**УДК 81**

# СИСТЕМА POLYANALYST КАК ИНСТРУМЕНТ ЛИНГВИСТИЧЕСКОГО АНАЛИЗА (НА МАТЕРИАЛЕ СООБЩЕНИЙ О СПЕЦИАЛЬНОЙ ВОЕННОЙ ОПЕРАЦИИ)

***Стародубцева Е.А.***
*Кандидат филологических наук, доцент,*
*доцент кафедры иностранных языков и профессиональной коммуникации*
*e-mail: e.starodubtseva@mail.ru*

*Курский государственный университет*

***Пикалов И.Ю.***
*Кандидат педагогических наук, доцент, доцент кафедры компьютерных технологий и информатизации образования, руководитель научно-методического центра разработки информационных систем и анализа данных*
*e-mail: pikalov@kursksu.ru*

*Курский государственный университет*

***Кривко И.П.***
*Кандидат филологических наук, доцент, заместитель директора по социальной политике и продвижению Института цифровых технологий*
*e-mail: irin-krivko@yandex.ru*

*Калининградский государственный технический университет*

*Статья посвящена описанию возможностей системы Polyanalyst в качестве инструмента анализа лингвистического материала на примере текстов о специальной военной операции (СВО). Показано, что предлагаемый авторами метод обладает возможностями текстовой аналитики: рассматривается как один из результативных инструментов, который целесообразно использовать при выделении основных тем и «сущностей» сообщений, семантических связей между ними, а также определения тональности высказываний, что позволяет оценить степень воздействия дискурсивных моделей. Статья сопровождается иллюстративным материалом, который способствует раскрытию специфики и возможностей предлагаемого инструмента. Проведенный анализ позволяет отнести систему Polyanalyst к эффективным методам отслеживания воздействующего ресурса через вербальные модели высказываний и настроений пользователей медиапространства в разные временные отрезки.*

***Ключевые слова:*** *текстовый анализ, определение тональности, сущность, сообщение, специальная военная операция, социальные сети.*

## Введение.

В последние годы исследования постов в социальных сетях являются актуальной проблемой в связи с обостряющейся необходимостью обеспечения мониторинга общественных мнений и настроений. Анализ

сообщений может найти применение в политических и социальных исследованиях, а также в отслеживании потребительских предпочтений в коммерческих целях.

На сегодняшний день существует достаточное количество инструментов анализа публикаций в Интернет и определения их тональности. Обзорные работы освещают преимущества и недостатки того или иного инструмента анализа текстов и описывают результаты исследований, в которых применялся инструмент.

В статье группы исследователей (А. Коршунов, И. Белобородов, В. Авансов и др.) описаны задачи, методы и приложения анализа сетевых и текстовых данных: определение демографических атрибутов пользователей, поиск описаний событий в корпусах сообщений, идентификация пользователей различных сетей, поиск сообществ пользователей и измерение информационного влияния между пользователями. Кроме того, рассмотрены подходы к получению исходных данных для анализа [Коршунов, Белобородов 2014: 439-455]. В статье В.С. Мошкина рассматриваются группы методов оценки тональности текстовых данных, а также описывается разработанный метод с использованием нечеткой лексической онтологии, полученной из словаря SentiWordNet 3.0 [Мошкин 2019]. Массивы больших данных и возможности их обработки в своей работе применяет Е.В. Дмитриева, анализируя особенности виртуальной коммуникации специалистов IT-сферы на материале англоязычных интернет-форумов [Дмитриева 2022]. В статье С. Сметанина и М. Комарова представлен подход к анализу тональности отзывов о продуктах на русском языке с использованием свёрточных нейронных сетей [Smetanin, Komarov 2019]. Авторы использовали предварительно обученные векторы Word2Vec в качестве входных данных для нейронных сетей, а набор обучающих данных был собран из отзывов о товарах с самым высоким рейтингом на крупнейшем сайте электронной коммерции в России, где оценки пользователей использовались в качестве меток классов. В обзорной работе «Семантический анализ для автоматической обработки естественного языка» представлено описание моделей и приложений обработки текста, описан их функционал и примеры использования, а также преимущества и недостатки инструментов [Корешкова 2021 http].

Целью данной работы является описание возможностей такого инструмента анализа, как PolyAnalyst, который применяется для извлечения полезной информации как из структурированных, так и неструктурированных данных, а также для интерпретации информации в бизнес-решениях. Функционал данной системы разнообразен, и мы посчитали возможным использовать её для проведения лингвистического исследования и оценки его потенциала в этой сфере [PolyAnalyst http].

PolyAnalyst обладает возможностями текстовой аналитики: извлечение сущностей, проверка орфографии и исправление ошибок, автоматическое определение языка, классификация текстов, определение тональности сообщений, лексикографическая, морфологическая и семантическая обработка текстов. Система содержит десятки словарей на разных языках и позволяет дополнять имеющиеся словари или создавать свои собственные. Благодаря использованию языка поисковых запросов PDL и языка извлечения информации XPDL можно производить анализ текстов для решения любых задач пользователя. PolyAnalyst предоставляет эффективные средства лингвистического и семантического анализа, а также алгоритмы машинного обучения, статистические инструменты анализа и эффективную визуализацию отчётов.

PolyAnalyst уже использовался в некоторых исследованиях. Группа ученых рассмотрела потенциал динамического подхода к анализу данных в изучении поведения пользователей в социальных сетях с использованием системы PolyAnalyst [Александрова, Лебедкина, Орлова 2021]. В работе Е.В. Митягиной, Е.В. Конышева, К.А. Чернышева и др. представлен анализ миграции выпускников вузов, расположенных в регионах с высоким оттоком населения; обработка данных также производилась с помощью этой системы [Митягина, Конышев, Чернышев 2021].

**Материалы и методы.**

Для сбора информации об объекте исследования из социальных медиа использовалась система Крибрум (Kribrum). Крибрум – это сервис медиа-аналитики, позволяющий осуществлять мониторинг интереса пользователей социальных сетей и онлайн-СМИ к информационному объекту, автоматизируя сбор информации и помогая представить данные в визуальном представлении. Среди анализируемых системой Kribrum СМИ – основные социальные сети, блог-платформы, сервисы микроблогов, форумы и интернет-СМИ [Крибрум http].

Для определения объекта использовались следующие поисковые запросы:

- спецоперация & россия;
- неонацист & украина;
- россия & сво;
- донбасс & спецоперация.

Сообщения собирались, начиная с августа 2022 года. По состоянию на середину февраля 2023 г. собрано 516 509 сообщений. Тексты представляют собой как комментарии и посты в социальных сетях, так и новостные публикации. Длина и стилистика сообщений разная.

Для анализа были выбраны сообщения с 16.08.2022 г. по 16.09.2022 г. Изначально общее количество сообщений составило 80 423.

Предобработка полученных сообщений и их последующий анализ был выполнен в PolyAnalyst. Вся работа в системе сводится к добавлению

необходимых узлов на рабочее поле, их настройке и выполнению. На рисунке 1 показан вид рабочей области проекта, который выполнял предварительную обработку полученных сообщений.

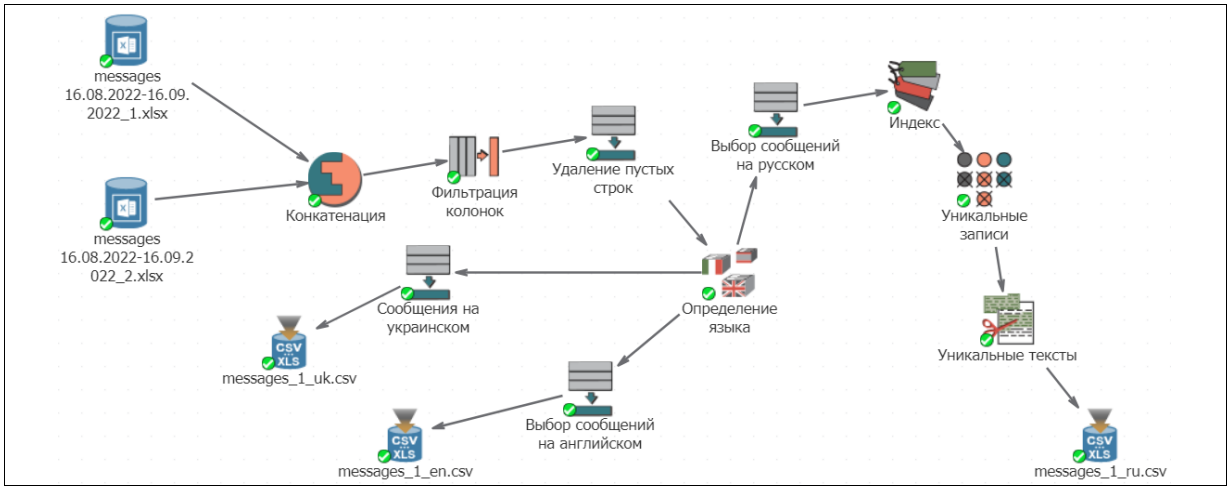



Рис. 1. Вид проекта по предобработке полученных сообщений в PolyAnalyst

Рассмотрим выполненные этапы предварительной обработки сообщений:

- Загрузка полученных файлов с сообщениями (узел Microsoft Excel);
- Объединение загруженных сообщений в один датасет (узел Конкатенация);
- Выбор полей для исследования (узел Фильтрация колонок);
- Удаление пустых сообщений (узел Фильтрация строк);
- Выбор сообщений на русском языке (узлы Определение языка и Фильтрация строк);
- Индексирование полей с текстом сообщений (узел Индекс);
- Отбор уникальных записей (узел Уникальные записи);
- Удаление дубликатов по установленному порогу схожести (узел Фильтрация строк);
- Сохранение полученного очищенного датасета (узел Экспорт в файл).

Настройка некоторых узлов заключается только в указании необходимых для работы узла файлов или полей с данными, некоторые узлы требуют дополнительной настройки. Остановимся подробнее на описании некоторых использованных узлов.

Для удаления пустых сообщений был использован узел Фильтрация строк, в котором в качестве фильтра использовался запрос not(isnull([Текст])), где Текст – это колонка, содержащая текст сообщения.

Узел Индекс используется для эффективной работы последующих узлов.

Узел Уникальные тексты использовался для того, чтобы исключить из рассмотрения перепечатанные, перенаправленные и немного изменённые сообщения. В настройках узла мы указали порог схожести 90% и необходимость сохранения самой длинной записи.

Результаты предварительной обработки сообщений показаны в таблице 1. После предварительной обработки можно сделать вывод, что собранные сообщения почти не содержали пустых значений, почти все были на русском языке (99 %) и содержали небольшое количество повторяющихся сообщений (1 %).

Таблица 1. Результаты предварительной обработки сообщений

| | **Название этапа** | **Количество записей** |
|---|---|---|
| 1. | Объединение загруженных сообщений в один датасет | 80 423 |
| 2. | Удаление пустых сообщений | 80 420 |
| 3. | Выбор сообщений на русском языке | 79 727 |
| 4. | Отбор уникальных записей | 78 655 |
| 5. | Удаление дубликатов по установленному порогу схожести | 77 063 |

Анализ текстовых сообщений начинается с узла Индекс. С одной стороны, его результаты используются последующими узлами, с другой стороны, уже в нём можно выполнить определённые настройки для нахождения лексем, работы с сокращениями и специальными символами, а также посмотреть первый анализ текстовых сообщений. В результатах работы узла можно увидеть все найденные лексемы, их принадлежность к частям речи, поддержку (в скольких загруженных текстах встречаются) и частоту. Можно посмотреть размеченный тегированный текст (рис. 2) и получить статистическую информацию по загруженным текстам.

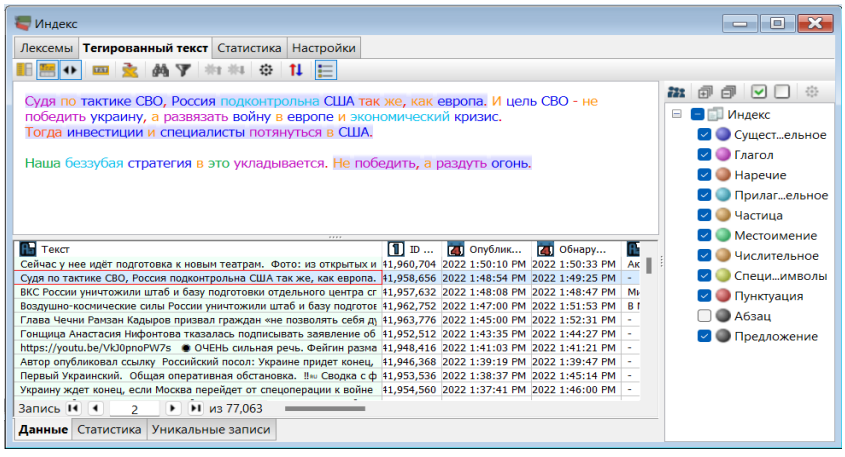



Рис. 2. Размеченный тегированный текст

Для анализа загруженных текстов можно использовать узел Извлечение сущностей, который позволяет аналитику быстро представить неструктурированные данные большого объема в структурированном виде и понять, о ком или о чем говорится в тексте. PolyAnalyst определяет большое количество сущностей, т.е. людей, организации и компании, геоадминистративные, географические и другие объекты, почтовые адреса, номера счетов, даты и т.д.

Полученное структурное представление данных с именами людей, названиями компаний, брендов, стран, адресов, номеров телефонов и др. может стать отправной точкой для дальнейшего исследования текста. Для каждой сущности существует список атрибутов, например, для сущности, обозначающей организацию, это может быть тип, расположение, отрасль деятельности, аббревиатура и т.д. Кроме того, как и многие другие объекты системы PolyAnalyst, сущности являются интерактивными: выделяя сущность в таблице, мы получаем таблицу с текстами, где она встречается.

Узел Анализ тональности отвечает на вопрос «Какие чувства испытывал говорящий или автор текста в отношении определенного объекта или ситуации?». В системе PolyAnalyst анализ тональности текста выполняется с помощью специальных правил на языке XPDL и словарей. Домен текста (его тематическая область) определяет выбор правил, используемых для извлечения тональностей.

Этот узел определяет субъект речи, объект оценки и характер самой оценки в тексте. Субъектом оценки может быть автор текста или другое лицо, чье мнение озвучивается в тексте. Субъект оценки далеко не всегда упоминается в записях, особенно если речь идет о текстах неформальной стилистики (форумы, блоги, социальные сети и др.). Оценка – это эмоциональное суждение субъекта о какой-то теме, предмете, описание эмоциональной реакции на что-то или общая тональность высказывания. Оценка выражает отношение субъекта к некоторому объекту, эмоциональную реакцию на что-либо, при этом объект тональности может быть выражен эксплицитно или подразумеваться. Объект оценки – то, что оценивается. Объект оценки так же, как и субъект, в тексте присутствует не всегда. Из выявленных объектов автоматически выделяются так называемые головные объекты.

Узел Извлечение ключевых слов генерирует интерактивный отчет, содержащий ключевые слова и прочие статистические данные, которые были извлечены из текстовой колонки исходной таблицы данных. Результаты извлечения позволяют быстро понять суть данных. В системе PolyAnalyst имеется несколько узлов для исследования ключевых слов и фраз. В частности, узел Извлечение ключевых слов – это эффективный,

простой и быстрый способ изучения некоторых часто упоминаемых в тексте терминов.

Узел Извлечение ключевых слов производит несколько видов отчета. Основным является интерактивный отчет, представленный в окне просмотра результатов. Также узел Извлечение ключевых слов создает еще и модель, которая может использоваться многими другими узлами.

Значимость ключевых слов рассчитывается по шкале от 0 до 100. Данное значение показывает, насколько уникально конкретное ключевое слово для всех текстов в исследуемой колонке. Некоторые ключевые слова выделяются не благодаря частоте, а благодаря тому, что слово встречается чаще других в данном тексте, чем в среднем по колонке. Чем больше значимость, тем больше вероятность того, что понятия в исследуемых данных тесно связаны с этим словом.

Другими словами, значимость слова показывает «аномальность» распределения данного слова по всем анализируемым текстам. Такое «аномальное» слово встречается: а) чаще, чем другие слова; б) в меньшем количестве текстов.

В большинстве случаев мера значимости может быть полезнее, чем частота, поскольку некоторые слова в тексте имеют бóльшую частоту, чем другие, но эти слова не всегда важны для понимания сути текста. Некоторые слова изначально используются часто в любом тексте, но это не значит, что они несут особую важность, поэтому можно использовать значимость в качестве более точной меры.

Следует помнить, что расчет значимости слова производится в контексте анализируемых документов. При этом значимость ключевого слова определяется в его соотношении со сбалансированным корпусом того или иного языка, т.е. зафиксированного объема письменных и устных текстов различного происхождения и различной тематики.

Показатель значимости слова также будет увеличиваться, если такое слово реже встречается в сбалансированном корпусе соответствующего языка. Иначе говоря, если слово в данном конкретном тексте встречается чаще, чем в среднем по языковому корпусу, то такое слово является ключевым.

Узел Связь терминов используется для визуализации ассоциативных связей между извлеченными терминами или ключевыми словами. Он дает возможность быстро установить, какие отношения между словами хорошо представлены в ряду документов или в колонке текстовых значений. Узел Связь терминов не просто исследует частотные ключевые слова, а изучает их ассоциации. Связи между словами выражают взаимозависимость концептов, обозначенных в записях.

Для получения, сохранения и анализа результатов выполнения узлов Извлечение сущностей и Извлечение ключевых слов часто используется узел Производная таблица. В настройках узла Производная таблица

помимо выбора типа сущностей нужно также указать тип создаваемой таблицы, выбрав один из вариантов: Список сущностей, Строки с сущностями, Колонки с сущностями, Список размеченных текстов. В зависимости от выбранного типа таблицы появляются доступные для отображения колонки. Например, при выборе типа таблицы Строки с сущностями доступны для добавления следующие поля: Сущность, Тип сущности, Поддержка, Частота, Валидация, Позицию. Кроме того, можно менять доступные для отображения поля.

Для визуализации результатов узла Производная таблица можно использовать узел Граф, позволяющий выполнять графическую визуализацию с использованием фильтрации и детализации результата.

С учётом возможностей описанных узлов возможный вид рабочего поля проекта по анализу текстовых сообщений показан на рисунке 3.

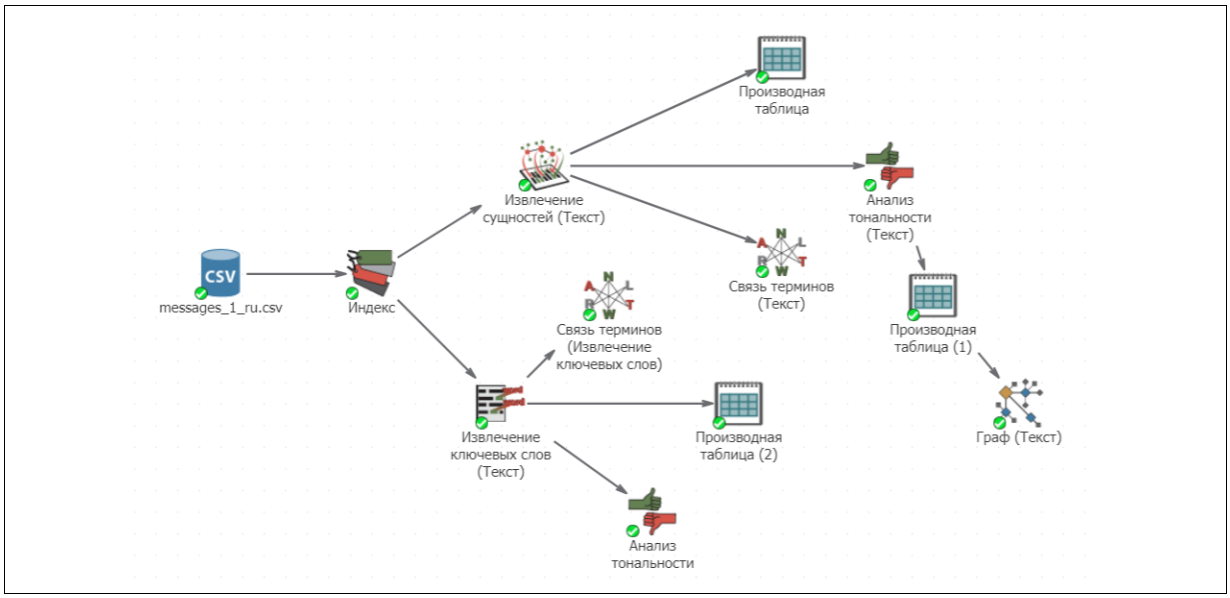



Рис. 3. Вид проекта по анализу текстовых сообщений в PolyAnalyst

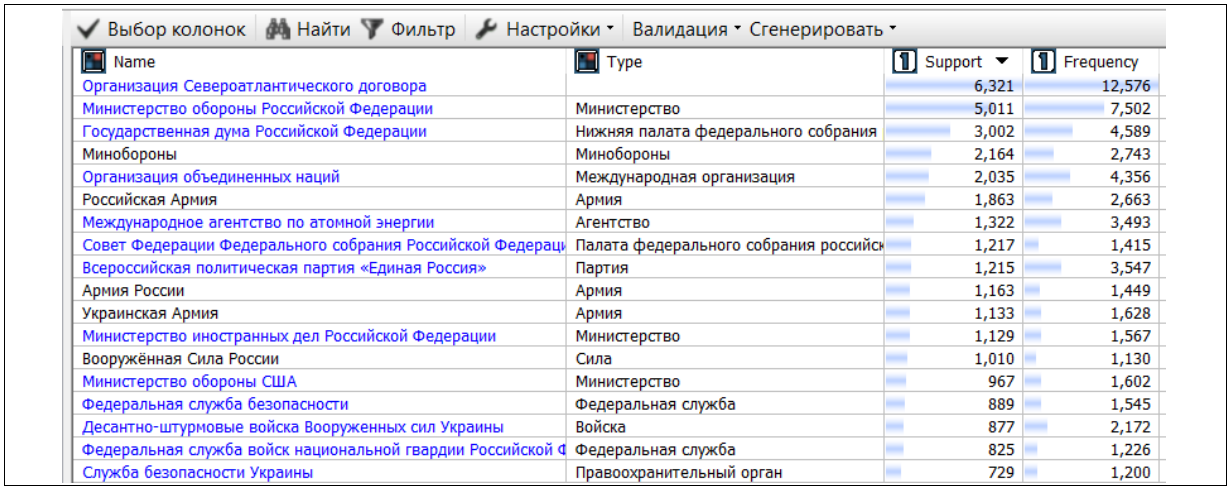
**Обсуждение результатов.**

В целях нашей работы были выделены сущности, проведен анализ тональности и определены ключевые слова. По умолчанию PolyAnalyst выделяет 18 типов сущностей, но, используя язык XPDL, можно создавать и свои сущности для поиска и анализа.

***Узел Извлечение сущностей.***

Для нашего исследования нам не нужны такие сущности, как идентификаторы, почтовые адреса, телефонные номера и e-mail-адреса. Эти сущности мы исключили в настройках узла. Итак, было выделено 11 типов сущностей, количество которых составило 101 640. Выделяя соответствующий тип сущности, можно посмотреть на их таблицу с указанием частоты и поддержки. Приведём фрагмент таблицы с

найденными организациями, расположенными по убыванию поддержки в текстах (рис. 4).



Выбор колонок | Найти | Фильтр | Настройки | Валидация | Сгенерировать

| Name | Type | Support | Frequency |
|---|---|---|---|
| Организация Североатлантического договора | | 6,321 | 12,576 |
| Министерство обороны Российской Федерации | Министерство | 5,011 | 7,502 |
| Государственная дума Российской Федерации | Нижняя палата федерального собрания | 3,002 | 4,589 |
| Минобороны | Минобороны | 2,164 | 2,743 |
| Организация объединенных наций | Международная организация | 2,035 | 4,356 |
| Российская Армия | Армия | 1,863 | 2,663 |
| Международное агентство по атомной энергии | Агентство | 1,322 | 3,493 |
| Совет Федерации Федерального собрания Российской Федерации | Палата федерального собрания российск | 1,217 | 1,415 |
| Всероссийская политическая партия «Единая Россия» | Партия | 1,215 | 3,547 |
| Армия России | Армия | 1,163 | 1,449 |
| Украинская Армия | Армия | 1,133 | 1,628 |
| Министерство иностранных дел Российской Федерации | Министерство | 1,129 | 1,567 |
| Вооружённая Сила России | Сила | 1,010 | 1,130 |
| Министерство обороны США | Министерство | 967 | 1,602 |
| Федеральная служба безопасности | Федеральная служба | 889 | 1,545 |
| Десантно-штурмовые войска Вооруженных сил Украины | Войска | 877 | 2,172 |
| Федеральная служба войск национальной гвардии Российской Ф | Федеральная служба | 825 | 1,226 |
| Служба безопасности Украины | Правоохранительный орган | 729 | 1,200 |

Рис. 4. Таблица сущностей типа Организация

Для визуализации связей между сущностями можно использовать узел Связь терминов (рис. 5). Из графа связей между сущностями видим один центр – «Россия» (поддержка 43 567 текстов, относится к типу сущностей GeoAdministrative). Граф также является интерактивным: выделяя сущность или связь, мы будем видеть перерисованный граф, показывающий только связи с выделенной сущностью. Так, если мы выделим сущность Россия, то граф примет вид, показанный на рисунке 6.

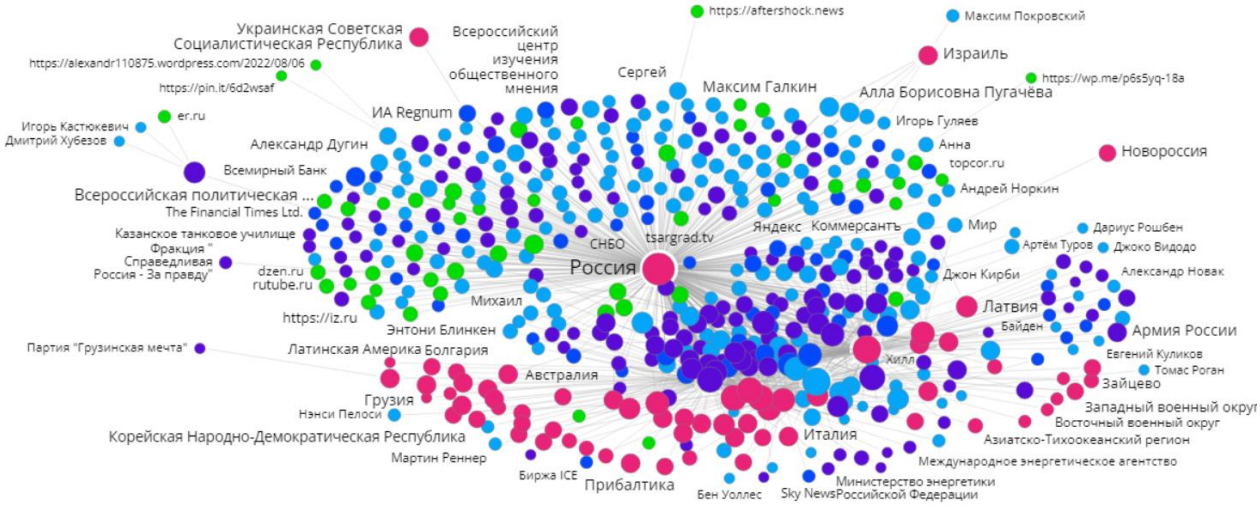



Рис. 5. Граф связей сущностей

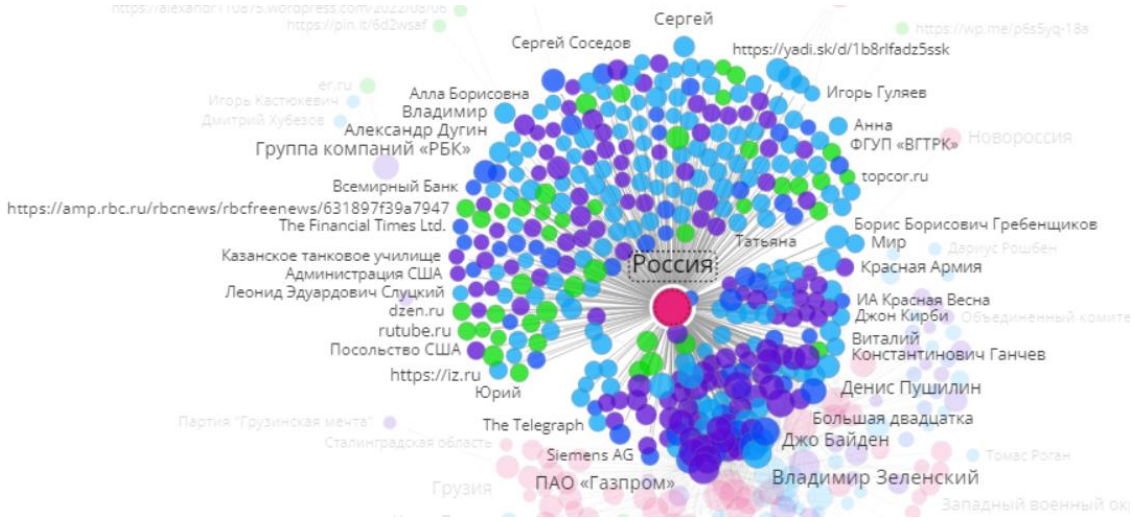



Рис. 6. Граф связей сущностей с выделенной сущностью

В окне представления графа можно также менять такие параметры отображения, как сила связи и поддержка узла. Если настроить отображение узлов с поддержкой от 3 700 сообщений, то на графе будут показаны связи сущности Россия с такими сущностями, как Соединённые Штаты Америки, Владимир Путин, Европа, Европейский союз, Министерство обороны Российской Федерации (рис. 7).

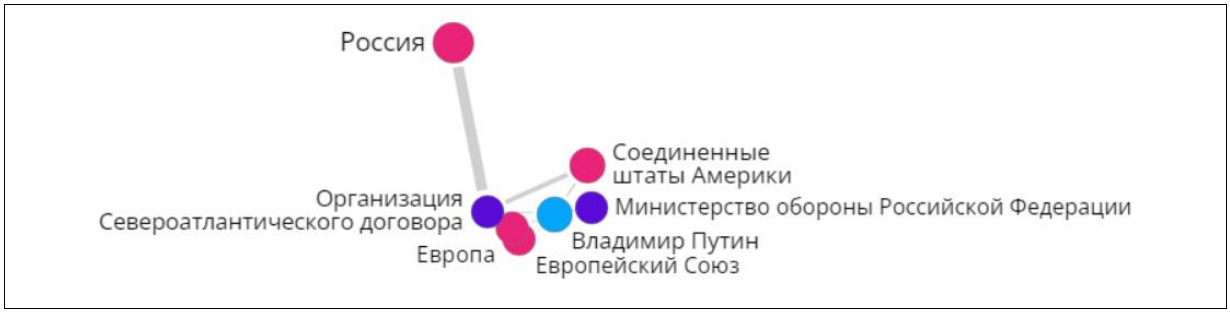



Рис. 7. Граф связей сущностей, поддержка которых от 3 700

Вкладка Связь терминов позволяет посмотреть таблицу, где также указана сила связи и поддержка (рис. 8). Наиболее сильная связь наблюдается между сущностями Организация Североатлантического договора – Россия (сила связи – 1, поддержка 6 226 текстов). Далее идет связь между РИА Новости – Россия (сила связи – 0,48, поддержка 2 848 текстов). Следом с одинаковой силой связи идут узлы Организация Североатлантического договора – Соединенные Штаты Америки (сила связи – 0,39, поддержка 3 163 текста) и Государственная Дума Российской Федерации – Россия (сила связи – 0,39, поддержка 2 703 текста). В 15 первых по силе связей попали также связи узлов Организация Объединённых Наций – Россия, Владимир Зеленский – Россия, Украинская армия – Россия, Владимир Путин – Донбасские республики.

Визуально связь терминов можно посмотреть на вкладке Облако терминов (рис. 9).

| # | Первый узел | Второй узел | Сила связи | Поддержка |
|---|---|---|---|---|
| 1 | Организация Североатлантического договора | Россия | 1.00 | 6,226 |
| 2 | РИА Новости | Россия | 0.48 | 2,848 |
| 3 | Организация Североатлантического договора | Соединенные штаты Америки | 0.39 | 3,163 |
| 4 | Государственная дума Российской Федерации | Россия | 0.39 | 2,703 |
| 5 | Минобороны | Россия | 0.35 | 2,080 |
| 6 | Организация объединенных наций | Россия | 0.34 | 2,002 |
| 7 | Владимир Зеленский | Россия | 0.33 | 1,878 |
| 8 | ИТАР-ТАСС | Россия | 0.31 | 1,767 |
| 9 | Международное агентство по атомной энергии | Россия | 0.23 | 1,303 |
| 10 | Российская Армия | Россия | 0.22 | 1,608 |
| 11 | Совет Федерации Федерального собрания Российской | Россия | 0.21 | 1,189 |
| 12 | Министерство обороны США | Соединенные штаты Америки | 0.20 | 755 |
| 13 | Украинская Армия | Россия | 0.20 | 1,121 |
| 14 | Владимир Путин | Донбасские республика | 0.20 | 561 |
| 15 | Государственная дума Российской Федерации | Донбасские республика | 0.19 | 557 |

Рис. 8. Связь терминов

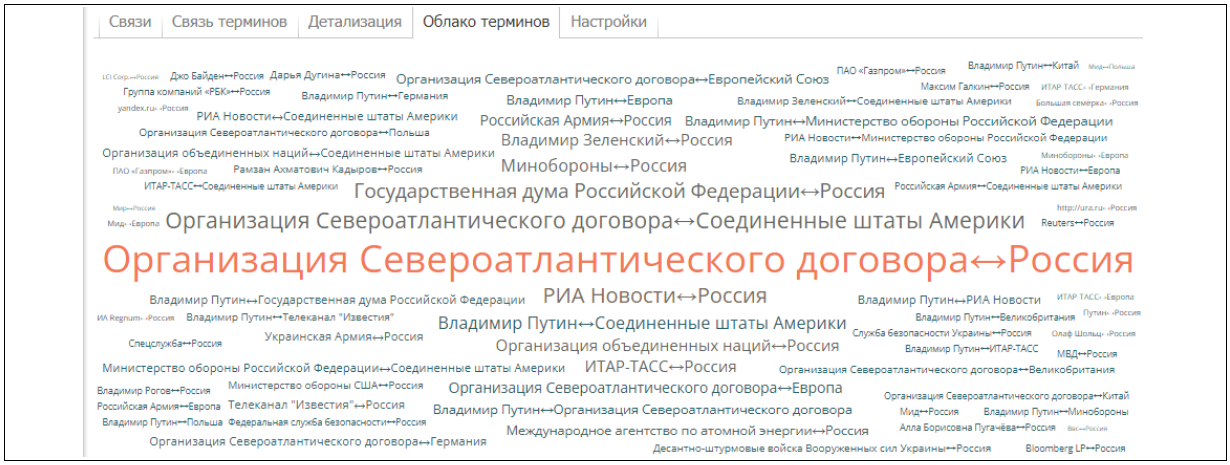



Рис. 9. Облако терминов

Извлечение сущностей помогает составить представление об общей тематике новостей, сообщений и постов. Самыми обсуждаемыми темами являются непосредственно политические события, связанные с центральными игроками – Россией, Украиной, США, а также президентами стран Путиным и Зеленским, международными (НАТО), российскими организациями (Минобороны, Государственная Дума РФ).

Помимо наиболее частотных сущностей с высокой текстовой поддержкой, интересными показались следующие объединения: Максим Галкин-Россия-Израиль, Алла Борисовна Пугачёва-Россия-Израиль, связи Организации Североатлантического договора с такими странами и регионами, как Армения, Канада, Литва, Прибалтика, Польша, Италия, Сирия, Латвия, Индия, и такими персонами, как Сергей Кужугетович Шойгу, Дмитрий Анатольевич Медведев, Рамзан Ахматович Кадыров, Владимир Рогов. Новороссия (поддержка 388) на построенной диаграмме имеет связь только с термином Мир (поддержка 153).

Подобные связи и без анализа самих текстовых сообщений показывают тематику новостей. Такая функция может быть очень удобной в поиске и фильтре текстовых сообщений определенной направленности.

***Узел Анализ тональности.***

Ниже представлены результаты анализа тональности нескольких выделенных сущностей: Россия, Украина, США, Путин, Зеленский. Россия как объект высказывания связан с такими оценочными лексическими единицами, как *справедливый* (поддержка 219, позитивное отношение), *санкция* (поддержка 168, негативное отношение) *критиковать* (поддержка 117, негативное отношение), *ненависть* (поддержка 102, негативное отношение), *любить* (поддержка 99, позитивное отношение), *неприемлемая угроза* (поддержка 71, негативное отношение), *ненавидеть* (поддержка 68, негативное отношение), *уважение* (поддержка 60, позитивное отношение), *проблема* (поддержка 48, негативное отношение).

Если рассматривать связи с поддержкой от 20 высказываний, то негативные высказывания, где объектом является Россия, составляют 60%, 40 % – позитивных. Соответствующий граф тональности показан на рисунке 10.

Украина как объект высказывания с поддержкой связи тональности от 20 высказываний связана только с оценкой *проблема* (поддержка 25, негативное отношение). Если перечислить связи с поддержкой от 7 высказываний и выше, то можно увидеть среди негативных такие оценки, как *проблема*, *опасный район*, *ненавидеть*, *враждебный, ненависть*. Среди позитивных оценок с поддержкой от 7 высказываний найдены: *любить, любовь, квалифицированный лётный состав, одобрить, важный военный объект, уверенно освобождать*. Все связи приведены в порядке уменьшения поддержки.

Соединенные Штаты Америки имеют максимальную поддержку связи тональности – только 13 высказываний. Приведём первые 4 связи: *опасаться* (поддержка 13), *проблема* (поддержка 11), *бояться* (поддержка 10), *грубо вмешиваться* (поддержка 7). Все перечисленные связи являются негативными.

Объект «Владимир Владимирович Путин» имеет при определении тональности связи с поддержкой от 1 до 18. Самая сильная связь с оценкой *недовольство* (поддержка 18, негативное отношение). Приведём также связи с поддержкой от 8 до 5: *великий* (позитивная), *молодец* (позитивная), *плохой* (негативная), *злой* (негативная), *любить* (позитивная), *спасибо* (позитивная), *смеяться* (негативная), *справедливый* (позитивная).

Объект «Владимир Зеленский» имеет при определении тональности связи с поддержкой от 1 до 7. Все связи с поддержкой от 7 до 3 являются негативными, укажем их в порядке убывания силы: *недовольный, критиковать, разочароваться, возмутить, неопытный, грязно играть.*

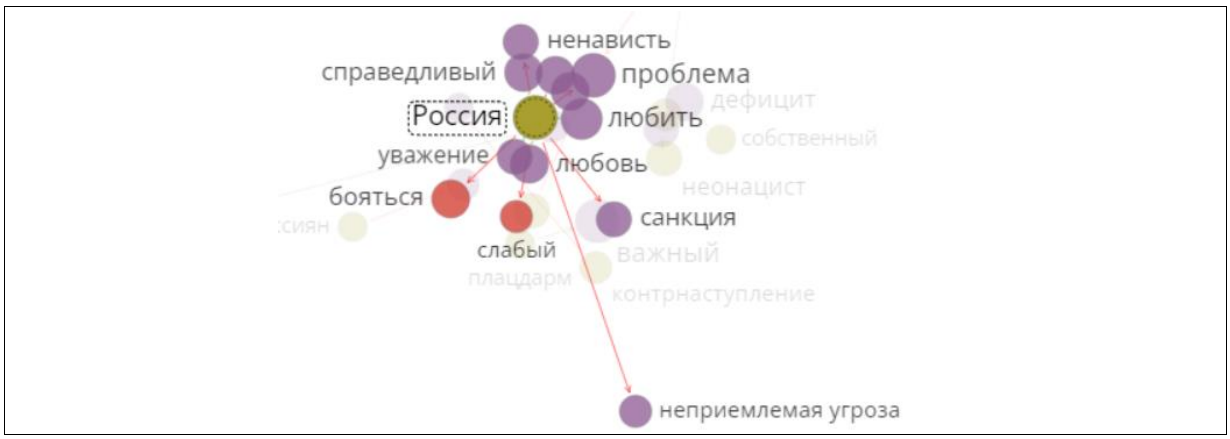



Рис. 10. Граф тональности, связывающий объект Россия и оценку (поддержка связи больше 20)

В большинстве случаев оценочная характеристика субъектно-объектных отношений определена правильно, обращение к тестам это подтверждает:

– ***«США** и их **вассалы грубо вмешиваются во внутренние дела суверенных государств:** организуют провокации, государственные перевороты, гражданские войны»;*

– *«… во всём виноват **плохой Путин»;***

– *«Во многих отношениях **Путин – великий лидер** и человек мира…»* [Polyanalyst http].

Но не везде, распознав и определив оценку отношения к объекту как отрицательную, само предложение или полностью сообщение имеет отрицательный смысл по отношению к объекту высказывания. Например, сущность *Путин* выступает объектом отрицательной оценки, но контекст всего сообщения показывает положительную оценочную окраску:

*«В ходе спецоперации на Украине Россия защищает своё будущее и предотвращает большую войну, несмотря на информационную агрессию и преднамеренные попытки создать у россиян негативное отношение к тому, что происходит в стране, вызвать **недовольство политикой Путина В.В.**»* [Polyanalyst http].

Фраза *«недовольство политикой Путина В.В.»* отмечена как негативное отношение к субъекту, в то время как это не так.

***Узел Извлечение ключевых слов.***

В настройках узла мы отметили, что нас интересуют такие части речи, как существительное и прилагательное. В нашем исследовании программа выделила 23 168 ключевых слов, из них 60 % существительных, 40 % прилагательных, что говорит об описательном и нарицательном характере языковых средств, используемых при обсуждении СВО.

Наиболее частотные из выделенных ключевых слов приведены в таблице (табл. 2):

Табл. 2 Таблица частотности ключевых слов

| № | Ключевое слово | Частотность | Поддержка текстов |
|---|---|---|---|
| 1 | Военный | 31 922 | 15 950 |
| 2 | Украинский | 31 405 | 13 178 |
| 3 | Боевой | 12 713 | 7 282 |
| 4 | Мирный | 10 357 | 7 099 |
| 5 | Западный | 10 356 | 6 283 |
| 6 | Специальная военная | 8391 | 6 324 |
| 7 | Населённый | 7 788 | 3 554 |
| 8 | Американский | 6 849 | 4 191 |
| 9 | Народный | 6 381 | 4 070 |
| 10 | Киевский | 6 177 | 4 423 |

По степени значимости в текстах система определила ключевые слова несколько иным образом (рис. 11).

| # | Ключевое слово | Значимость | Поддержка | Частота |
|---|---|---|---|---|
| 1 | военный | 100.00 | 15,950 | 31,922 |
| 2 | специальная военная | 94.65 | 6,324 | 8,391 |
| 3 | украинский | 89.77 | 13,178 | 31,405 |
| 4 | украинские военные | 51.33 | 1,321 | 1,414 |
| 5 | боевой | 42.75 | 7,282 | 12,713 |
| 6 | мирный | 38.09 | 7,099 | 10,357 |
| 7 | западный | 34.99 | 6,283 | 10,356 |
| 8 | американский | 23.66 | 4,191 | 6,849 |
| 9 | населённый | 23.49 | 3,554 | 7,788 |
| 10 | специальная военная автомобильная | 22.30 | 340 | 344 |
| 11 | народный | 19.74 | 4,070 | 6,381 |
| 12 | политический | 19.31 | 3,429 | 5,555 |

Рис. 11. Результаты по обработке степени значимости ключевых слов

Можно заметить, что ключевое слово *военный* занимает первое место по частотности и по степени значимости, а вот далее идут расхождения. Частотность не всегда определяет важность понятия. Так, по степени значимости система на второе место ставит ключевое слово *специальная военная*, далее – *украинский*, *украинские военные*, *боевой*, *мирный*. Ключевое слово *народный* занимает одиннадцатое место по значимости, в то время как по частотности – девятое.

Вкладка Облако ключевых слов в данном узле позволяет визуализировать полученный результат для большей репрезентативности. Наведение курсора на каждую единицу из Облака показывает значимость и текстовую поддержку, а в отдельном окне открывает примеры из текстов (рис. 12).

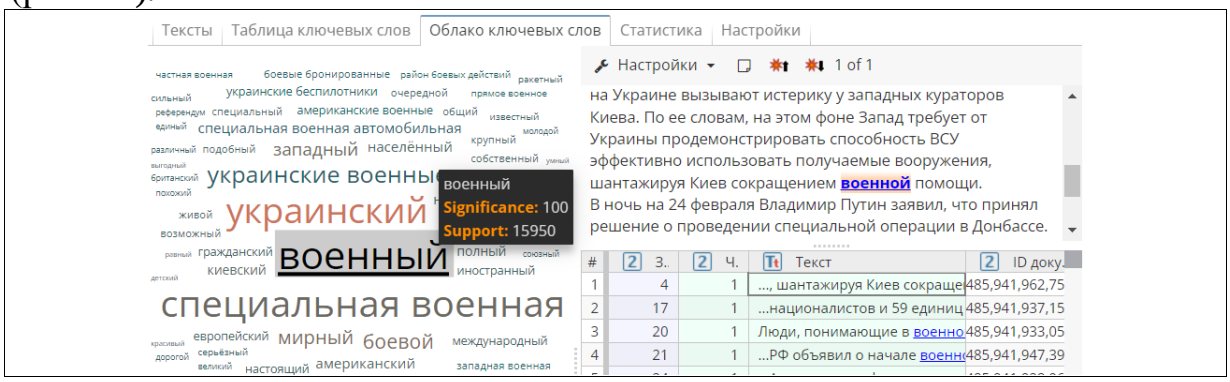



Рис. 12. Вкладка *Облако ключевых слов*

***Узел Связь терминов***

Применяя узел Связь терминов после узла Извлечение ключевых слов, мы указываем, что узел будет строить связи между извлеченными нами терминами. На созданном системой графе связей терминов выделены доминантные центры: *военный*, *украинский*, *боевой, мирный, западный* (рис. 13).

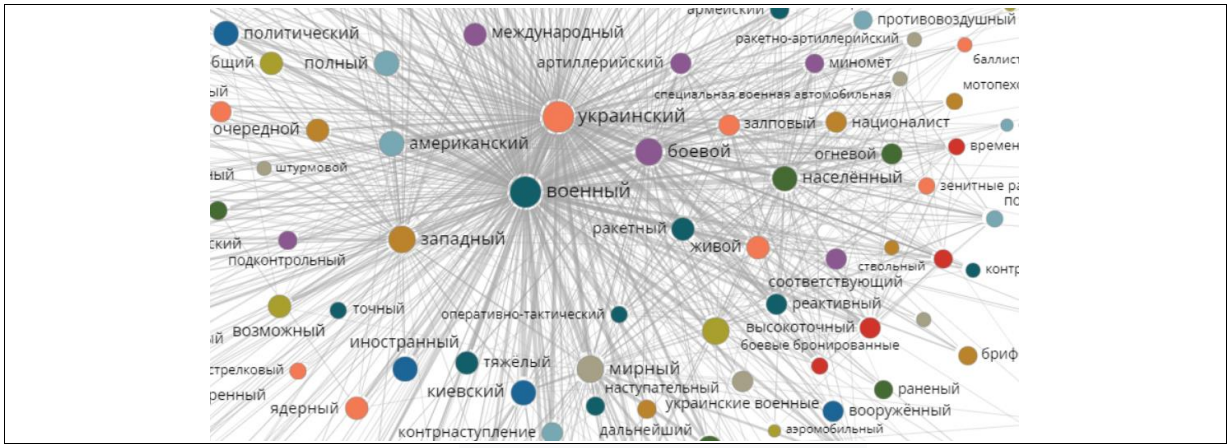



Рис. 13. Связь ключевых слов

Самая сильная кластеризация заметна вокруг ключевого слова *украинский*, связь которого можно подробно изучить, используя настройки по отображению с учётом силы связи. Установив силу связи 0,46, мы видим, что термин наиболее сильно связан с такими ключевыми словами, как *военный, населённый, американский, западный, киевский, мирный, ракетный, артиллерийский.*

Второй доминантный кластер формируется вокруг ключевого слова *военный*, которое образует связи со следующими понятиями (сила связи

также от 0,46): *украинский, боевой, населённый, американский, западный, киевский, демилитаризация, мирный, специальная военная, артиллерийский, живой, высокоточный, ракетный*.

Интерес представляют не только доминантные центры, но и периферийные, которые позволяют анализировать специфические связи, ассоциации или представления людей о важных понятиях, которые редко обсуждаются публично.

Ключевое слово *ядерный* на рис. 14 (поддержка 2 253 текста) имеет связи с ключевыми словами: *военный, украинский, атомный, западный, американский, боевой, мирный*. Рассматриваемый граф дает представление о том, в каких контекстах фигурирует слово *ядерный*, с какими понятиями оно связано.

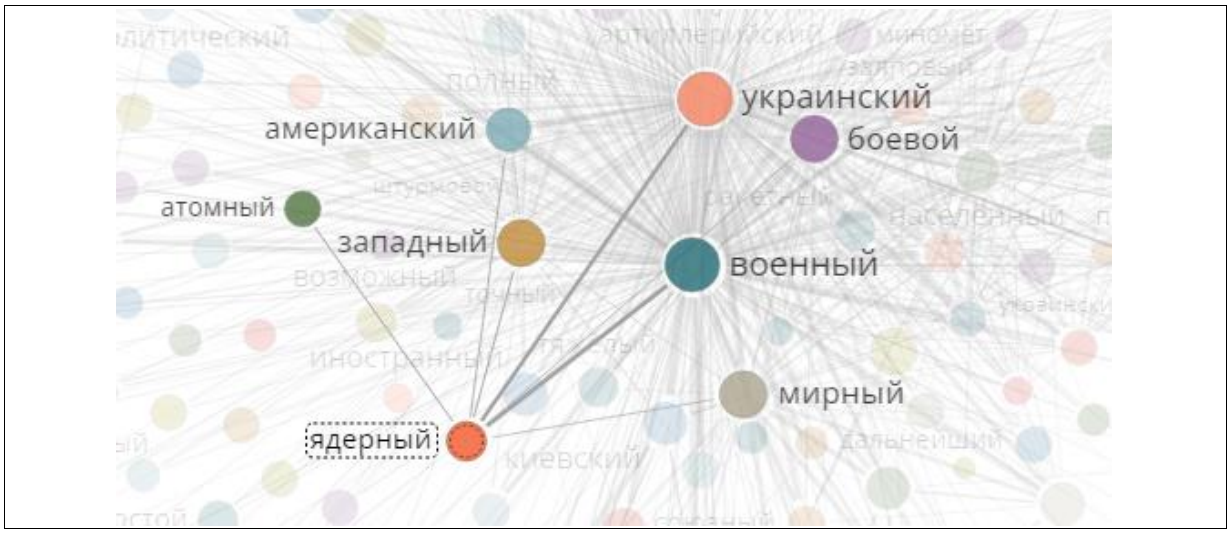



Рис.14. Граф связи ключевых слов со словом ядерный

Ключевое слово *террористический* имеет всего две связи с ключевыми словами *украинский* (поддержка связи 459) и *военный* (поддержка связи 468), но эти связи образованы с самыми значимыми понятиями, выделенными системой (рис. 15). Такие связи поддерживаются текстами, в которых ярко отражается характер и эмоциональная окраска высказывания, например:

*«Сегодня коллективный Запад продолжает игнорировать преступную, **террористическую** суть киевского режима»;*

*«Мы бы столкнулись с проблемами у себя в тылу - Украина стала бы местом **террористической** активности» (выгрузка из системы Polyanalyst);*

*«Обязанность каждого - мониторинг и выявление пособников **террористической** неонацисткой русофобской секты Украина»* [Polyanalyst http].

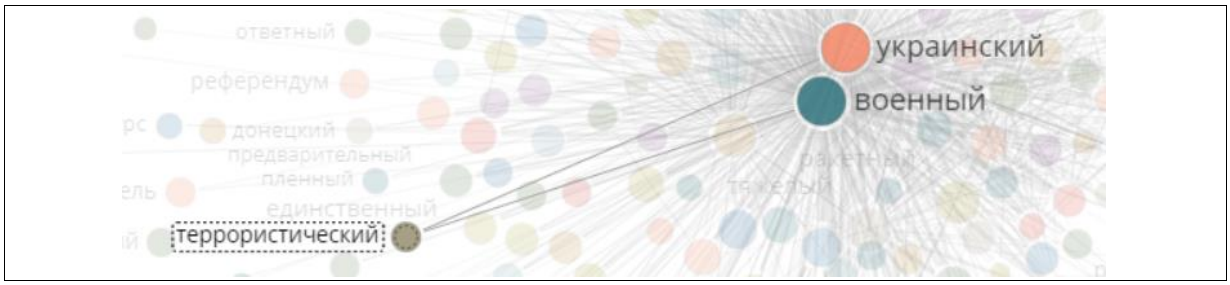



Рис.15. Граф связи ключевого слова террористический

Ключевое слово *диверсант* имеет всего одну связь с ключевым словом *украинский* с поддержкой связи 230 и силой связи 0,08 (рис. 16). Это говорит о том, что в медиапространстве понятие диверсии и диверсантов обсуждается только в контексте, связанном с Украиной, что подтверждается текстами, например:

*«Но разведка с беспилотника вовремя обнаружила врага, а костромские десантники точными ударами 100-миллиметровых пушек боевых машин уничтожили до 20* ***диверсантов****»;*

*«Силы (ВКС) России сбили украинский вертолет Ми-8 с* ***диверсантами*** *в Херсонской области»;*

*«Нам не нужны* ***диверсанты*** *на нашей территории, поэтому на трассах будем дежурить вооружённые МЫ – мужики с оружием на каждом километре»* [Polyanalyst http].


Рис.16. Граф связи ключевого слова диверсант

**Выводы**

Инструменты и функционал системы PolyAnalyst предоставляет обширные возможности для работы с массивами текстов и большими данными. Узел Извлечение сущностей позволяет осуществлять свертывание огромного массива сообщений/текстов для дальнейшего их анализа. Эта функция может быть полезна для определения наиболее частотных персоналий, стран, компаний, военных операций и международных организаций, установления связей между ними, а также отношений к этим сущностям других объектов. В отличие от целей, которые ставит перед собой узел Извлечение ключевых слов, данный узел

определяет не частоту и поддержку извлеченных терминов, а просто выявляет физическое присутствие этих терминов в тексте.

Узел Извлечение ключевых слов позволяет выделить ключевые слова, имеющие большое значение именно для рассматриваемого множества текстов, и анализировать их по таким характеристикам, как значимость, поддержка и частотность, а также связь с другими ключевыми словами с учётом их поддержки и силы связи.

Удобные интерактивные результаты работы узлов позволяют увидеть статистику, визуализацию результатов анализа, использовать поиск и детализацию по самым разным параметрам. Все эти возможности дают неисчерпаемый материал для исследований не только в лингвистическом поле, но и на стыке наук. Для данного исследования видим перспективу во временном трекинге материала, отслеживании изменений уровня агрессивности высказываний и настроений пользователей медиапространства в разные временные отрезки. Подобное исследование позволит проанализировать динамику изменений в отношении пользователей к СВО.

## ***Библиографический список***

*Александрова Ю.К., Лебедкина Н.С., Орлова В.В*. Исследование информационного поля на основе динамического подхода к анализу данных социальной сети «ВКонтакте» (Кейс г. Томск) // Векторы благополучия: экономика и социум. 2021. № 3 (42). С. 33-42.

*Дмитриева Е.В*. Автоматический анализ виртуальной коммуникации специалистов IT-сферы в психолингвистическом аспекте (на материале англоязычных интернет-форумов) // Теория языка и межкультурная коммуникация. 2022. №1 (44). С. 64-77 [Электронный ресурс]. URL: https://tl-ic.kursksu.ru/magazine/archive/number/194 (дата обращения: 17.02.2023).

*PolyAnalyst* – Информационно-аналитическая платформа PolyAnalyst [Электронный ресурс]. URL: https://www.megaputer.com/ru/polyanalyst/ (дата обращения: 20.10.2022).

*Корешкова Т*. Семантический анализ для автоматической обработки естественного языка». 2021 [Электронный ресурс]. URL: https://rdc.grfc.ru/2021/09/semantic_analysis/#post-1707-_Toc69397632 (дата обращения: 10.02.2023)

*Коршунов А., Белобородов И.* Анализ социальных сетей: методы и приложения / Коршунов А., Белобородов И., Бузун Н., Аванесов В., Пастухов Р., Чихрадзе К., Козлов И., Гомзин А. // Институт системного программирования им. В.П. Иванникова РАН. 2014. С.439-456.

*Крибрум* – сервис медиа-аналитики [Электронный ресурс]. URL: https://soware.ru/products/kribrum (дата обращения: 20.10.2022).

*Митягина Е.В., Конышев Е.В., Чернышев К.А.* Цифровые следы выпускников вузов при исследовании миграции из регионов-доноров / Митягина Е.В., Конышев Е.В., Чернышев К.А., Сайфулин Э.Р. // Вестник Томского государственного университета. 2021. № 467. С. 144–155.

*Мошкин В.С.* Алгоритм анализа эмоциональной окраски текстовых ресурсов социальных сетей на основе онтологии / Мошкин В.С., Андреев И.А., Ярушкина Н.Г. // Семнадцатая Национальная конференция по искусственному интеллекту с международным участием: КИИ-2019 (21–25 окт.): сборник научных трудов. Ульяновск: УлГТУ, 2019. Т. 1. С. 171– 184.

*Smetanin S., Komarov M.* Sentiment analysis of product reviews in Russian using convolutional neural networks // Proc. IEEE 21st Conf. Bus. Informat. (CBI). Vol. 1, July 2019. Pp. 482-486.

**References**

Aleksandrova YU.K., Lebedkina N.S., Orlova V.V. Issledovanie informacionnogo polya na osnove dinamicheskogo podhoda k analizu dannyh social'noj seti «VKontakte» (Kejs g. Tomsk) // Vektory blagopoluchiya: ekonomika i socium. 2021. № 3 (42). S. 33-42.

Dmitrieva E.V. Avtomaticheskij analiz virtual'noj kommunikacii specialistov IT-sfery v psiholingvisticheskom aspekte (na materiale angloyazychnyh internet-forumov) // Teoriya yazyka i mezhkul'turnaya kommunikaciya. 2022. №1 (44). S. 64-77 [Elektronnyj resurs]. URL: https://tl-ic.kursksu.ru/magazine/archive/number/194 (data obrashcheniya: 17.02.2023).

PolyAnalyst – Informacionno-analiticheskaya platforma PolyAnalyst [Elektronnyj resurs]. URL: https://www.megaputer.com/ru/polyanalyst/ (data obrashcheniya: 20.10.2022).

Koreshkova T. Semanticheskij analiz dlya avtomaticheskoj obrabotki estestvennogo yazyka». 2021 [Elektronnyj resurs]. URL: https://rdc.grfc.ru/2021/09/semantic_analysis/#post-1707-_Toc69397632 (data obrashcheniya: 10.02.2023)

Korshunov A., Beloborodov I. Analiz social'nyh seteĭ: metody i prilozheniya / Korshunov A., Beloborodov I., Buzun N., Avanesov V., Pastuhov R., CHihradze K., Kozlov I., Gomzin A. // Institut sistemnogo programmirovaniya im. V.P. Ivannikova RAN. 2014. S.439-456.

Kribrum – servis media-analitiki [Elektronnyj resurs]. URL: https://soware.ru/products/kribrum (data obrashcheniya: 20.10.2022).

Mityagina E.V., Konyshev E.V., CHernyshev K.A. Cifrovye sledy vypusknikov vuzov pri issledovanii migracii iz regionov-donorov / Mityagina E.V., Konyshev E.V., CHernyshev K.A., Saĭfulin E.R. // Vestnik Tomskogo gosudarstvennogo universiteta. 2021. № 467. S. 144–155.

Moshkin V.S. Algoritm analiza emocional'noj okraski tekstovyh resursov social'nyh setej na osnove ontologii / Moshkin V.S., Andreev I.A., YArushkina N.G. // Semnadcataya Nacional'naya konferenciya po iskusstvennomu intellektu s mezhdunarodnym uchastiem: KII-2019 (21–25 okt.): sbornik nauchnyh trudov. Ul'yanovsk: UlGTU, 2019. T. 1. S. 171–184.

Smetanin S., Komarov M. Sentiment analysis of product reviews in Russian using convolutional neural networks // Proc. IEEE 21st Conf. Bus. Informat. (CBI). Vol. 1, July 2019. Pp. 482-486.