

Научная статья

УДК 614.2

doi:10.25742/NRIPH.2023.02.010

Методика контекстно-семантического анализа: возможности применения на примере исследования термина ЗОЖ

Светлана Юрьевна Горячкина

ФГБНУ «Национальный научно-исследовательский институт общественного здоровья имени Н. А. Семашко» Минобрнауки России, 105064, г. Москва, Российская Федерация
syg66@yandex.ru, <https://orcid.org/0000-0002-9938-4168>

Аннотация. В статье приводится разбор методик анализа текстов с применением программного обеспечения для анализа текстовых данных. Приводятся примеры существующих видов анализа текстов и программ. На примере анализа термина «здоровый образ жизни» при помощи КН Coder — v.3.1 приводится пример проведенного контекстно — семантического анализа. В качестве заключения предлагается методика (последовательность шагов) для проведения аналогичного анализа.

Ключевые слова: анализ текстовых данных, контекстно — семантический анализ, КН-Coder

Для цитирования: Горячкина С. Ю. Методика контекстно — семантического анализа: возможности применения на примере исследования термина ЗОЖ // Бюллетень Национального научно-исследовательского института общественного здоровья имени Н. А. Семашко. 2023. № 2. С. 69—75. doi:10.25742/NRIPH.2023.02.010.

Original article

The procedure of contextual semantic analysis: the possibilities of application on the example of the study of the term «healthy lifestyle»

Svetlana Yu. Goryachkina

N. A. Semashko National Research Institute of Public Health, 105064, Moscow, Russian Federation
syg66@yandex.ru, <https://orcid.org/0000-0002-9938-4168>

Annotation. The article provides an analysis of the methods of text analysis using software for analyzing text data. The existing types of text analysis and programs are given. Using the example of the analysis of the term «healthy lifestyle» with the help of KH Coder — v.3.1, an example of a contextual semantic analysis is given. As a conclusion, a methodology (sequence of steps) for conducting a similar analysis is proposed.

Keywords: text data analysis, contextual semantic analysis, KH-Coder

For citation: Goriachkina S. Y. The procedure of contextual semantic analysis: the possibilities of application on the example of the study of the term «healthy lifestyle». *Bulletin of Semashko National Research Institute of Public Health*. 2023;(2):69–75. (In Russ.). doi:10.25742/NRIPH.2023.02.010.

Введение

Современному человеку ежедневно требуется обрабатывать большое количество информации, преимущественно поступающей в электронном виде.

В случае с научными трудами некоторые тематические поиски приводят современного исследователя к многотысячным результатам, среди которых требуется найти нужный участок информации и проанализировать его.

Непрерывный поток информации, ввиду развития информационного общества, основанного на развитии информационных технологий, социальных сетей приводит к тому, что ежегодный мировой прирост информации составляет более 30% объема информации к предыдущему. Так в 2020 году объем информации составил 39 ZB (что составляет 10^{21}

байта), огромный поток текстовых документов, в том числе на бумажных носителях и погрешности из-за «человеческого фактора» (усталость, ограниченность во времени, поверхностное отношение, отсутствие собственной мотивации в глубине исследования, либо, наоборот — ввиду наличия собственного непоколебимого убеждения, когда результаты «подгоняются» под необходимые факты и / или воспринимаются через призму собственной перцепции) ведут к погрешностям в исследованиях и сложности в том, чтобы не упустить (пропустить) нужную информацию). В январе 2021 года ключевой причиной выхода в интернет в мире людей явилась: поиск нужной информации (обращались в более, чем в 60% случаев)¹.

¹Digital 2021: Global overview report. Available at: <https://datareportal.com/reports/digital-2021-global-overview-report>.

Поиск методов и методик для быстрой обработки большого потока информации обусловлен такими явлениями как:

1. Цифровизация:

Цифровизация является нашим настоящим. Она затронула практически все сферы жизнедеятельности человека (в том числе, медицину и социальную сферу) и приводит к новым задачам и возможностям, например, возможности анализировать качественные данные при помощи автоматизированных алгоритмов. В настоящее время активно внедряются продукты цифровизации, в том числе с использованием искусственного интеллекта в медицине согласно стратегии развития информационного общества в России на 2017—2030 году². Так, в 2023 году планируется дополнительное финансирование в рамках национального проекта «Здравоохранение» на развитие платформы ЕГИСЗ³.

2. Прирост информации:

Ежегодный пророст информации можно выразить в объеме памяти цифровых носителей информации, серверов.

Уровень проникновения интернет во всем мире составляет более 59%. Более 85% городского и более 80% сельского населения РФ в возрасте 15—74 года активно пользуются интернетом в 2020 году, что приводит к возрастанию скорости обмена информацией^{4, 5}.

3. Сложности в анализе текстов:

Разная структура текстов (в том числе, длина предложений, абзацев, таблиц и прочих структурных элементов).

- Разный контекст текстов и лексикон.
- Разный формат и уровень значимости информации:
- Разное количество информации в источнике.
- Разное количество источников информации.
- Необходимость отследить имеющиеся связи внутри документа для установления контекста.

4. Финансовые ограничения исследователей:

Как правило, для разработки программного обеспечения требуется финансирование. Существующие зарубежные программы для анализа большого количества текстовой информации, ставшие практически недоступными в наши дни как с финансовой, так и с организационных точек зрения, приводят к необходимости поиска альтернативных путей решения задач.

5. Проблематика контекстно-семантического анализа

В каждой сфере, содержащий человеческую речь (устную, печатную, письменную) контекст может

изменяться в зависимости от профессиональной или бытовой направленности. Необходимо учитывать онтологию с учетом предметной области. Что требует от программистов не только знаний языков программирования, но и понимания глубины проблемы, системного взгляда, знаний в области онтологии [1].

6. Сложность в выборе необходимой программы для анализа.

Большое количество различных программ с различными ограничениями, стоимостью и пр. Большинство программ не содержит словарей, справочников и т. п. для русского языка, а также ограничены по числу обрабатываемой информации [1].

Существует два основных подхода, используемых при обработке текстовой информации: символический и семантический.

Символический подход представления знаний появился благодаря тому, что символы (переменные), описывались именами и значениями с дальнейшим использованием в различных вычислениях для получения желаемых результатов.

Данный подход явился основой семантического подхода. Минусами такого подхода являются: зависимость от качества программирования (программиста); ограниченность в использовании — возможно применение лишь для конкретной области, для которой проводились разработки. В случае добавления предметной области, изменения языка со временем — требуются значительные доработки программного обеспечения, а иногда и перепрограммирование. Из-за возросшей сложности роль программиста в символическом программировании двояка: роль разработчика онтологий, ответственного за описание предметной области приложения, и роль прикладного программиста, отвечающего за обработку представленных знаний [1].

Семантический подход позволяет решать проблемы, выявленные при использовании символического подхода. Для этого применяется составление иерархий взаимоотношений между буквами, затем слогами, затем морфемами и так далее, с учетом естественного восприятия языка, знаний о исследуемом языке, составлении вопросов и ответов на естественном языке. Требуется установление существующих концептов в языке, а также связей между этими концептами. По сути, буквы в данном случае, рассматриваются как атомы, когда понятия — это сложносоставные структуры, состоящие из элементов меньшей сложности: фразы состоят из слов, слова из слогов, слоги из букв [1].

Естественный язык является основой для обучения искусственного интеллекта. Основываясь на лингвистике, при помощи компьютерных программ проводится обработка, узнавание, создание, анализ речи человека (устной, либо письменной).

Методы понимания естественного языка позволяют компьютерам понимать утверждения, вопросы, ответы, команды и так далее на естественном языке [1].

Одним из примеров использования семантического анализа является аннотирование тегов, что

² Указ Президента РФ от 9 мая 2017 г. N 203 «О Стратегии развития информационного общества в Российской Федерации на 2017—2030 годы».

³ Федеральный закон от 05.12.2022 № 466-ФЗ «О федеральном бюджете на 2023 год и на плановый период 2024 и 2025 годов» (Приложение 17) // Available at: <http://actual.pravo.gov.ru/text.html#pnum=0001202212050007>.

⁴ Российский статистический ежегодник. 2021: Стат.сб./Росстат. — P76 М., 2021 — 692 с., с 476.

⁵ Digital 2021: Global overview report. Available at: <https://datareportal.com/reports/digital-2021-global-overview-report>.

важно для интерпретации данных, полученных в ходе социологических исследований, анализе литературы и пр.

Возможность заменять теги аннотациями, основанными на онтологии, облегчают задачу исследователю. Современные системы позволяют аннотировать теги произвольно, что позволяет упорядочивать по категориям, в зависимости от задач пользователя [2].

По частоте встречаемости слов в связке с определяемым словом (словосочетанием) можно сделать вывод о контексте, заложенного в определяемое слово (словосочетание). Для этого в различные программы заложены различные методики обработки и «обучения» программы понимать и анализировать смысл.

Программы можно разделить на условные группы по целям:

1. Для морфологического, синтаксического и семантического анализа (лингвистический анализ русскоязычных текстов: примерами такой программы является AskNet, Russian Context Optimizer — программы распространяются на коммерческой основе)⁶.
2. Графематический и морфологический анализ (например, Mini Project 3 (Google): Text Mining and Analysis).
3. Синтаксический и морфологический анализ: Link Grammar Parser — для работы с английским языком⁷. Программа ищет синтаксические связи для слова, которое анализируется, значения определяются на основе частоты выявленных связей; АОТ (автоматическая обработка текстов) — имеются ограничения по количеству анализируемой информации, имеются морфологические словари в том числе русский⁸.
4. Программы для статистического анализа текстов:

NetXtract — возможно использовать в MS Internet Explorer. Подразумевает индексирование слов, и создание автором контекста и формирование персональной базы данных (есть возможность установить бесплатную персональную версию)¹⁰;

WordStat — программа позволяет создать частотный словарь для анализируемого текста.

Существует множество других утилит и программ для подсчета частоты встречаемости слов, нахождения словоформ в словарях и пр. Например, программа Atlas.ti (позволяет проводить авторское кодирование с последующей обработкой данных полученных данных и связей)¹¹.

⁶ Корпоративная поисковая система AskNet Search Corporate. Available at: URL: <http://asknet.ru/SearchCorporate.htm>; <http://www.RCO.ru>

⁷ URL: <https://sites.google.com/site/sd16spring/home/assignments-and-mini-projects/text-mining>

⁸ Link Grammar Documentation, 2015. Available at: URL: <http://www.abisource.com/projects/link-grammar>

⁹ URL: http://www.aot.ru/online_demo.html

¹⁰ URL: <https://www.netxtract.com/info>

¹¹ URL: <https://atlasti.com/research-hub/qualitative-data-analysis>

Выбор программы зависит от целей, который ставит перед собой исследователь.

Рассмотрим более детально контекстно — семантический анализ.

Целями контекстно — семантического анализа могут быть:

- Количественный анализ качественных данных и графическое представление данных.
- Узнавание, понимание контекста (например, можно выявить что понимается под определенным словом, термином среди целевой аудитории или в литературе).
- Понимание запросов своей целевой аудитории (врачей, пациентов, и пр.).
- Выявление наиболее частых конструкторов (оборотов) профессиональной речи для создания шаблонов для автоматизации процессов, обучения искусственного интеллекта и др.

Что можно анализировать:

- Диалоги.
- Тексты книг, публикаций, журналов, нормативно-правовых актов.
- Тексты песен.
- Тексты мессенджеров, социальных сетей.
- Результаты опросов.
- Обратную связь пациентов.
- Записи амбулаторных карт.
- Запросы пациентов (в том числе, в поисковиках).

Любые другие источники текстовой информации, которые необходимо проанализировать исследователю.

Подробно, применяемые в различных программах методики для анализа значений больших текстов, в зависимости от вида (научная литература, социальные сети, блоги и т. п. проводятся авторами Hassani H, et all (2020) [3].

В статье рассмотрены возможности контекстно — семантического анализа на примере анализа термина «здоровый образ жизни» при помощи КН Coder — v.3.1 с описанием этапов, включая подготовительный.

Исследование выполнено в рамках Исследования выполнено в рамках темы фундаментальных научных исследований ФГБНУ «Национального НИИ общественного здоровья имени Н. А. Семашко» (FGFR-2022-0012, госрегистрация № 1021062512027-9).

Материалы и методы

Для проведения анализа использованы методы: аналитический, статистический, контекстно — семантический. На подготовительном этапе использовались: метод PRISMA — Scr^[4] и кластеризация. Контекстно — семантический анализ, обработка результатов и статистический анализ проводились с использованием КН — Coder v.3.1¹².

Материалами послужили данные 1С КонсультантПлюс: нормативно — правовые акты, содержа-

¹² Koichi Higuchi. КН Coder 3 Reference Manual, 2016. Available from: <https://kncoder.net/en/> [Accessed 17 February 2023]

и раннее выявление таких заболеваний, а также на снижение риска их развития, предупреждение и устранение отрицательного воздействия на здоровье факторов внутренней и внешней среды, формирование здорового образа жизни. 3. Формирование здорового образа жизни у граждан начиная с детского возраста обеспечивается путем проведения мероприятий, направленных на информирование граждан о факторах риска для их здоровья, формирование мотивации к ведению здорового образа жизни и создание условий для ведения здорового образа жизни, в том числе для занятий физической культурой и спортом. 4. Порядок организации и осуществления профилактики

Рис. 1. Пример кластера (Федеральный закон от 21.11.2011 N 323-ФЗ)

щие термин «здоровый образ жизни». Рассмотрено применение контекстно-семантического анализа на примере исследования термина Здоровый образ жизни (детально результаты исследования были опубликованы в Уральском медицинском журнале (выпуск 10, 2019 г.) [5]), а также мануал (руководство) к программе КН — Coder v.3.1¹³.

Результаты

Для проведения исследования с использованием методики контекстно-семантического анализа и выбора подходящей программы нами было выполнено несколько шагов (этапов).

1. Этап: Формулирование исследовательского вопроса.

Например: установить определение Здорового образа жизни в Нормативно — правовых документах (НПА), отобранных по заданным критериям.

Критерии включения документов НПА: действующие в РФ на 01 марта 2019 года, содержащие понятие «Здоровый образ жизни».

Критерии исключения: недействующие НПА; документы не содержащие термин «здоровый образ жизни»;

Также, при систематизации документов использовалась методика PRISMA-Scr.

2. Выбор наиболее подходящей программы (утилиты) для анализа текстов.

Подбор программы зависит от вида анализа, наличия в программе возможности работать с русскоязычными текстами, возможности программы работать с нужным количеством информации (некоторые программы позволяют обрабатывать не более 1500 символов).

В нашем исследовании требовалось провести контекстно — семантический, статистический анализ для выявления определения понятия здоровый образ жизни и того, что подразумевается под данным определением в НПА. Данным критериям на момент исследования соответствовала программа КН — Coder v.3.1.

3. Уточнение требований программы к формату имеющейся для анализа информации (формат файла, размер и пр.).

Например, в случае с КН Coder 3 — v.3.1 для загрузки информации необходим файл в формате *.txt.

4. Подготовка информации.

На данном этапе, перед исследованием встанет дополнительный вопрос по выбору дополнительных методов систематизации текстов. Иногда (при небольшом количестве данных) данная обработка возможна вручную.

В случае исследования термина здоровый образ жизни, ввиду большого количества текстовой информации, потребовалось проведение кластеризации (представление данных в формате интересующего для исследования фрагмента текста).

Для детального изучения термина здоровый образ жизни в имеющихся НПА требовалась та часть документа, которая связана с термином «здоровый образ жизни».

При помощи автоматической кластеризации (свидетельство о регистрации «АВКТ — v.01» № 2022614391) проведена обработка НПА для выделения кластеров с целью дальнейшего сохранения информации в формате *.txt для последующей обработки при помощи программы КН-Coder — v.3.1.

Пример кластера представлен на рисунке 1.

5. Выбор параметров для анализа в программе:

В программе необходимо выбрать исследуемое слово/слова/морфемы, установить параметры, соответствующие задачам исследования. В некоторых случаях, программы предполагают кодирование (индексирование и аннотирование) слов вручную.

Например, в исследовании термина Здоровый образ жизни требовалось определить, что же подразумевается под данным понятием в полученных кластерах документов (по контексту, в котором оно употребляется) при помощи КН-Coder. Для определения связей в программе используются коэффициенты:

Jaccard Co-efficient и Jacard Distance. Данные коэффициенты измеряются от 0 до 1. Чем сильнее связь между словами, тем коэффициент ближе или равен 1.

6. Получение результатов

Контекстно-семантический анализ — это методика, позволяющая установить семантические связи и выявить контекст (в том числе, скрытый контекст, который неявно выражен).

Программа КН — Coder позволяет получить статистический анализ данных: статистический (рис. 2) в виде частоты встречаемости слов и коллокаций — слов, которые часто встречаются вместе (рис. 3).

Также возможно посмотреть участки (примеры) текста с интересующим термином — рис. 4, который можно детально просмотреть (рис. 5).

Полученные данные формируются, например, в виде bubble-diagram, что наглядно демонстрирует полученные результаты и облегчает выводы — рисунок 6.

Диаграмма, представленная на рисунке 6 формируется на основании заданного Jaccard Coefficient (в

¹³ Koichi Higuchi. КН Coder 3 Reference Manual, 2016. Available from: <https://khcoder.net/en/> [Accessed 17 February 2023]

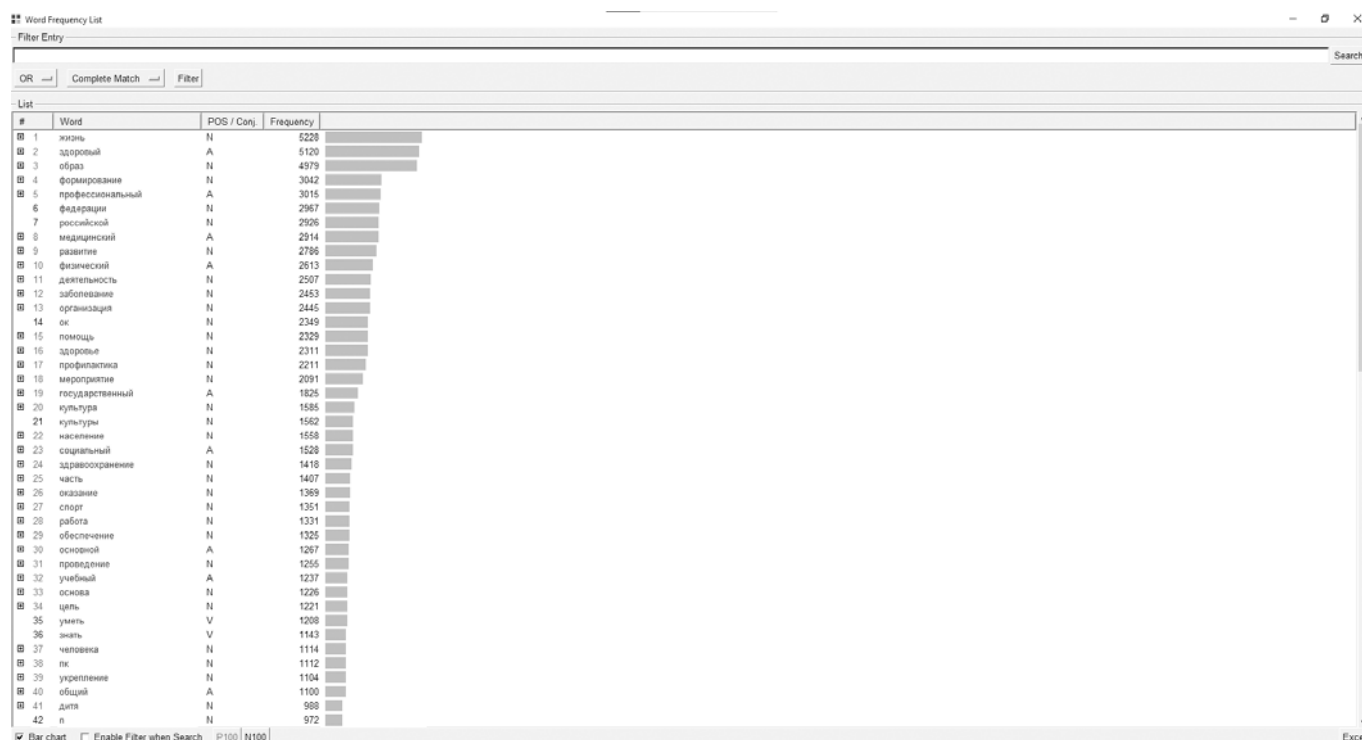


Рис. 2. Лист частот встречаемости слов

нашем исследовании был задан параметр не менее 0,4) и отображает связь на основании Jaccard Distance, а также отображает частоту встречаемости слов (отображено размером пузырей (bubbles)).

Обсуждение

Благодаря контекстно — семантическому анализу и обработке первичных данных при помощи кластеризации и КН — Coder v.3.1 в 2019 году было проанализировано более 1500 нормативно — правовых документов; найден пробел в определении термина «здоровый образ жизни» (данный термин введен в НПА приказом Министерства здравоохранения от 15 января 2020 года № 8)¹⁴.

Результаты проведенного исследования позволяют предложить комбинированную методику для анализа русскоязычных текстов (статей, публикаций, книг, нормативно-правовых документов), состоящую из следующих этапов, возможную к применению в научно — практических целях:

Примеры применения практического использования семантического подхода описаны в Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Ap-

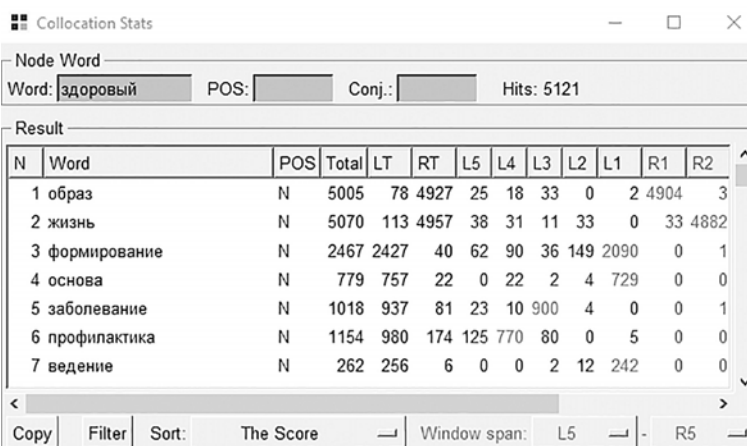


Рис. 3. Коллокации — статистический анализ

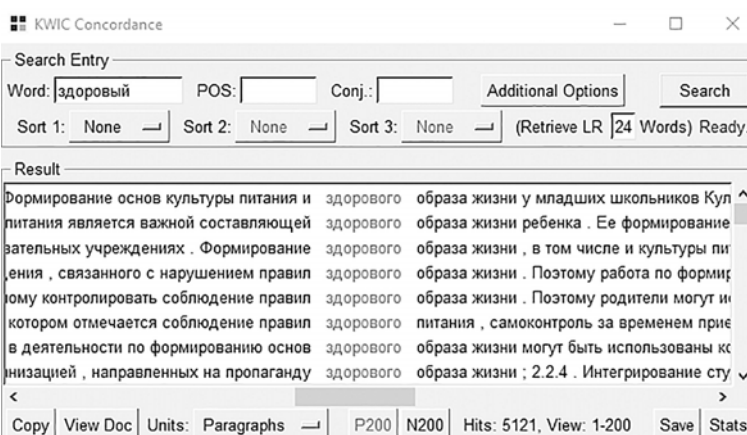


Рис. 4. Список примеров употребления слова

¹⁴ Приказ Министерства здравоохранения от 15 января 2020 года № 8 «Об утверждении Стратегии формирования здорового образа жизни населения, профилактики и контроля неинфекционных заболеваний на период до 2025 года». Available at: URL https://www.consultant.ru/document/cons_doc_LAW_344362/2ff7a8c72de3994f30496a0ccbb1ddafaddf518/



Стоит отметить, что при использовании данного метода в существующих программах имеются возможности (обработка разнообразных текстовых данных, включая большие объемы информации; объективизация контекста; возможность систематизации и классификации данных; статистическая обработка данных и пр.) и ограничения (по количеству анализируемой информации; не всегда имеются нужные морфологические словари на исследуемом языке; необходимость учитывать формат исходного файла; знание специфики семантики слов исследователем; дороговизна (стоимость) программы, что существенно ограничивает возможности исследователей), что также находит отражение в других исследованиях:

- специфика и возможности контент — анализа в отношении работы с большими качественными данными приводятся А. Н. Олейником (2019). Отмечается существующая сложность в объективной обработке большого количества текстовых данных, тогда как контент — анализ может являться источником информации для банка данных [6].
- В. С. Корецкая (2013) в своей статье приводит пример обработки качественных данных социологических исследований при помощи Atlas.ti с целью возможности использования полученных данных для принятия решения. Описываются также возможности контекстно-

- 1 Определить достоверный источник данных необходимого для анализа материала.
При работе с большим количеством источников данных рекомендуется придерживаться последовательности шагов PRISMA-Scr
- 2 Определить цель исследования текста
- 3 Определить вид анализа (статистический, синтаксический, семантический и т. п.)
- 4 Подобрать необходимую программу для анализа (Atlas.ti, KH Coder и т. п.)
Внимание! Требуется удостовериться, что выбранная программа позволяет обрабатывать тексты на языке анализируемого материала
- 5 Ознакомиться с инструкцией соответствующего программного обеспечения. Определить необходимый формат данных, которые будет необходимо загрузить / использовать при работе с программой
- 6 Провести подготовку материала в необходимом для программы формате
- 7 Ввести условия для анализа (согласно инструкции к программе)
- 8 Вывести результаты анализа, предполагаемые программой
- 9 Проанализировать полученные данные
- 10 Сделать необходимые выводы. При необходимости вернуться и повторить шаги 7—10

ориентированного подхода в работе с естественным языком при работе с документами. Проводится сравнение обычного анализа качественных данных с контекстно — ориентированным при помощи программы Atlas.ti. Приводится пример кодировки и обработки данных при помощи данного программного продукта. Но также отмечаются сложности в отсутствии качественных, многогранных программных продуктов для решения социологических задач при обработке качественных данных. Описаны существующие ограничения данной программы: например, невозможность коллективной работы и пр. [7].

- Возможности применения контекстного анализа при изучении психологических явлений, а также применения контекстного подхода для

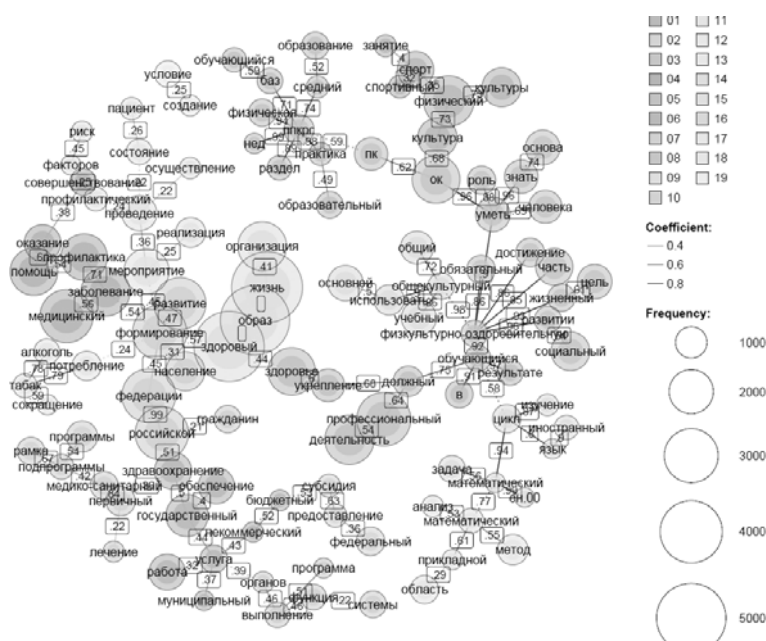


Рис. 6. Bubble diagram для термина «здоровый образ жизни».

обучения представлены в работе В.Г.Калашникова (2013, 2018). Отмечается многомерность методологии контекстного подхода с использованием метода контекстного анализа [8, 9].

- Подробный обзор программ «Vega», «TextAnalyst», «MaxQDA», «NVivo», «Ethnograph», «Atlas.ti», их возможностей, ограничений, а также о сложности в использовании ввиду дороговизны приведены Гегером А. Э. (2015) [10].

Заключение

Проведенный контекстно-семантический анализ позволяет выявить связь между словами, заложенный контекст, найти/ опровергнуть исследовательскую гипотезу, выявить пробелы, а также определить связь и т. п.

На основании проведенного анализа предложена комбинированная методика для анализа русскоязычных текстов при помощи контекстно — семантического анализа.

Разбор методик анализа текстов с применением программного обеспечения для анализа текстовых данных (при должной предварительной подготовке данных) говорит о возможности проводить анализ большого массива текстовых данных в автоматическом режиме. Но требуется внимательный подбор программного обеспечения, соответствующего исследовательским целям и задачам, а также имеющего возможности обработки языка источников данных.

Результаты, получаемые при контекстно — семантическом анализе можно применять с целью: управления социологическими процессами в группах общества (например, выявив спрос — создать предложение; выявить связанные запросы); с целью анализа литературных данных для поиска наиболее значимых источников, образов, идей; с целью анализа медицинских первичных записей — для поиска паттернов, ключевых фраз, что можно использовать для создания шаблонов; для анализа нормативно — правовых документов — для поиска терминов и пробелов в их определении; для использования естественного языка при работе с искусственным интеллектом; изучения качественных данных в социологии, психологии и других науках, что имеет научно — практическое значение.

СПИСОК ИСТОЧНИКОВ

1. Cruz-Cunha, Maria Manuela, Oliveira, Eva F., Tavares, Antonio J., et al. Handbook of Research on Social Dimensions of Semantic

Technologies and Web Services. IGI Global, Hershey, New York. 2009.

2. Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (2 Volumes). San Murugesan (Ed). IGI Global, 2009.
3. Hassani H, Beneki C, Unger S, Mazinani MT, Yeganegi MR. Text Mining in Big Data Analytics. *Big Data and Cognitive Computing*. 2020;4(1):1. doi: 10.3390/bdcc4010001.
4. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi: 10.7326/M18-0850.
5. Горячкина С. Ю. Контекстный семантический анализ нормативно-правовой базы регулирования здорового образа жизни. *Уральский медицинский журнал*. 2019;178(10):161—168.
6. Олейник А. Н. Контент-анализ больших качественных данных. *International Journal of Open Information Technologies*. 2019;(10):36-49.
7. Корецкая В. С. Контекстно-ориентированная концептуализация понятий. *Петербургская социология сегодня*. 2013;(4):267-280.
8. Калашников В. Г. Контекстный анализ как инструмент образовательной технологии. *Вестник ВГУ*. 2013;5(2):56-61.
9. Калашников В. Г. Метод контекстного анализа в методологии контекстного подхода. *Педагогика и психология образования*. 2018;(2):48-61.
10. Гегер А. Э., Чупахина Ю. А., Гегер С. А. Компьютерные программы для анализа качественных и смешанных данных. *Петербургская социология сегодня*. 2015;(6):374-383.

REFERENCES

1. Cruz-Cunha, Maria Manuela, Oliveira, Eva F., Tavares, Antonio J., et al. Handbook of Research on Social Dimensions of Semantic Technologies and Web Services. IGI Global, Hershey, New York. 2009.
2. Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (2 Volumes). San Murugesan (Ed). IGI Global, 2009.
3. Hassani H, Beneki C, Unger S, Mazinani MT, Yeganegi MR. Text Mining in Big Data Analytics. *Big Data and Cognitive Computing*. 2020;4(1):1. doi: 10.3390/bdcc4010001.
4. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi: 10.7326/M18-0850.
5. Goryachkina SYu. Contextual semantic analysis of the legal framework for regulating a healthy lifestyle. *Ural Medical Journal. [Ural'skij medicinskij zhurnal]*. 2019;178(10):161—168. (In Russian).
6. Olejnik AN. Content analysis of big qualitative data. *International Journal of Open Information Technologies*. 2019;(10):36-49. (In Russian).
7. Koreckaya VS. Context-oriented conceptualization of concepts. *Petersburg Sociology Today. [Peterburgskaya sociologiya segodnya]*. 2013;(4):267-280. (In Russian).
8. Kalashnikov VG. Context analysis as a tool of educational technology. *VGTU Bulletin. [Vestnik VGTU]*. 2013;5(2):56-61. (In Russian).
9. Kalashnikov VG. Method of contextual analysis in the methodology of the contextual approach. *Pedagogy and Psychology of Education. [Pedagogika i psihologiya obrazovaniya]*. 2018;(2):48-61. (In Russian).
10. Geger AE, Chupakhina YuA, Geger SA. Computer programs for the analysis of qualitative and mixed data. *Petersburg Sociology Today. [Peterburgskaya sociologiya segodnya]*. 2015;(6):374-383. (In Russian).

Автор заявляет об отсутствии конфликта интересов. Автор выражает благодарность ООО «ЕВРОМЕД».

The author declares no conflicts of interests. The author expresses his gratitude to LLC «Euromed»

Статья поступила в редакцию 27.02.2023; одобрена после рецензирования 11.04.2023; принята к публикации 17.05.2023.

The article was submitted 27.02.2023; approved after reviewing 11.04.2023; accepted for publication 17.05.2023.