ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ

УДК: 004.912: 81'33 DOI: 10.18287/2223-9537-2020-10-4-449-462

Компьютерный анализ эмоциональной компоненты научных публикаций на примерах в физике и экономике

В.С. Крылов[1], А.А. Кудрявцев[2], Л.Н. Абдурайимов[1]

[1]Крымский инженерно-педагогический университет имени Февзи Якубова, Симферополь, Россия
[2]Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

Аннотация

Представлены результаты применения системы автоматизированного извлечения и визуализации метаданных эмоциональности текстов научных публикаций в области инновационных исследований в физике и экономике. В физике рассматриваются работы по прямому преобразованию светового излучения в электрическую энергию на основе фотоплазмы - эффекта возникновения разности потенциалов в плазме щелочных металлов. В этой области практически завершён этап научно-исследовательских работ, и происходит переход к опытно-конструкторским работам с последующей оценкой экономической эффективности производства и эксплуатации фотопреобразователей на основе фотоплазмы. Такие метаданные необходимы для разработки и внедрения в практику систем интеллектуального анализа новостных текстовых данных, аналитических публикаций для повышения эффективности прогноза и оценки перспектив развития в разных сферах, например, в финансово-экономической деятельности для прогноза котировок на рынке ценных бумаг. В работе приведены примеры сентимент-анализа и его визуализация экономических научных текстов и текста из области исследований газоразрядной плазмы. Инструментальные методы эмоционального анализа текстов реализованы на языке *R*, который позволяет в короткий срок формировать необходимые пакеты программ анализа текстов не только профессиональным программистам, но и аналитикам.

Ключевые слова: *компьютерный анализ текстов, сентимент-анализ, эмоциональная компонента, физика плазмы, цифровая экономика, R-программирование, искусственный интеллект.*

Цитирование: *Крылов, В.С.* Компьютерный анализ эмоциональной компоненты научных публикаций на примерах в физике и экономике / В.С. Крылов, А.А. Кудрявцев, Л.Н. Абдурайимов // Онтология проектирования. – 2020. – Т.10, №4(38). - С.449-462. – DOI: 10.18287/2223-9537-2020-10-4-449-462.

Введение

Компьютерный анализ текстов позволяет получить новые, не лежащие на поверхности, знания. Научные тексты представлены профессиональным языком, который является специфическим подмножеством естественного базового языка (ЕЯ) общения. Поэтому его синтаксис, грамматика и наборы слов получают эмоциональные оценки с точки зрения базового языка. Эмоционально окрашенные названия характеристик объектов или процессов зачастую специально присваиваются, чтобы подчеркнуть их особенность. Например, в физике элементарных частиц некоторые характеристики получили явно эмоциональные названия: «странность», «очарование», «цвет» и т.д. Похожее происходит и в других разделах физики, в т.ч. и в физике плазмы. Так, первоначально воспринимавшееся эмоционально окрашенным название явления в газоразрядной пылевой плазме – плазменный кристалл, оказалось не метафорой, а реально наблюдаемой упорядоченной структурой.

В экономике и финансах получают вполне эмоционально окрашенные названия показателей состояния, например, рынков ценных бумаг – «голубые фишки», «золотая акция»» и т.п. Математики очень часто используют эстетические эмоциональные оценки в отношении получаемых результатов, такие как: красивое, прозрачное (доказательство, …) и т.д. Поэтому востребована разработка инструментов многостороннего анализа текстов, в том числе, инструменты анализа их эмоциональной компоненты.

Технология сентимент-анализа (СА) широко используется корпорациями, которые владеют брендами для анализа социальных медиа с целью оценки коммерческих результатов. Обзор предлагаемых приложений СА представлен в [1]. Считается, что имеется возможность не только оценить тональность высказываний о бренде, но и получить целый ряд дополнительных инструментов управления социальной аудиторией, интересующейся брендом, установления контактов, обмена информацией, влияния на социальный контент, поиска лидеров мнений социального сообщества и снабжения их информацией для привлечения к продвижению бренда. Подобные системы малопригодны для целей СА научных публикаций, в связи с тем, что они узкоспециализированы и с закрытым кодом. Последнее не позволяет их адаптировать для решения задач интеллектуального анализа текстов научных публикаций. Значительная часть этих систем является платной, а предлагаемый функционал предварительно бесплатного использования малопригоден для анализа текстов на этапах научно-исследовательских работ (НИР) и опытно-конструкторских работ (ОКР) [1, 2].

В тоже время, платформа *R*-языка содержит пакеты программ, которые позволяют собирать алгоритмы анализа текстов, ориентированных на решение поставленных аналитиками конкретных исследовательских задач. В данном случае разработка алгоритмов анализа и их реализация во многом похожа на экспериментальные исследования, например, в физике плазмы газового разряда: установка для исследований собирается из готовых компонент и дополняется нестандартным оборудованием [3].

С появлением широко используемых открытых данных о методах обработки ЕЯ можно легко сравнивать различные доступные наборы инструментов аналитиков, которые позволяют выполнить обработку текстов ЕЯ, в том числе и текстов научных публикаций [4-6].

На платформе *R*-языка предлагается несколько пакетов программ с открытым кодом:

- *textrank* - составление резюме текста;
- *crfsuite* - распознавание сущностей, разбиение на части и моделирование последовательности;
- *BTM* - тематическое моделирование битермов или очень коротких текстов (например, ответы на опрос / данные твиттера);
- *ruimtehol* - нейронные текстовые модели, нейронные модели для категоризации текста, встраивания слов/предложений/документов, рекомендаций по документам, завершения ссылок на объекты и внедрения объектов);
- *udpipe* - общий пакет обработки текстов ЕЯ для токенизации, лемматизации, тегов частей речи, морфологических аннотаций, синтаксического анализа зависимостей, извлечения ключевых слов и потоков;
- *tidytext*: - функция *unnest_tokens*() предварительно переводит текст в упорядоченный *tidy* формат, который позволяет выделить эмоциональную компоненту текста с помощью специальных словарей.

Пакеты программ позволяют получить не только многостороннюю качественную оценку эмоциональной компоненты, но и её количественную характеристику грамматических и синтаксических структур [7, 8]. Например, с помощью разработанных алгоритмов на основе *R*-пакетов программ была изучена тональность в текстах писем крупнейшего в мире инвестора У.Э. Баффетта акционерам за период с 1977 по 2016 год [8, 9].

Цифровая экономика - новый этап развития информационного общества, в котором экономические отношения основываются на новых методах генерирования, обработки, хранения, передачи данных. Здесь возникают проблемы извлечения необходимой для конкретной деятельности информации из информационных потоков очень больших объёмов. Например, пока не решена задача по установлению связи и прогноза влияния эмоционально окрашенных текстов экспертов на котировки ценных бумаг на финансовых рынках [10, 11].

Экономический кризис во многом «обнулил» все прогнозы экономического развития [12]. Глава международного валютного фонда эмоционально оценила перспективы развития экономики в условиях кризиса, в том числе обусловленного и пандемией Covid-19: «экономические прогнозы, перевернулись «с ног на голову» и мировая экономика вместо роста будет сокращаться» [12, 13].

В условиях выхода из кризиса экономики большое значение имеют инновации – научные открытия, изобретения, которые имеют практическое применение и удовлетворяют социальным, экономическим требованиям, которые дают эффект в соответствующих сферах деятельности [14, 15]. Плазменные технологии газового разряда являются одним из важнейших инновационных направлений во многих областях практического применения [3]. Например, инновационные системы генерации электроэнергии на основе плазмы газового разряда (фотоплазмы), которые из стадии НИР переходят в ОКР, могут кардинально изменить глобальные и региональные рынки солнечной энергетики. Причём они позволят отказаться от льготного налогообложения и специальных зелёных тарифов [16]. СА такого инновационного объекта должен включать в себя метаданные как для оценки научной или технической новизны инновационного объекта, так и метаданные оценки эмоциональности его экономико-финансовых свойств [17]. Эти метаданные могут стать важным показателем оценки этапа перехода от НИР к ОКР и прогноза внедрения инновационных разработок.

Цель работы – представить результаты интеллектуального анализа и визуализации компонентов текстов научных публикаций из области экономической теории и инновационных исследований преобразования световой энергии в электрическую с помощью газоразрядной плазмы (фотоплазмы).

1 Автоматизированный анализ естественных профессиональных текстов

Сообщество разработчиков программ для научных исследований, в том числе и текстов, предлагают инструменты извлечения, анализа и визуализации данных для широкого круга исследователей [4, 7, 18]. С их помощью проведено исследование по автоматизированному извлечению и взаимосвязи ключевых терминов из правительственных документов, выпущенных в 2013-2018 годах и связанных с направлением «цифровая экономика» [10]. Исследование было выполнено с помощью графоориентированных методов алгоритма, реализованного в пакете *textrank*. Выбранный алгоритм был протестирован на 13 правительственных документах. В результате анализа каждого текста были построены взвешенные графы семантических связей между ключевыми словами, на основании которых были выделены ключевые термины.

Существующие доступные пакеты программ анализа текстов позволяют решать определённый круг исследовательских задач. Однако в конкретном исследовании всегда возникают проблемы, которые не могут быть решены в представленных пакетах программ. Эти проблемы позволяет решить платформа программирования, на которой можно либо провести модификацию программы для решения исследовательской задачи, либо дополнить программами из других пакетов, не требуя от аналитика высокого уровня квалификации в программировании для разработки и использования инструмента в решении поставленной задачи.

Платформа программирования *R* обладает этими качествами. Она широко используется для работы с самыми разнообразными данными, визуализации данных, анализа текстов и представляет собой наборы разнообразных универсальных инструментов, из которых можно формировать требуемые наборы инструментов анализа. Платформа *R* включает в себя мощные вычислительные и графические возможности получения и визуализации результатов исследований [18].

R-язык является простым и эффективным инструментом для статистического анализа данных, анализа текстов, использования методологии объектно-ориентированного программирования, а также функционального программирования и других парадигм программирования систем анализа данных. Ведущие научно-исследовательские центры и университеты мира, аналитики крупнейших компаний широко используют разработанные пакеты программ на языке *R* для анализа больших данных и реализации крупных информационных проектов анализа текстов [18, 19].

Тексты научных публикаций, монографий, учебных пособий, также как и обычные неструктурированные литературные тексты, содержат много слов и знаков, затрудняющих последующий содержательный анализ и получение метаданных, в том числе об эмоциональной компоненте. Существуют разнообразные алгоритмы извлечения метаданных из текстов, реализованные в различных пакетах программ для решения задач исследования текстов [4, 18, 20]. Можно выделить общий для всех типов текстов алгоритм анализа, который состоит из следующих шагов:

1) извлечение данных,
2) очистка и предобработка,
3) представление, фильтрация и взвешивание,
4) результаты.

Блок-схема типичного алгоритма анализа текста представлена на рисунке 1.

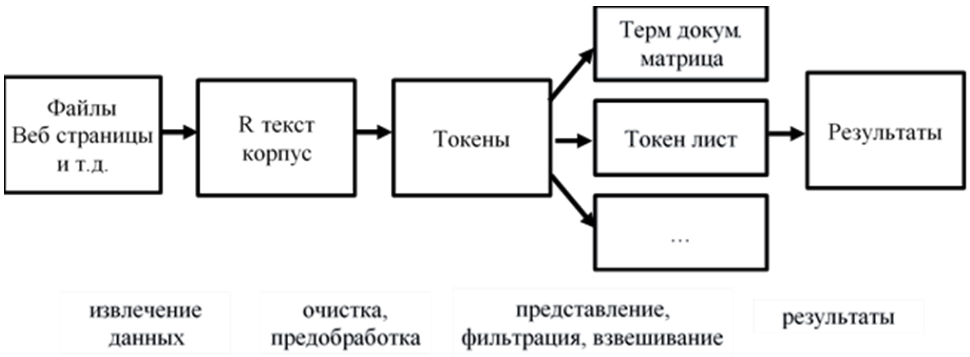


Рисунок 1 - Блок-схема типичного алгоритма анализа текста

Очистку текстов и приведение их к определённому типу представления можно выполнить специализированными пакетами программ. Разработанный на *R*-платформе пакет *tidytext* совместим с рядом других пакетов анализа текстов, позволяет выбирать и объединять в единый инструмент необходимые для решения конкретных задач модули. Этот пакет является основой для широко используемых инструментов компьютерного анализа текстов [4, 5]. В пакете осуществляется очистка текстов и представление их в форме фреймов из отдельных слов, пар слов и целых предложений или абзацев. Представление фреймами даёт возможность эффективно обрабатывать, обобщать и визуализировать характеристики текста, интегрировать обработку ЕЯ в эффективные рабочие процессы анализа.

Следует отметить, что эмоциональная компонента всегда присутствует в любом тексте. Не исключением являются тексты научных публикаций, монографий и документов, напри-

мер, таких как теоретические исследования холодной плазмы или исследования в сфере финансов и экономики. Так, действующий на бирже агент, обозначенный Адамом Смитом как *Homo economicus,* представляет собой холодную, лишённую эмоций расчётливую личность, которая использует исключительно интеллект, а не эмоции для принятия решений. Однако оказалось, что ключевые решения о финансах принимаются, в том числе, структурами мозга, ответственными за эмоции [21]. Иначе говоря, любая система искусственного интеллекта, разработанная для семантического анализа потоков публикаций и документов, в обязательном порядке должна учитывать эмоциональную компоненту их содержания. Особенно ярко эмоциональная компонента оценки научного сообщения по результатам исследования проявляется в обсуждениях на конференциях и форумах в социальных сетях.

Для оценки эмоциональности необходимы специальные словари, или лексиконы. В них каждое слово относится к той или иной эмоции. В представленной работе использовались следующие лексиконы [4, 7]:

- *bing* от Бин Лю и сотрудников,
- *nrc* от Саифа Мухаммеда и Питера Терни,
- *loughran* Лафрана и Макдоналда.

Лексиконы основаны на униграммах (отдельных словах) и содержат много английских слов, которым присваиваются баллы за позитивные/негативные эмоции, а также, например, за такие эмоциональные оценки, как радость, гнев, грусть и т.д. Лексикон *nrc* классифицирует слова в двоичном варианте (да/нет) в категориях: положительной, отрицательной, гнев, ожидание, отвращение, страх, радость, печаль, удивление и доверие. Лексикон *bing* классифицирует слова в двоичном представлении на положительные и отрицательные категории. Лексикон *loughran* ориентирован на финансовую лексику. В нём оценки эмоциональности, в сравнении с другими словарями, менее категоричны, более обтекаемы. Считается, что положительные оценки указывают на позитивные, а отрицательные - на негативные настроения. Эта информация сводится в таблицу набора данных, а пакет *tidytext* предоставляет функцию *get_sentiments*() для получения определённой лексики настроения [4, 8].

Основанные на лексиконах методы определяют общее настроение фрагмента текста, составляют индивидуальные оценки настроения для каждого слова в тексте. Лексиконы эмоциональности были составлены и проверены либо с помощью краудсорсинга, либо с помощью кропотливого труда и проверкой с использованием некоторой комбинации краудсорсинга на основании авторских обзоров впечатлений от ресторанов или кинофильмов, либо данных *Twitter* и т.д. Отсюда следует, что в выводах в оценке эмоциональности следует учитывать отличия профессиональных текстов от материалов, на котором лексиконы были проверены. В настоящее время это единственный способ измерить содержание настроений для слов, общих для лексикона и текста базового языка, например, литературных произведений [4]. Полученные с их помощью результаты следует понимать как оценку восприятия профессионального текста обычным человеком. Для оценки восприятия специалистом необходимо выполнить специальное исследование и работу по каждому профессиональному языку.

СА информационных потоков имеет большой потенциал применения для мониторинговых, аналитических и интеллектуальных систем, систем документооборота и рекламы, специализированных по тематике веб-страниц и т.д. Основанные на лексиконе методы позволяют определить общее настроение фрагмента текста, складывая индивидуальные оценки настроения для каждого слова в тексте [4, 8, 11].

Не каждое слово попадает в специализированный лексикон, потому что многие слова достаточно нейтральны. Методы, основанные исключительно на униграммах, не учитывают квалификаторы перед словом, например, «не хорошо» или «не соответствует действительности». Для многих видов текста не существует устойчивых методов различения сарказма от

негативного текста. В этом случае можно использовать подход с очищенным упорядоченным форматом для того, чтобы начать понимать, какие слова отрицания важны в данном тексте. Размер фрагмента текста, который используется для суммирования оценок настроения по отдельным словам, может повлиять на конечный результат анализа. В целом текст с множеством абзацев часто может иметь положительное или отрицательное к нему отношение, усреднённое примерно до нуля, в то время как текст размером с предложение или размером с абзац часто будет отличаться от общей оценки [4].

Положительное или отрицательное значение слова может зависеть от контекста. Слово «риск» имеет отрицательное значение в большинстве общих контекстов, но может быть более нейтральным для финансовой отчётности. Контекстно-специфические лексиконы настроений, такие как словарь *loughran* Лафрана-Макдональда, даёт возможность решить эту проблему. Предлагаемая финансовая лексика обозначает слова с шестью возможными чувствами [7, 18].

В [8, 22] представлены результаты анализ тональностей в текстах писем У.Э. Баффетта акционерам за период с 1977 по 2016 год. Использовался метод выявления и количественной оценки общего настроения текста писем - определение того, насколько положительным или отрицательным в целом является конкретный текстовый документ. Для этого текстовый документ разбивался на набор отдельных слов, а затем для каждого слова определялось, является ли оно положительным, отрицательным или нейтральным с помощью лексикона оценки эмоциональности *bing*. Вычислялся коэффициент тональности как отношение количества положительных слов минус количество отрицательных слов к общему количеству слов. В целом письма У.Э. Баффета оценивались как положительные. За сорок лет только пять писем получили отрицательную чистую оценку настроения. Как оказалось, эти отрицательные оценки тесно связаны с крупными негативными экономическими событиями.

2 Сентимент-анализ экономических научных текстов

С помощью предлагаемых платформой *R*-пакетов анализа текстов был «собран» алгоритм комплексного анализа текстов. В общую схему (рисунок 1) были включены модули: создание корпуса выбранной предметной области (пакет *tm*), приведение текста к упорядоченному формату (пакеты *tidytext*), визуализация промежуточных и окончательных результатов анализа, построение семантических сетей (пакеты *ggraph* и *igraf*).

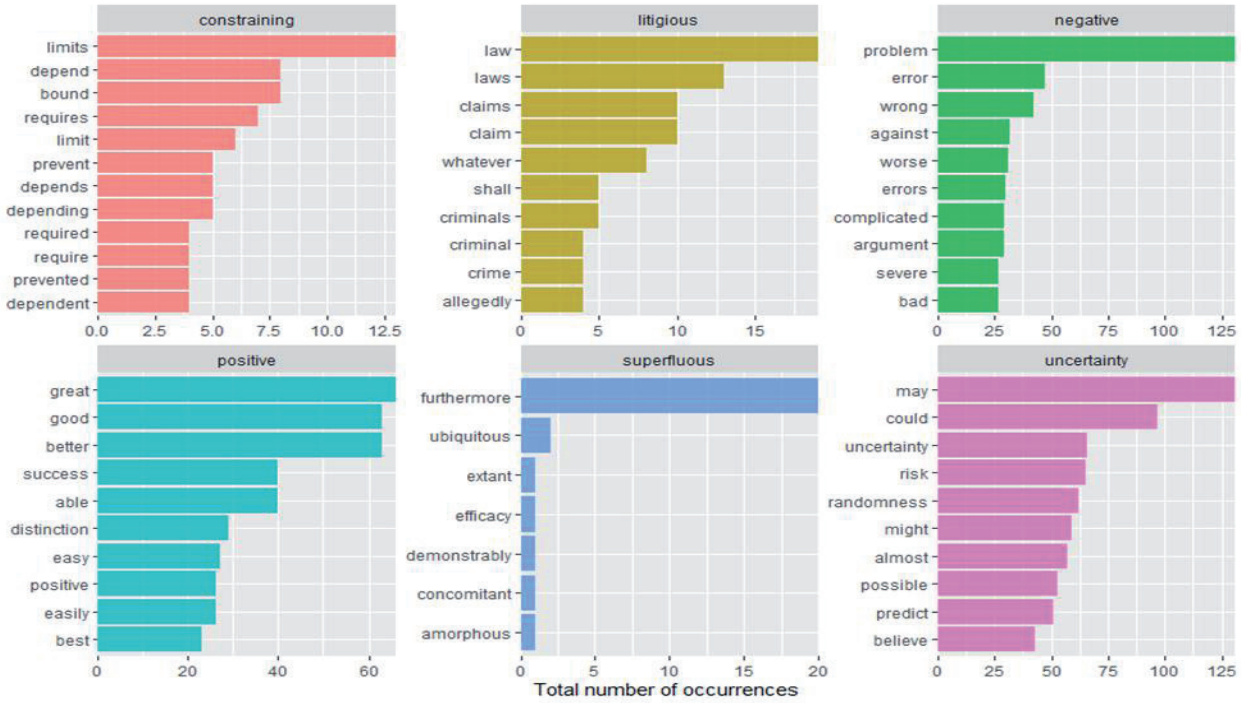
Набор пакетов *R*-платформы позволяет программно реализовать алгоритм разработки терминологических онтологий для различных предметных областей [23].

На рисунках 2, 3 приведены некоторые сравнительные результаты применения программных модулей эмоциональной оценки текстов двух значимых для экономической теории и практики финансов книг [24] и [25]. На рисунке 2 представлена визуализация результатов распределения слов книг по категориям. Из рисунка 2 видно, что слова с идентичной эмоциональной оценкой имеют разные частоты встречаемости.

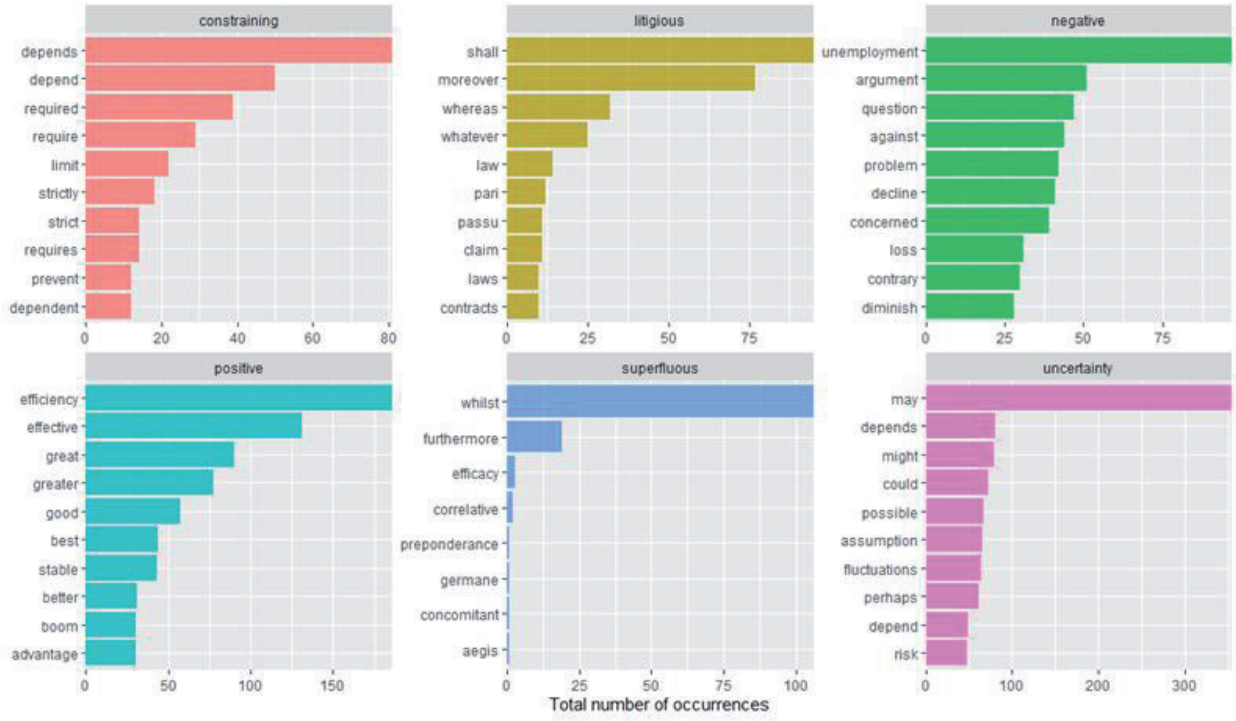
Рассмотренный пример эмоционального анализа содержания для каждой книги может быть представлен облаком слов - визуальное представление эмоциональности списка категорий слов (рисунок 3). В данном случае эта визуализация показывает отдельные слова, а важность каждого слова обозначается размером шрифта и цветом. Такое представление удобно для быстрого целостного восприятия наиболее часто встречающихся терминов и для представления распределения терминов по частоте встречаемости относительно друг друга.

На рисунке 3 представлено распределение положительных и отрицательных слов. Размер слова на рисунке пропорционален его частоте в пределах его эмоционального настроения. На

этой визуализации видно, какие из слов самые важные с их положительной и отрицательной оценкой. Следует отметить, что размеры слов не отражают оценки их эмоциональности.



а)



б)

Рисунок 2 – Распределение слов по категориям эмоциональности лексикона *loughran,* проведённый по текстам книг: а) Нассим Талеб «Черный лебедь» [24] и б) Джон Мейнард Кейнси «Общая теория занятости, процента и денег» [25]

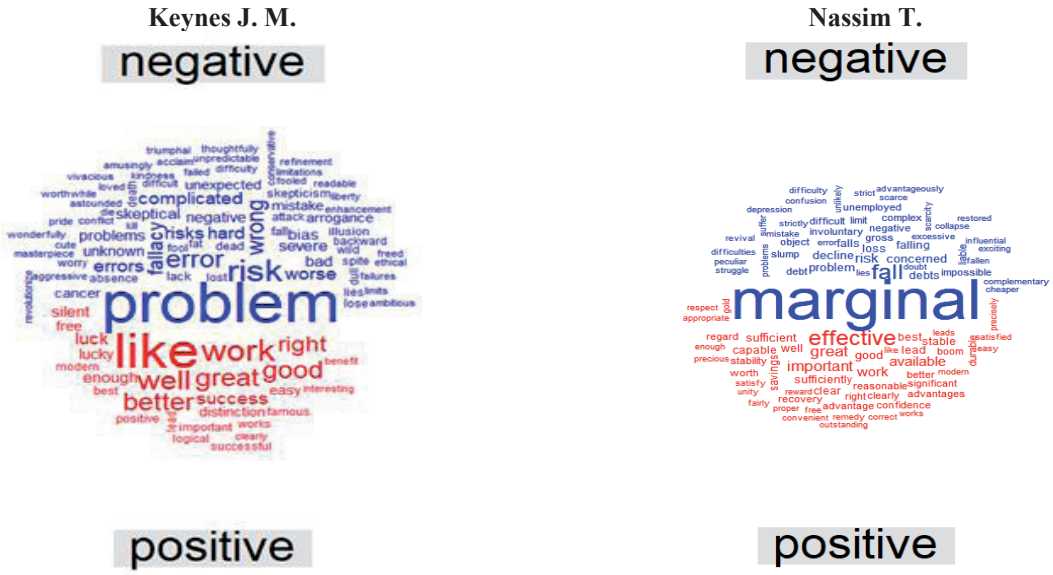


Рисунок 3 - Наиболее распространённые положительные и отрицательные слова текстов книг: Нассим Талеб «Черный лебедь» [24] и Джон Мейнард Кейнси «Общая теория занятости, процента и денег» [25]

3 Сентимент-анализ текста из области инновационных исследований газоразрядной плазмы

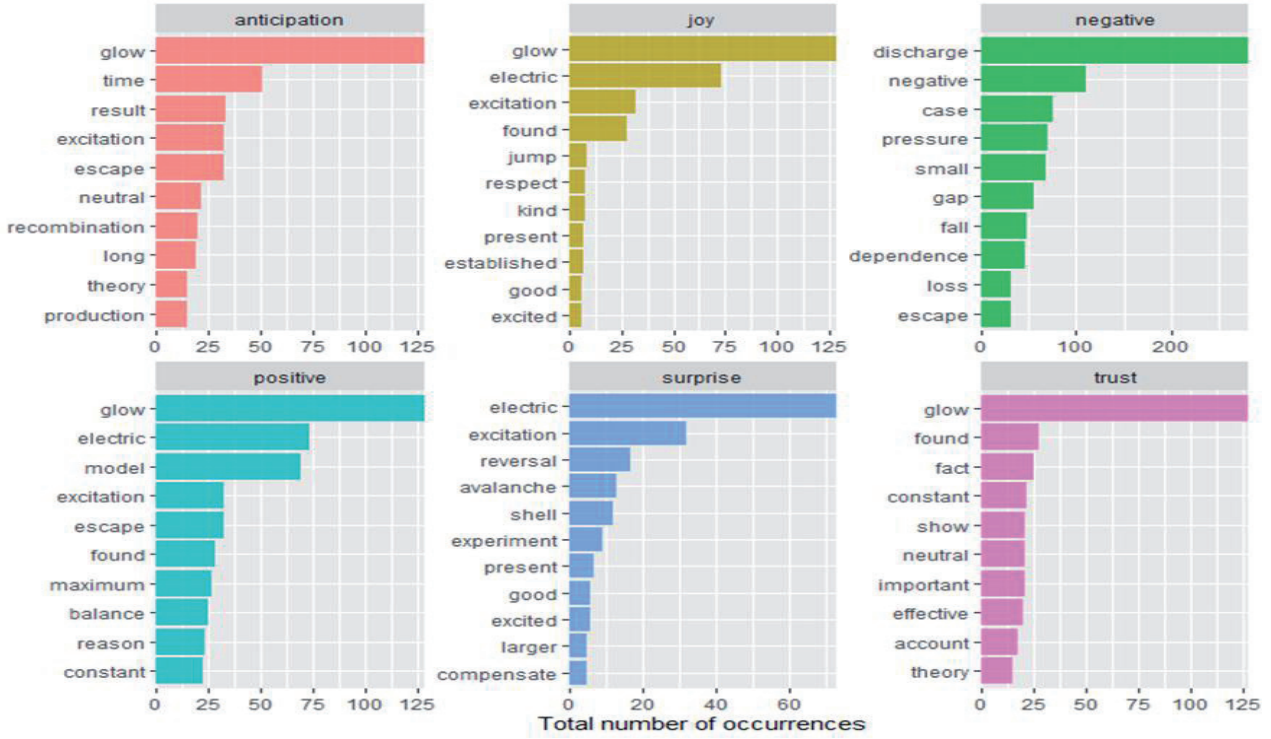
Результаты анализа эмоциональности текста по направлению теоретических исследований газоразрядной плазмы представлены на примере одной из самых современных монографий [3]. Эмоциональная компонента оценивалась с помощью лексиконов *nrc* и *loughran*. Последний ориентирован на финансовую лексику. То есть в данном случае представленный текст получает оценку с позиций финансиста, экономиста или бухгалтера. И если слово «отрицательный» в тексте публикации из области исследований газоразрядной плазмы не соответствует негативной эмоции, то оценка с помощью лексикона *loughran* указывает скорее на негативные настроения, положительный заряд - на позитивные настроения.

На рисунке 4а представлена визуализация результатов распределения слов книги по указанным базовым эмоциональным категориям лексикона *nrc*, а на рисунке 4б - распределения слов по эмоциональным категориям, отнесённым к финансам и экономике.

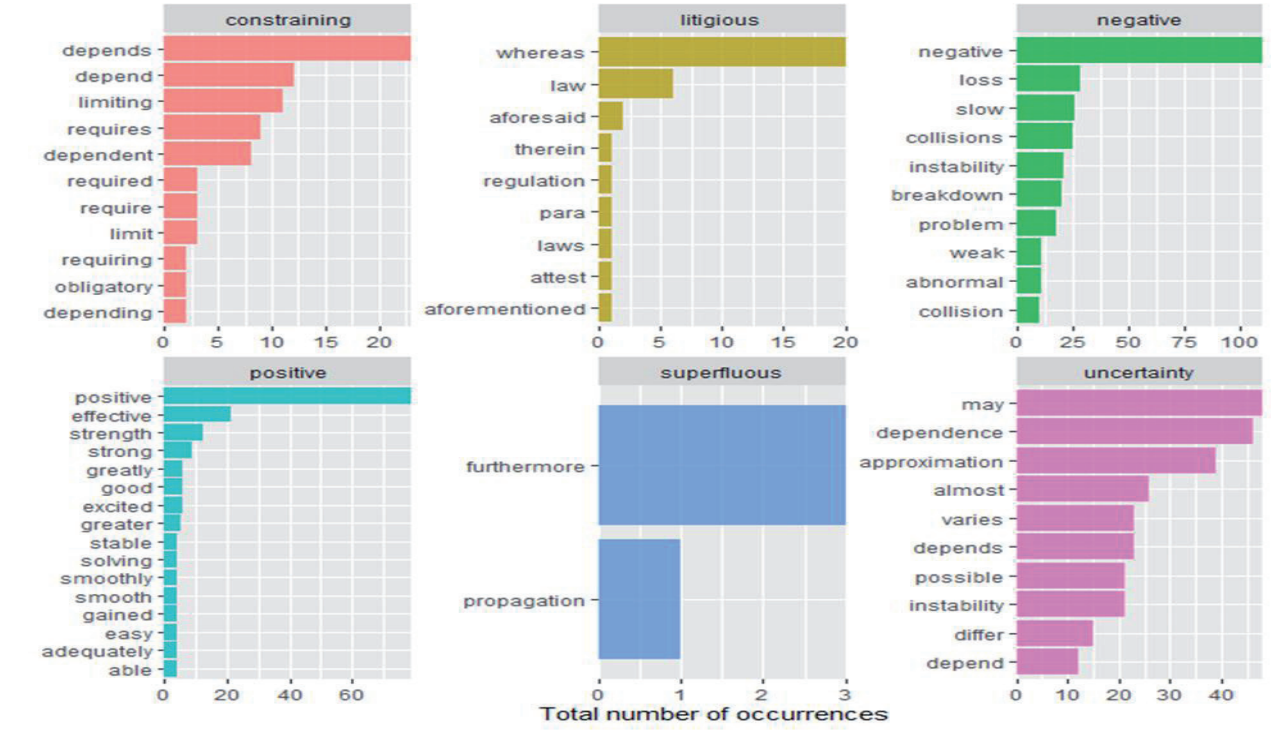
Лексикон *loughran* специально формировался для оценки эмоциональности текстов из области финансов и экономики. В настоящее время отсутствуют лексиконы для СА эмоциональности текстов научно-технического направления, в том числе текстов публикаций, связанных с тематикой исследований газоразрядной плазмы. Поэтому оценку эмоциональности таких текстов лексиконом *loughran* надо рассматривать как первый шаг в понимания того, каким должен быть лексикон оценки публикаций об инновационных объектах и процессах.

На рисунке 5 представлены гистограммы слов, которые соответствуют позитивным и негативным настроениям согласно категориям лексикона *bing*.

На этих гистограммах отчётливо видно, что наиболее часто встречаются слова «отрицательный» (*negative*) и «свечение» (*glow*). Эти слова являются «визитной карточкой» представленного в книге направления исследования плазмы. В данном случае слово «отрицательный» несёт негативную эмоциональную оценку только в обычных текстах. В данном случае это слово является эмоционально нейтральным и означает знак заряда электрона (он отрицательный) и отрицательных зарядов в плазме. Термин «свечение» также эмоционально нейтрален и соответствует излучению плазмы в видимой части спектра.



а)



б)

Рисунок 4 - Распределение слов в книге [3]: а) по категориям эмоциональности лексикона *nrc,* б) по эмоциональным категориям, отнесенным к финансам и экономике

На рисунке 6 представлено облако слов, которое даёт визуальное представление эмоциональности слов по их положительной и отрицательной оценке. Размер слова пропорционален

его частоте в пределах его настроения. На этой визуализации видны самые важные положительные и отрицательные слова.

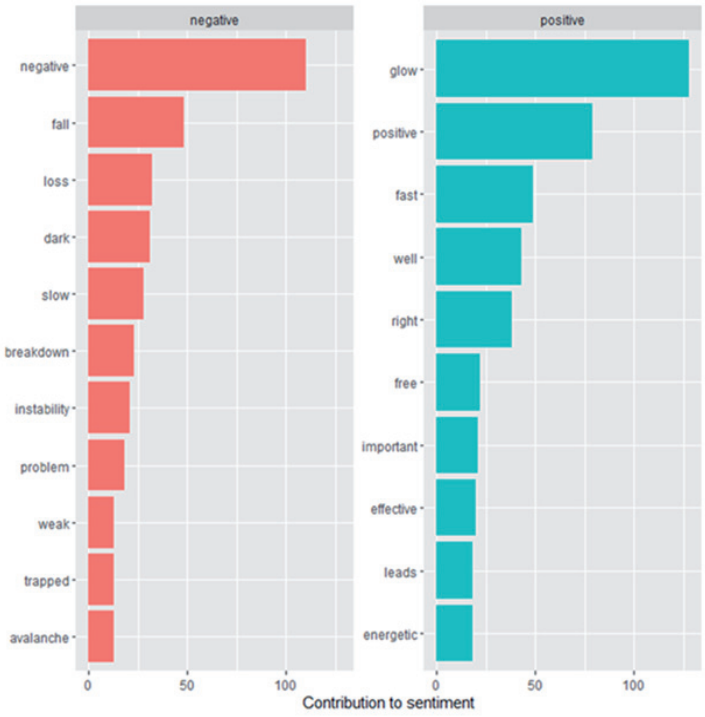


Рисунок 5 - Слова, которые соответствуют позитивным и негативным настроениям в книге [3]

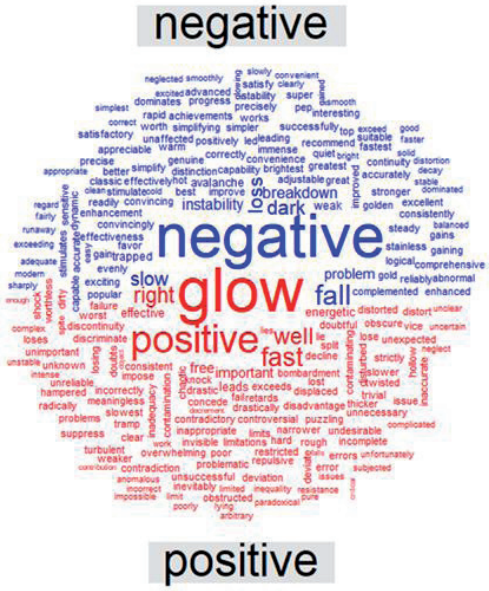


Рисунок 6 - Наиболее распространённые положительные и отрицательные слова в тексте книги [3]

Заключение

Компьютерный анализ эмоциональных оценок результатов научных исследований позволяет получить метаданные, которые не видны непосредственно в исходных текстах публикаций. Эти метаданные представляют не только исходные данные для систем искусственного интеллекта, но и создают новое интуитивное представление о предмете исследования. Они необходимы для повышения эффективности прогноза перспектив развития разных сфер финансово-экономической деятельности, например, котировок рынка ценных бумаг. Кроме того, они играют важную роль в разработке и внедрении интеллектуальных информационных систем.

Список источников

[1] ***Прохоров, А.*** Сентимент-анализ и продвижение в социальных медиа / А. Прохоров, А. Керимов // КомпьютерПресс. 07'2012. - https://compress.ru/article.aspx?id=23115#4.

[2] Практическое руководство. Анализ тональности и интеллектуальный анализ мнений // Документация по API Анализа текста. 04.12.2020. https://docs.microsoft.com/ru-ru/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis?tabs=version-3-1.

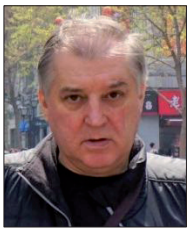
[3] ***Yuan, C.*** Introduction to the Kinetics of Glow Discharges (IOP Concise Physics) Kindle Edition / C. Yuan, A. Kudryavtsev, V. Demidov // Eurospan, 2019, - 168 p.

[4] ***Silge, J.*** Text Mining with R. A Tidy Approach / J. Silge, D. Robinson. ISBN-13: 978-1491981658 - https://www.tidytextmining.com/index.html.

[5] ***Ward, B.*** A Light Introduction to Text Analysis in R / B. Ward // Towards Data Science, May 3, 2019. - https://towardsdatascience.com/a-light-introduction-to-text-analysis-in-r-ea291a9865a8.
[6] ***Mihalcea, R.*** TextRank: Bringing Order into Texts / Rada Mihalcea and Paul Tarau / University of North Texas. 8 p. - https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf
[7] ***Fridolin, W.*** CRAN Task View: Natural Language Processing / W. Fridolin. Oxford Brookes University, UK. 2020-12-09. - https://cran.r-project.org/web/views/NaturalLanguageProcessing.html.
[8] ***Toth, M.*** Sentiment Analysis of Warren Buffett's Letters to Shareholders / M. Toth. 20 March 2017. http://michaeltoth.me/sentiment-analysis-of-warren-buffetts-letters-to-shareholders.html.
[9] ***Шрёдер, Э.*** У. Баффет. Лучший инвестор мира: пер. с англ. / Э. Шрёдер. М.: Изд. «Манн, Иванов и Фербер», 2013. — 800 с. https://www.mann-ivanov-ferber.ru/assets/files/bookparts/warlife2/warlife_read.pdf.
[10] ***Милкова, М.А.*** Извлечение ключевых терминов направления «Цифровая экономика»: графориентированный подход / М.А. Милкова // Цифровая экономика. 4(4) 2018. – с.57-65. DOI: 10.34706/DE-2018-04-06. http://digital-economy.ru/images/easyblog_articles/524/DE-2018-04-06.pdf.
[11] ***Андрианова, Е.Г.*** Роль методов интеллектуального анализа текста в автоматизации прогнозирования рынка ценных бумаг / Е. Г. Андрианова, О. А. Новикова // Cloud of Science, 2018. Т.5. № 1, с.196–206
[12] ***Георгиева, К.*** Экономические прогнозы перевернулись «с ног на голову»: мировая экономика в этом году начнет резко сокращаться / К. Георгиева: Новости ООН, 9.04.2020. - https://news.un.org/ru/story/2020/04/1375882.
[13] World Economic Situation and Prospects 2020. United Nations. New York, 2020. 236 p. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/WESP2020_FullReport_web.pdf.
[14] ***Глазьев, С.Ю.*** Управление развитием экономики / С. Ю. Глазьев. Факультет государственного управления МГУ. 2019. - 759 с. - https://aurora.network/images/Учебник_файл.pdf.
[15] ***Лесков, С.*** Академик Александр Дынкин: Нефть, бриллианты и мозги - главная ценность по всему миру. Известия, 13.03.2009. - http://www.ras.ru/digest/showdnews.aspx?id=f650b9ec-befd-4f49-be03-c07fbb1b87fd.
[16] ***Крылов, В.С.*** Перспективы и тенденции развития солнечной энергетики в условиях кризиса / В.С. Крылов, А.А. Кудрявцев, Н.Б. Косых // Ученые записки Крымского инженерно-педагогического университета, 2020, № 2(68), с.124-131.
[17] ***Крылов, В.С.*** R: компьютерный анализ эмоциональности текстов статей исследований холодной плазмы / В.С. Крылов // Информационно-компьютерные технологии в экономике, образовании и социальной сфере. – Симферополь, 2019. № 2 (24) – с. 129-136.
[18] ***Wickham, H.*** R for Data Science: Import, Tidy, Transform, Visualize, and Model Data / H. Wickham, G. Grolemund. 1st Edition. ISBN-13: 978-1491910399 - https://r4ds.had.co.nz/.
[19] Top 63 Software for Text Analysis, Text Mining, Text Analytics: https://www.predictiveanalyticstoday.com/top-software-for-text-analysis-text-mining-text-analytics/.
[20] BNOSAC Open Analytical Helpers. https://www.bnosac.be/index.php/blog/84-starspace-for-nlp-nlproc.
[21] ***Крылов, В.С.*** Homo economicus Адама Смита не холодный и не расчётливый интеллектуал // В.С. Крылов // Информационно-компьютерные технологии в экономике, образовании и социальной сфере: тез. док. IV Всеукраинской науч.-практ. конф. – Симферополь, КРП Крымучпедгиз, 2010. – С.50 – 51.
[22] ***Gebeyaw, M.*** Parsing Text for Emotion Terms: Analysis & Visualization Using R / M. Gebeyaw // R bloggers. May 11, 2017. https://www.r-bloggers.com/2017/05/parsing-text-for-emotion-terms-analysis-visualization-using-r/.
[23] ***Ландэ, Д.В***. Подход к созданию терминологических онтологий / Д.В. Ландэ, А.А. Снарский // Онтология проектирования, 2(12), 2014, с. 83 -91.
[24] ***Nassim, T.*** The black swan: the impact of the highly improbable / T. Nassim // Random House N.Y., 2007. 401 p.
[25] ***Keynes, J.M.*** The General Theory of Employment, Interest, and Money / J. M. Keynes // Macmillan Cambridge University Press, for Royal Economic Society in 1936 https://www.marxists.org/reference/subject/economics/keynes/general-theory/.

Сведения об авторах

Крылов Владимир Сергеевич, 1952 г. рождения. Окончил Ленинградский государственный университет им. А.А. Жданова в 1976, к.б.н. (1993). Доцент кафедры прикладной информатики Крымского инженерно-педагогического университета. Специализируется в области информационно-коммуникационных технологий, междисциплинарных исследований приложений информационных технологий. Автор 2-х монографий и 3-х учебных пособий, более 120 статей и докладов на конференциях. Author ID (РИНЦ): 911682, ORCID: 0000-0003-3419-6307, ResearcherID (WoS): I-6750-2017. *vskrylov@gmail.com.*



Кудрявцев Анатолий Анатольевич, 1953 г. рождения. Окончил Ленинградский государственный университет им. А.А. Жданова в 1976, к.ф-м.н. (1983). Доцент кафедры оптики физического факультета Санкт-Петербургского государственного университета. Специализируется в области физики газового разряда и плазмы. Автор 4 монографий и более 150 статей и докладов на конференциях. Author ID (РИНЦ): 20106, ORCID: 0000-0002-2232-2954, SCOPUS ID: 57203208658, ResearcherID (WoS): I-3413-2012. *akud53@mail.ru.*

Абдурайимов Ленмар Нариманович, 1983 г. рождения. Окончил Крымский инженерно-педагогический университет в 2004 , к.т.н. (2013). Доцент кафедры прикладной информатики Крымского инженерно-педагогического университета. Специализируется в области генеративных интегрированных технологий, встроенных систем и микроконтроллеров. Автор учебного пособия, более 40 статей и докладов на конференциях. Author ID (РИНЦ): 911885; Researcher ID (WoS): B-2423-2019. *abdurayimov@gmail.com.*

Поступила в редакцию 21.09.2020, после рецензирования 20.12.2020. Принята к публикации 25.12.2020.

Computer analysis of the emotional component of scientific publications using examples in physics and economics

V.S. Krylov[1], A.A Kudryavtsev[2], L.N. Abduraimov[1]

[1]*Crimean Engineering and Pedagogical University named after Fevzi Yakubov, Simferopol, Russia*

[2]*St. Petersburg State University, St. Petersburg, Russia*

Abstract

The results of the application of the system of automated extraction and visualization of metadata of emotionality of texts of scientific publications in the field of innovative research in physics and economics are presented. In physics, works on the direct conversion of light radiation into electrical energy are considered based on photoplasma, which is an effect of the appearance of a potential difference in the plasma of alkali metals. In this area, the stage of research work has practically been completed, and a transition to experimental design work is taking place with a subsequent assessment of the economic efficiency of the production and operation of photoconverters based on photoplasma. Such metadata is necessary for the development and implementation into practice of systems for the intellectual analysis of news text data, analytical publications to increase the efficiency of forecasting and assess development prospects in various areas, for example, in financial and economic activities to predict quotations on the securities market. The paper provides examples of sentiment analysis and its visualization of economic scientific texts and text from the field of gas-discharge plasma research. Instrumental methods of emotional analysis of texts are implemented in the R language, which allows in a short time to form the necessary packages of text analysis programs not only by professional programmers, but also by analysts.

Key words: *computer analysis of texts, sentiment analysis, emotional component, plasma physics, digital economy, R programming, artificial intelligence.*

Citation: *Krylov VS, Kudryavtsev AA, Abduraimov LN.* Computer analysis of the emotional component of scientific publications using examples in physics and economics [In Russian]. *Ontology of designing*. 2020; 10(4): 449-462. DOI: 10.18287/2223-9537-2020-10-4-449-462.

List of figures

References

[1] ***Prokhorov A, Kerimov A.*** Sentiment Analysis and Social Media Promotion [In Russian]. ComputerPress. 07'2012. https://compress.ru/article.aspx?id=23115#4.
[2] A practical guide. Sentiment Analysis and Intelligent Opinion Analysis [In Russian]. Text Analysis API Documentation. 12/04/2020. https://docs.microsoft.com/ru-ru/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis?tabs=version-3-1.
[3] ***Yuan C, Kudryavtsev A, Demidov V***. Introduction to the Kinetics of Glow Discharges (IOP Concise Physics) Kindle Edition. Eurospan, 2019, 168 p.
[4] ***Silge J, Robinson D.*** Text Mining with R. A Tidy Approach. ISBN-13: 978-1491981658.
[5] ***Ward B.*** A Light Introduction to Text Analysis in R: Towards Data Science, May 3, 2019. https://towardsdatascience.com/a-light-introduction-to-text-analysis-in-r-ea291a9865a8.
[6] ***Mihalcea Rada and Tarau Paul.*** TextRank: Bringing Order into Texts: University of North Texas. 8 p https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf.
[7] ***Fridolin W.*** CRAN Task View: Natural Language Processing. Oxford Brookes University, UK. 2020-12-09. https://cran.r-project.org/web/views/NaturalLanguageProcessing.html.
[8] ***Toth M.*** Sentiment Analysis of Warren Buffett's Letters to Shareholders. 20 March 2017. http://michaeltoth.me/sentiment-analysis-of-warren-buffetts-letters-to-shareholders.html.
[9] ***Schroede A***. The Snowball. Warren Buffett and the Business of Life, Bantam Books. NY. 2008.
[10] ***Milkova MA.*** Extraction of Key Terms of the Digital Economy: Graph-Oriented Approach [In Russian]. Digital economy. 2018; 4(4): 57-65. DOI: 10.34706/DE-2018-04-06. http://digital-economy.ru/images/easyblog_articles/524/digital_economy-number-4-0-6.pdf.
[11] ***Andrianova YeG, Novikova OA.*** The role of text mining methods in the automation of forecasting the securities market [In Russian]. Cloud of Science, 2018; 5(1): 196–206.
[12] ***Georgiyeva K.*** Economic forecasts have turned upside down: the world economy will begin to contract sharply this year [In Russian]. News UN, 9.04.2020. - https://news.un.org/ru/story/2020/04/1375882.
[13] World Economic Situation and Prospects 2020: United Nations. New York, 2020. 236 p. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/WESP2020_FullReport_web.pdf.
[14] ***Glaziyev SYu.*** Management of economic development [In Russian]. Faculty of Public Administration, Moscow State University 2019. 759 p. - https://aurora.network/images/Uchebnik_faĭl.pdf.
[15] ***Leskov S.*** Academician Alexander Dynkin: Oil, diamonds and brains are the main value around the world [In Russian]. Izvestiya, 13.03.2009. - http://www.ras.ru/digest/showdnews.aspx?id=f650b9ec-befd-4f49-be03-c07fbb1b87fd.
[16] ***Krylov VS, Kudryavtsev AA, Kosykh NB***. Prospects and trends in the development of solar energy during the crisis [In Russian]. Scientific notes of the Crimean Engineering Pedagogical University, 2020; 2(68): 124-131.
[17] ***Krylov VS***. R: computer analysis of the emotionality of the texts of articles of cold plasma research [In Russian]. Information and computer technologies in economics, education and social sphere. Simferopol, 2019; 2(24): 129-136.
[18] ***Wickham H, Grolemund G.*** R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 1st Edition. ISBN-13: 978-1491910399. - https://r4ds.had.co.nz/.
[19] Top 63 Software for Text Analysis, Text Mining, Text Analytics: https://www.predictiveanalyticstoday.com/top-software-for-text-analysis-text-mining-text-analytics/.
[20] BNOSAC Open Analytical Helpers. https://www.bnosac.be/index.php/blog/84-starspace-for-nlp-nlproc.
[21] ***Krylov VS***. Homo economicus of Adam Smith is not a cold and calculating intellectual [In Russian]. Information and computer technologies in economics, education and social sphere: abstracts. IV all-Ukrainian scientific-practical. conf. - Simferopol, KRP Krymuchpedgiz, 2010. P.50-51.

[22] ***Gebeyaw M.*** Parsing Text for Emotion Terms: Analysis & Visualization Using R. R bloggers. May 11, 2017. https://www.r-bloggers.com/2017/05/parsing-text-for-emotion-terms-analysis-visualization-using-r/.
[23] ***Lande DV, Snarskii AA.*** Approach to the creation of terminological ontologies [In Russian]. Ontology of designing. 2014; 2(12): 83 -91.
[24] ***Nassim T.*** The black swan: the impact of the highly improbable. Random House N.Y., 2007, 401 p.
[25] ***Keynes JM.*** The General Theory of Employment, Interest, and Money. Macmillan Cambridge University Press, for Royal Economic Society 1936. https://www.marxists.org/reference/subject/economics/keynes/general-theory/.

About the authors

Vladimir Sergeevich Krylov (b. 1952). Graduated from Leningrad State University named after A.A. Zhdanov in 1976, Ph.D. (1993). Associate Professor of the Department of Applied Informatics of the Crimean Engineering and Pedagogical University. He specializes in information and communication technology, interdisciplinary research of information technology applications. Author of 2 monographs and 3 textbooks, more than 120 journal articles and reports at conferences. Author ID (RSCI): 911682; ORCID: 0000-0003-3419-6307, ResearcherID (WoS): I-6750-2017. *vskrylov@gmail.com.*

Anatoly Anatolyevich Kudryavtsev (b. 1953). Graduated from Leningrad State University named after A.A. Zhdanov in 1976, Ph.D. (1983). Associate Professor of the Department of Optics, Physics Faculty, St. Petersburg State University. He specializes in the physics of gas discharge and plasma. Author of 4 monographs and more than 150 journal articles and conference reports. Author ID (RSCI): 20106, ORCID: 0000-0002-2232-2954, SCOPUS ID: 57203208658, ResearcherID (WoS): I-3413-2012. *akud53@mail.ru.*

Lenmar Narimanovich Abduraimov (b. 1983). Graduated from the Crimean Engineering and Pedagogical University in 2004, Ph.D. (2013). Associate Professor of the Department of Applied Informatics of the Crimean Engineering and Pedagogical University. Specializes in generative integrated technologies, embedded systems and microcontrollers. Author of textbook, more than 40 journal articles and conference reports. Author ID (RSCI): 911885; Researcher ID (WoS): B-2423-2019. *abdurayimov@gmail.com*.

Received September 21, 2020. Revised December 20, 2020. Accepted December 25, 2020.