The case of movie appreciation and live comments. *Marketing Science*, vol. 39, no. 2, pp. 285–295. <https://doi.org/10.1287/mksc.2019.1215>
13. Petrova D.A., Trunin P.V. (2021) Analysis of the impact of central bank communications on money market indicators. *Business Informatics*, vol. 15, no. 3, pp. 24–34. <http://doi.org/10.17323/2587-814X.2021.3.24.34>
14. Hansen S., McMahon M. (2016) Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, vol. 99, pp. 114–133. <https://doi.org/10.1016/j.jinteco.2015.12.008>
15. Liu Y., Lopez R.A. (2016) The impact of social media conversations on consumer brand choices. *Marketing Letters*, vol. 27, p. 1–13. <https://doi.org/10.1007/s11002-014-9321-2>
16. Austin L.L., Gaither B.M. (2016) Examining public response to corporate social initiative types: A quantitative content analysis of Coca-Cola's social media. *Social Marketing Quarterly*, vol. 22, no. 4, pp. 290–306. <https://doi.org/10.1177/1524500416642441>
17. Reid E., Duffy K. (2018) A netnographic sensibility: Developing the netnographic/social listening boundaries. *Journal of Marketing Management*, vol. 34, nos. 3–4, pp. 263–286. <https://doi.org/10.1080/0267257X.2018.1450282>
18. De Luca F., Iaia L., Mehmood A., Vrontis D. (2022) Can social media improve stakeholder engagement and communication of Sustainable Development Goals? A cross-country analysis. *Technological Forecasting and Social Change*, vol. 177, article 121525. <https://doi.org/10.1016/j.techfore.2022.121525>
19. Vitellaro F., Satta G., Parola F., Buratti N. (2022) Social Media and CSR Communication in European ports: the case of Twitter at the Port of Rotterdam. *Maritime Business Review*, vol. 7, no. 1, pp. 24–48. <https://doi.org/10.1108/MABR-03-2021-0020>
20. Töllinen A., Järvinen J., Karjaluo H. (2012) Social media monitoring in the industrial business to business sector. *World Journal of Social Sciences*, vol. 2, no. 4, pp. 65–76.
21. Galaskiewicz J. (2011) Studying supply chains from a social network perspective. *Journal of Supply Chain Management*, vol. 47, no. 1, pp. 4–8. <https://doi.org/10.1111/j.1745-493X.2010.03209.x>
22. Gal-Tzur A., Grant-Muller S.M., Kuflik T., Minkov E., Nocera S., Shoor I. (2014) The potential of social media in delivering transport policy goals. *Transport Policy*, vol. 32, pp. 115–123. <https://doi.org/10.1016/j.tranpol.2014.01.007>
23. Güner S., Taşkın K., Cebeci H.İ., Aydemir E. (2022) Service quality in rail systems: listen to the voice of social media. *PREPRINT (Version 1) available at Research Square*. <https://doi.org/10.21203/rs.3.rs-1980183/v1>
24. Jing P., Cai Y., Wang B., Wang B., Huang J., Jiang C., Yang C. (2023) Listen to social media users: Mining Chinese public perception of automated vehicles after crashes. *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 93, pp. 248–265. <https://doi.org/10.1016/j.trf.2023.01.018>
25. Bhattacharjya J., Ellison A., Tripathi S. (2016) An exploration of logistics-related customer service provision on Twitter: The case of e-retailers. *International Journal of Physical Distribution & Logistics Management*, vol. 46, nos. 6/7, pp. 659–680. <https://doi.org/10.1108/IJPDLM-01-2015-0007>
26. Chae B.K. (2015) Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, vol. 165, pp. 247–259. <https://doi.org/10.1016/j.ijpe.2014.12.037>

27. Ahmadi S., Shokouhyar S., Shahidzadeh M.H., Papageorgiou E.I. (2022) The bright side of consumers' opinions of improving reverse logistics decisions: a social media analytic framework. *International Journal of Logistics Research and Applications*, vol. 25, no. 6, pp. 977–1010. <https://doi.org/10.1080/13675567.2020.1846693>
28. Ahmadi S., Shokouhyar S., Amerioun M., Tabrizi N.S. (2024) A social media analytics-based approach to customer-centric reverse logistics management of electronic devices: A case study on notebooks. *Journal of Retailing and Consumer Services*, vol. 76, article 103540. <https://doi.org/10.1016/j.jretconser.2023.103540>
29. Borisova L., Kostyukevich Y. (2020) Digitalization of logistics: what is the role of social networks? *Logistics and Supply Chain Management*, no. 3, pp. 44–50 (in Russian).
30. Borisova L. (2021) Logistics – the Eurasian Bridge. Materials of the XVI International scientific and practical conference (April 28 – May 01, 2021, Krasnoyarsk, Yeniseisk). Krasnoyarsk: Krasnoyarsk State University, pp. 15–18 (in Russian).

About the authors

Ludmila A. Borisova

Cand. Sci. (Econ.);

Associate Professor, Department of Operations Management and Logistics, Graduate School of Business, HSE University, 26–28, Shabolovka St., Moscow 119049, Russia;

E-mail: la.borisova@hse.ru

ORCID: 0000-0003-0691-4519

Yury I. Kostyukevich

Dr. Sci. (Chem.);

Associate Professor, Skolkovo Institute of Science and Technology, the territory of the Skolkovo Innovation Center, Bolshoy Blv., 30, p. 1, Moscow 121205, Russia;

E-mail: y.kostyukevich@skoltech.ru

ORCID: 0000-0002-1955-9336