

УДК 355/359.07

DOI: 10.22213/2410-9304-2025-2-104-111

Система анализа на основе открытых источников текстовых данных на базе отечественной среды визуального программирования PolyAnalyst

К. С. Корнеев, ВИТ «ЭРА», Анапа, Россия

М. С. Спирин, кандидат технических наук, ВИТ «ЭРА», Анапа, Россия

Е. М. Степанов, ВИТ «ЭРА», Анапа, Россия

Д. С. Дерунов, ВИТ «ЭРА», Анапа, Россия

Статья посвящена применению технологий обработки естественного языка (NLP) для усиления анализа открытых источников информации (OSINT) в государственных структурах, ответственных за обеспечение национальной безопасности. В условиях постоянно растущего потока данных из различных онлайн-источников эффективная обработка и анализ информации критически важны для своевременного выявления и предотвращения угроз. Для решения этой задачи используются передовые алгоритмы машинного обучения и искусственного интеллекта, реализованные с применением low-code-принципов. Это позволяет значительно ускорить разработку и внедрение аналитических решений, минимизируя потребность в традиционном программировании и привлекая специалистов с меньшим опытом в области программирования, но обладающих экспертизой в анализе данных и OSINT.

Центральной частью рассматриваемой системы является платформа визуального программирования PolyAnalyst. Ее архитектура, основанная на принципах low-code, предоставляет пользователям интуитивно понятный интерфейс для построения сложных аналитических конвейеров. Система включает в себя специализированные наборы функциональных узлов, предназначенных для выполнения различных этапов обработки данных. Модуль ETL (Extract, Transform, Load) обеспечивает загрузку, очистку и преобразование данных из различных источников. Модуль Data Mining позволяет применять алгоритмы машинного обучения для выявления закономерностей и прогнозирования. Ключевым элементом является модуль TextMining, который использует NLP-алгоритмы для анализа текстовой информации, извлечения ключевых слов, определения настроений и выявления сущностей, представляющих интерес для анализа OSINT. Наконец, модуль BI (Business Intelligence) обеспечивает визуализацию результатов анализа и построение интерактивных отчетов для удобства восприятия и принятия решений. В совокупности, эти модули позволяют создавать мощные и гибкие аналитические системы, способные эффективно обрабатывать большие объемы OSINT-данных и предоставлять ценную информацию для принятия стратегических решений в области национальной безопасности.

Ключевые слова: анализ на основе открытых данных, OSINT, обработка естественного языка, NLP, машинное обучение, искусственный интеллект, low-code, анализ данных, PolyAnalyst.

Введение

События, происходящие в мире, показывают, что обеспечение суверенитета и независимости страны невозможно без внедрения передовых решений и технологий в оборонную сферу. Необходимо мониторить весь спектр зарубежных разработок, технологий и инициатив для нужд ОПК.

Одним из наиболее действенных инструментов достижения и поддержания технологического паритета является OSINT (Open Source INTEllIGENCE) [1, 2] – конкурентная разведка по открытым источникам. В настоящее время значимую часть задач разведки по открытым источникам возможно решить применением технологии обработки текста на естественном языке NLP «NATURAL LANGUAGE PROCESSING» [3, 4].

Информационная война не является чем-то новым для человечества. Наоборот: если пушки

изредка всё же затихали, то словесные баталии между государствами не прекращались никогда.

Россия впутана в целый ряд информационных конфликтов с самыми разными странами, особенно сильно это начало проявляться в 2014 году. Главными противниками в данном противостоянии являются так называемые страны Запада, из которых следует особо выделить США и Великобританию. Помимо собственных СМИ, в своей информационной борьбе с Россией эти страны в последние десятилетия активно используют как спонсируемые ими силы внутри России, так и антироссийски настроенные соседние страны, такие как Украина, Польша, Эстония.

Главной целью информационной войны против России является дестабилизация ситуации внутри страны, в частности организация «оранжевой революции» и других негативных сценариев в России. За рубежом приоритетом являет-

ся ослабление России и порча ее репутации. В частности, Запад навязчиво пытается выставить Россию тиранической, отсталой и агрессивной, при этом агитация направлена как на население России, так и на жителей других стран.

Известно, что на сегодняшний день страны коллективного Запада ведут активную работу по созданию систем поддержки командных решений. Уже созданные и применяемые странами Запада программные решения позволяют осуществлять сбор информации с территории Российской Федерации.

Разработка отечественной аналитической системы на основе OSINT-технологии обусловлена необходимостью оперативного реагирования на действия противника с целью анализа обстановки [5]. Создание подобной системы позволяет обеспечить оперативное информирование, а также значительно упростит работу аналитических групп правительственные структур, занимающихся обеспечением безопасности государства.

Для решения данных задач была рассмотрена отечественная аналитическая платформа PolyAnalyst [6]. Использование данной программной среды позволило разработать проект для анализа по открытым источникам сети интернет «Оракул».

На рис. 1 представлена архитектура платформы PolyAnalyst.

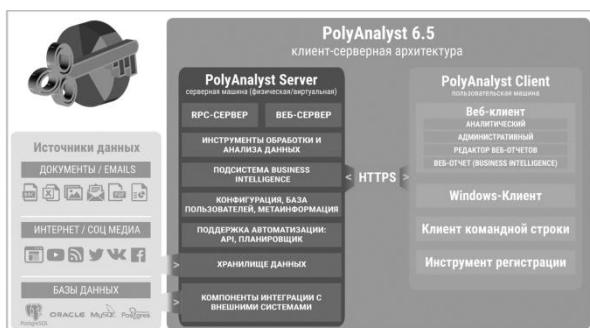


Рис. 1. Архитектура программы PolyAnalyst

Fig. 1. PolyAnalyst software architecture

Платформа PolyAnalyst 6.5 построена на основе клиент-серверной архитектуры, что обеспечивает работу системы как программного обеспечения корпоративного (enterprise) уровня. Это позволяет проведение совместной работы множества пользователей над одними аналитическими проектами и совместное использование различных связанных ресурсов, таких как сценарии анализа, словари, таксономии и многомерные матрицы. Также клиент-серверная архитектура обеспечивает высокую производитель-

ность системы за счет выполнения расчетов на наиболее мощных вычислительных ресурсах, сокращает нагрузку передаваемого трафика на корпоративную сеть, позволяет осуществлять отложенное на заданное время выполнение задач и создавать пользовательские отчеты с возможностью оповещения о наступлении событий (инцидентах) для разных групп бизнес-пользователей.

Доступно централизованное управление и контроль списка пользователей, их прав и действий в системе. Клиентский доступ осуществляется через запуск приложения в ОС, или через «тонкий клиент» в веб-браузере.

В системе PolyAnalyst представлены модели NLP для обработки и анализа текста.

Система интеллектуального анализа текста органично сочетает в себе передовые инструменты обработки естественного языка, средства семантического текстового анализа и алгоритмы машинного обучения, что позволяет пользователям получать результаты высокой точности. Мы предлагаем вам:

- средства для загрузки данных из всех популярных источников и документов разного формата;
- эффективные инструменты Text Mining для очистки и преобразования данных;
- непревзойденную систему интеллектуального анализа текста, включающую в себя:
 - инструменты лингвистического (лексического, морфологического и синтаксического) анализа;
 - поддержку онтологий и семантических словарей;
 - статистические инструменты и алгоритмы машинного обучения.

В данной работе представлены варианты использования NLP для анализа текста и формирования системы сбора, обработки и отображения требуемой информации.

Также в PolyAnalyst возможно использовать высокоуровневые языки программирования Python и R для выполнения специальных операций с данными:

- прогнозирование;
- регрессионный анализ;
- анализ статистических распределений;
- кластеризация;
- анализ покупательских корзин;
- деревья решений;
- анализ временных рядов;
- классификация;
- анализ связей;
- разрешение сущностей.

«Оракул» – анализ на основе открытых источников

На данный момент проект «Оракул» представляет собой гибкую систему, созданную в отечественной аналитической среде визуального программирования PolyAnalyst, способную в короткие сроки переориентироваться под обработку любой информации. В рамках проекта решались задачи сбора, обработки и отображения информации, получаемой из открытых источников сети Интернет, а также задачи обра-

ботки доверенных источников для получения полезной и достоверной информации [7].

Первоначально происходит поиск и выгрузка информации из открытых источников затем в основной части проекта данные обрабатываются и структурируются, после этого происходит формирование интерактивного веб-отчета для анализа результатов.

В текущей версии разработанный сценарий, частично представленный на рис. 2, состоит более чем из 300 функциональных узлов.

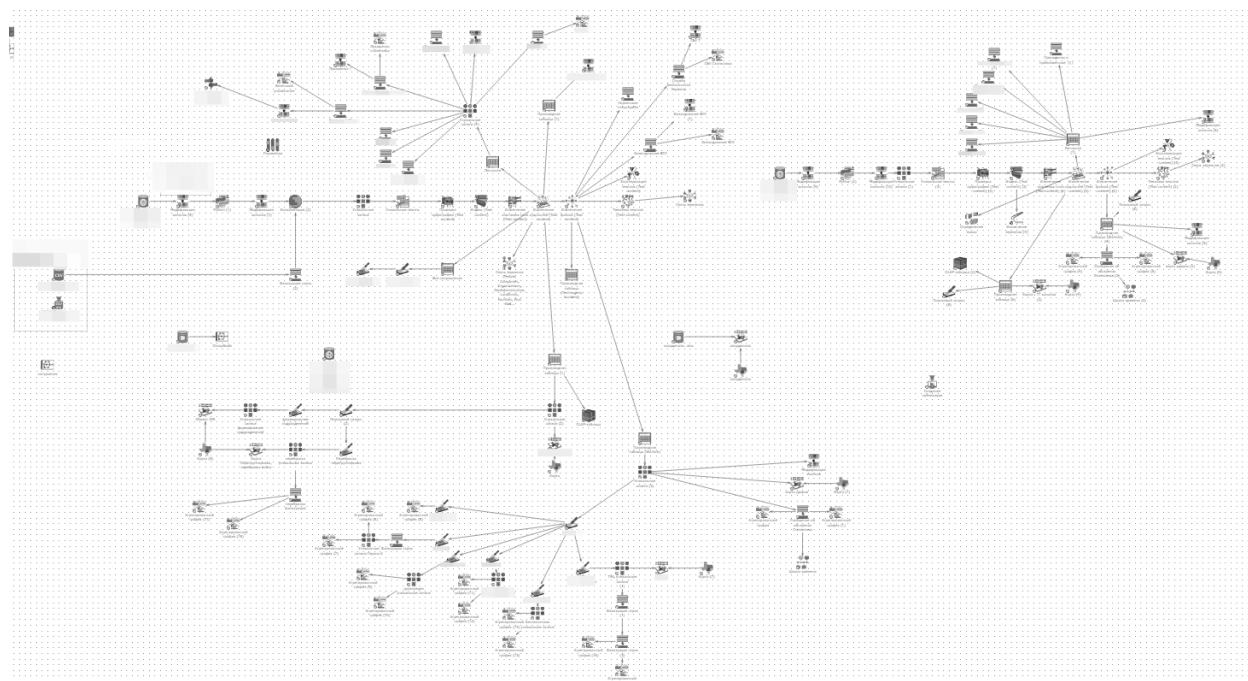


Рис. 2. Визуальный сценарий в программной среде PolyAnalyst

Fig. 2. Visual scenario in the PolyAnalyst software environment

Анализ интерактивных карт

Первое направление включает в себя четыре интерактивных веб-отчета:

- 1) интерактивную карту с отображением событий по населенным пунктам;
- 2) интерактивную карту с отображением населенных пунктов, по которым наносились удары разного рода;
- 3) интерактивную карту с отображением населенных пунктов, в которых были зафиксированы факты повреждения объектов инфраструктуры, энергосистемы и ТЭЦ;
- 4) интерактивную карту с отображением перемещения подразделений по населенным пунктам.

Все вышеприведенные веб-отчеты состоят из интерактивных карт, отображающих населенные пункты и интенсивность их упоминания

в информационном пространстве с помощью размера круговых зон. Примеры веб-отчетов представлены на рис. 3 и 4.

Также на данных веб-отчетах отображаются разного рода диаграммы и графики, позволяющие отслеживать статистические данные и динамику их изменения [8, 9]. Помимо этого, неотъемлемым атрибутом всех отчетов является зона с отображением текстовой информации из открытых источников сети Интернет, а именно Телеграм-каналов, по интересующим нас населенным пунктам в рамках поставленных задач, с отображением даты и источника публикации [10, 11]. Каждый веб-отчет поддерживает возможность отслеживания информации в конкретные даты или промежутки времени (годы, месяцы, часы, минуты, секунды). Примеры веб-отчетов представлены на рис. 2 и 3.



*Rис. 3. Веб-отчет с отображением информации о количестве событий в населенных пунктах
Fig. 3. Web report displaying information about occasion on populated areas*



*Rис. 4. Веб-отчет с отображением информации о перемещении подразделений
Fig. 4. Web report displaying information about the transfer of units*

Анализ личностей и связей

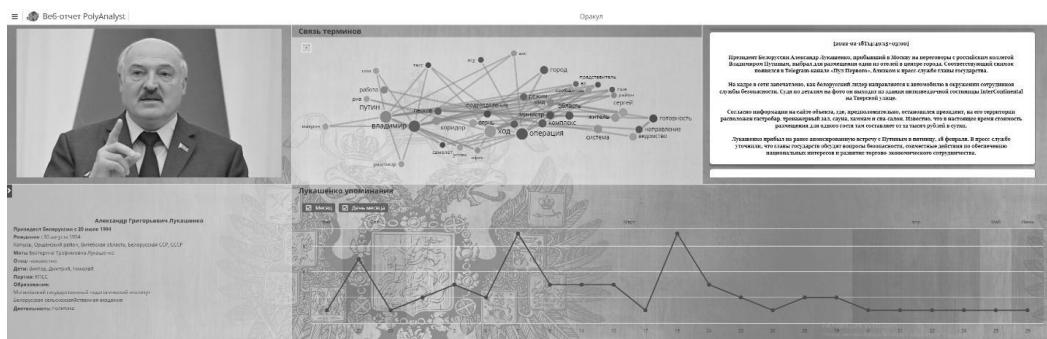
Второе направление включает в себя четыре вида интерактивных веб-отчетов:

1. Общая информация о выявленных в информационном пространстве личностях – должность, пол, количество упоминаний в источниках, текст и контекст упоминания.

2. Интерактивное дерево графов, отображающее наиболее значимые связи между организациями, государственными структурами, странами и личностями, выявленные в отслеживаемом информационном пространстве.

3. Персональные информационные карты личностей, государственных и частных структур, организаций. Данный тип веб-отчетов включают в себя персональное дерево графов, статистические данные, краткое досье со ссылками на внешние и внутренние источники, а также окно, отображающее сообщения из отслеживаемого информационного источника, включающее упоминание интересующего объекта.

Пример отображения данного вида веб-отчетов приведен на рис. 5.



*Rис. 5. Веб-отчет с отображением персональной карты интересуемого объекта, созданной на основе обработки информации из открытых источников сети Интернет
Fig. 5. Web – report displaying a personal map of the object of interest, created based on processing information from open sources on the Internet*

4. Отслеживание информации о повреждении инфраструктуры. Данный тип веб-отчетов включает в себя отображение статистики появления, перемещения и упоминания повреждения объектов инфраструктуры по населенным пунктам интересующей нас области.

Примеры данного вида веб-отчетов приведены на рис. 6.



Рис. 6. Веб-отчет с отображением детальной информации по повреждению ТЭЦ и другой инфраструктуры, созданный с помощью обработки информации, находящейся в открытых источниках сети Интернет

Fig. 6. Web-report displaying detailed information on damage to thermal power plants and other infrastructure, created by processing information found in open sources on the Internet

На рис. 6 представлен веб-отчет, отображающий информацию о поврежденных ТЭЦ и другой инфраструктуре. Круговая диаграмма позволяет нам увидеть количество поврежденных объектов инфраструктуры в различных городах. Временная диаграмма предоставляет статистику за рассматриваемый промежуток времени. Рассматриваемый промежуток времени изменяется с помощью диапазонного фильтра, что дает возможность провести более подробный анализ.

Анализ тональностей и настроений

Третье направление ориентировано на оценку тональностей сообщений и комментариев к ним, а также сравнение настроений и реакции на одни и те же темы в пророссийской и проза-

падной информационной среде. Данное направление включает в себя три типа веб-отчетов:

1. Отображение текста сообщений пророссийских и прозападных источников, а также сравнение закрепленных к ним комментариев.

2. Статистические и аналитические данные, отображающие количество и тональности сообщений по населенным пунктам, с отображением информации на интерактивной карте и текстовых табличках. Пример данного варианта веб-отчетов приведен на рис. 7.

3. Дерево графов и облако тегов, отображающие связь между объектами интереса и ключевыми словами, формирующими настроение в информационном пространстве.



Рис. 7. Веб-отчет отображающий статистику тональностей информационного пространства по областям и населенным пунктам, созданный на основе обработки информации из открытых источников сети Интернет

Fig. 7. Web report displaying statistics of sentiments in the information space by region and settlement, created based on processing information from open sources on the Internet

По результатам работы описанных модулей OSINT проекта «Оракул» можно утверждать, что применение отечественных аналитических платформ визуального программирования явля-

ется эффективным решением в области разработки программного обеспечения. В частности, платформа PolyAnalyst позволяет создавать эффективные и конкурентоспособные програм-

ные решения в области обработки и аналитики больших данных на естественном языке, значительно упрощающие работу аналитических групп [12, 13].

Заключение

На данный момент на рынке существуют решения от отечественных компаний Avalanche, «Крибрум» и американской компании Palantir. Представленные семейства продуктов являются коммерческими решениями, не предоставляющими исходный код, вследствие чего невозможно добиться качественного обеспечения информационной безопасности [14]. Данные программные продукты предоставляются заказчику по подписке, с начальной стоимостью от нескольких миллионов рублей в месяц за серверную лицензию и неполный набор доступного функционала.

Рассмотренный в статье проект «Оракул» позволяет обрабатывать открытые источники сети интернет с оперативным отображением результатов обработки в виде графиков, деревьев связи, текста и интерактивных карт местности, а также предоставлять оценки тональности на основе текстовых сообщений и комментариев к ним. Проект «Оракул», наглядно показывает эффективность применения отечественного программного обеспечения – PolyAnalyst, в основе которого лежит принцип Low-code [15] разработки.

Данное программное решение призвано повысить оперативность и осведомленность потребителей в правительственные и коммерческих структурах, что в глобальном плане позволит достичь повышения эффективности управления, а в локальном – позволит получать и оценивать любые сведения в режиме реального времени. Одним из главных потенциальных преимуществ системы перед конкурентами является необходимость приобрести только программную среду разработки, а все созданные в ней продукты будут являться собственностью организаций и правительственные структур РФ, что в свою очередь значительно снизит потенциальные расходы и обеспечит повышение информационной безопасности.

Библиографические ссылки

1. LebidAndrii E., StepanovVitalii V., NazarovMykola S. Use of the OSINT-technologies for civil society institutions // International Journal of Media and Information Literacy . 2023. № 1. URL: <https://cyberleninka.ru/article/n/use-of-the-osint-technologies-for-civil-society-institutions> (дата обращения: 17.01.2025).
2. Воложанина Д. Н., Жукова М. Н. Пересмотр процедуры оценки защищенности методами OSINT в условиях нестабильной международной политической обстановки // Актуальные проблемы авиации и космонавтики. 2022. URL: <https://cyberleninka.ru/article/n/peresmotr-protsedury-otsenki-zashchitnosti-metodami-osint-v-usloviyah-nestabilnoy-mezhdunarodnoy-politicheskoy-obstanovki> (дата обращения: 17.01.2025).
3. Gruetzemacher R. The Power of Natural Language Processing. – Harvard Business Review, 2022. URL: <https://hbr.org/2022/04/the-power-of-natural-language-processing> (дата обращения: 01.02.2022).
4. Zhang Zhipeng, Poguda Aleksey Research on the development of data augmentation techniques in the field of machine translation // International Journal of Open Information Technologies. 2023. № 5. URL: <https://cyberleninka.ru/article/n/research-on-the-development-of-data-augmentation-techniques-in-the-field-of-machine-translation> (дата обращения: 17.01.2025).
5. Смирнов В. И., Новоселова О. В. Обзор современных методов анализа больших данных для различных предметных областей // Вестник науки. 2024. № 6 (75). URL: <https://cyberleninka.ru/article/n/obzor-sovremennoy-metodov-analiza-bolshih-dannyh-dlya-razlichnyh-predmetnyh-oblastey> (дата обращения: 18.01.2025).
6. Саркисова А. Ю. Ресурс аналитической платформы polyanalyst в социогуманитарных научных исследованиях. Томск : Издательство Томского государственного университета, 2021. 12 с. URL: <https://vital.lib.tsu.ru/vital/access/services/Download/ko-ha:000897657/SOURCE1>
7. Соломонов А. А. Оптимизация ETL-процессов для больших данных // Вестник науки. 2024. № 9 (78). URL: <https://cyberleninka.ru/article/n/optimizatsiya-etl-protsessov-dlya-bolshih-dannyh> (дата обращения: 17.01.2025).
8. Русакович А. С. Интеллектуальный анализ данных как инструмент поддержки принятия решений // Современные инновации. 2022. № 1 (41). URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-kak-instrument-podderzhki-prinyatiya-resheniy> (дата обращения: 17.01.2025).
9. Кучкаров Т. С. О методах и инструментах анализа больших данных // Экономика и социум. 2023. №12 (115)-2. URL: <https://cyberleninka.ru/article/n/o-metodah-i-instrumentah-analiza-bolshih-dannyh> (дата обращения: 18.01.2025).
10. Статистическая модель поиска целевых объектов в социальной сети / Д. И. Сафиканов, А. А. Артамонов, Ю. Е. Фомина, А. И. Черкасский // International Journal of Open Information Technologies. 2024. № 10. URL: <https://cyberleninka.ru/article/n/statisticheskaya-model-poiska-tselevykh-obektov-v-sotsialnoy-seti> (дата обращения: 18.01.2025).
11. Арланова А. А., Нобатов А. М. Интеллектуальный анализ данных: виды и методы // Вестник науки. 2023. № 1 (58). URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-vidy-i-metody> (дата обращения: 18.01.2025).
12. Карпов Д. К. Обработка больших данных с использованием средств языка Python // StudNet. 2021. № 6. URL: <https://cyberleninka.ru>

/article/n/obrabotka-bolshih-dannyh-s-ispolzovaniem-sredstv-yazyka-python (дата обращения: 18.01.2025).

13. Басина П. А., Дунаева Д. О., Саркисова А. Ю. Валидация моделей машинного обучения для автоматизированного определения тональности русскоязычных текстов // Вестн. Том. гос. ун-та. 2022. №485.

URL: <https://cyberleninka.ru/article/n/validatsiya-modeley-mashinnogo-obucheniya-dlya-avtomatizirovannogo-opredeleniya-tonalnosti-russkoyazychnykh-tekstov> (дата обращения: 20.01.2025).

14. Статьев В. Ю., Докучаев В. А., Маклакова В. В. Информационная безопасность на пространстве «больших данных» // Т-Comm. 2022. № 4. URL: <https://cyberleninka.ru/article/n/informatsionnaya-bezopasnost-na-prostranstve-bolshih-dannyh> (дата обращения: 20.01.2025).

15. Магомадов В. С. Платформы LOW-CODE и NO-CODE как способ сделать программирование более доступным для широкой общественности // МНИЖ. 2021. № 6-1 (108). URL: <https://cyberleninka.ru/article/n/platformy-low-code-i-no-code-kak-sposob-sdelat-programmirovaniye-bolee-dostupnym-dlya-shirokoy-obschestvennosti> (дата обращения: 17.01.2025).

References

- Zhang Z., Poguda A. Research on the development of data augmentation techniques in the field of machine translation. In International Journal of Open Information Technologies, 2023, no. 5. URL: <https://cyberleninka.ru/article/n/research-on-the-development-of-data-augmentation-techniques-in-the-field-of-machine-translation> (accessed January 17, 2025).
- Volozhanina D.N., Zhukova M.N. [Review of the procedure for assessing security using OSINT methods in the conditions of an unstable international political environment]. Aktualnye problemy aviatsii i kosmonavtiki, 2022 (in Russ.). URL: [https://cyberleninka.ru/article/n/peresmotr-protcedury-otsenki-zashchitschennosti-metodami-osint-v-usloviyah-nestabilnoy-mezhdunarodnoy-politicheskoy-obstanovki](https://cyberleninka.ru/article/n/peresmotr-protsedury-otsenki-zashchitschennosti-metodami-osint-v-usloviyah-nestabilnoy-mezhdunarodnoy-politicheskoy-obstanovki) (accessed January 17, 2025).
- Lebid A.E., Stepanov V.V., Nazarov M.S. Use of OSINT technologies for civil society institutions. In International Journal of Media and Information Literacy, 2023, no. 1. URL: <https://cyberleninka.ru/article/n/use-of-the-osint-technologies-for-civil-society-institutions> (accessed January 17, 2025).
- Yakovlev G.S., Ivanov F.F. [Use of low-code platforms in the transition to a process approach in the creation of automated systems]. Vestnik KRaUNC. Fiz.-mat. nauki, 2020, no. 1. URL: <https://cyberleninka.ru/article/n/ispolzovanie-low-code-platform-pri-perehode-na-protsessnyy-podhod-v-sozdaniy-avtomatizirovannyh-sistem> (accessed January 17, 2025).
- Magomadov V.S. [Low-code and no-code platforms as a way to make programming more accessible to the wider public]. MNIZH. 2021. No. 6-1 (in Russ.). Available at: <https://cyberleninka.ru/article/n/platformy-low-code-i-no-code-kak-sposob-sdelat-programmirovaniye-bolee-dostupnym-dlya-shirokoy-obschestvennosti> (accessed January 17, 2025).
- Rusakovich A.S. [Intelligent data analysis as a decision support tool]. Sovremennye innovatsii, 2022, no. 1 (in Russ.). Available at: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-kak-instrument-podderzhki-prinyatiya-resheniy> (accessed January 17, 2025).
- Solomonov A.A. [Optimization of ETL processes for big data]. Vestnik nauki, 2024, no. 9 (in Russ.). Available at: <https://cyberleninka.ru/article/n/optimizatsiya-etl-protsessov-dlya-bolshih-dannyh> (accessed January 17, 2025).
- Rusakovich A.S. [Data mining as a decision support tool]. Sovremennye innovacii. 2022. No. 1 (in Russ.). Available at: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-kak-instrument-podderzhki-prinyatiya-resheniy> (accessed: 17.01.2025).
- Kuchkarov T.S. [On the methods and tools of big data analysis]. Ekonomika I sotsium. 2023. No. 12 (in Russ.). Available at: <https://cyberleninka.ru/article/n/o-metodah-i-instrumentah-analiza-bolshih-dannyh> (accessed January 18, 2025).
- Safikanov D.I., Artamonov A.A., Fomina Yu.E., & Cherkassky, A. I. [Statistical model for searching target objects in a social network]. International Journal of Open Information Technologies, 10. 2024 (in Russ.). Available at: <https://cyberleninka.ru/article/n/statisticheskaya-model-poiska-tselevykh-obektov-v-sotsialnoy-seti> (accessed: 18.01.2025).
- Arlanova A.A., & Nobatov A.M. [Intelligent data analysis: Types and methods]. Vestnik nauki. 2023. No. 1 (in Russ.). Available at: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-vidy-i-metody> (accessed: 18.01.2025).
- Karpov D.K. [Big data processing using Python]. student. 2021. No. 6 (in Russ.). Available at: <https://cyberleninka.ru/article/n/obrabotka-bolshih-dannyh-s-ispolzovaniem-sredstv-yazyka-python> (accessed: 18.01.2025).
- Basina P.A., Dunayeva D.O., & Sarkisova A.Yu. [Validation of machine learning models for automated determination of the tonality of Russian-language texts]. Vestnik Tomskogo gosudarstvennogo universiteta. 2022. No. 485 (in Russ.). Available at: <https://cyberleninka.ru/article/n/validatsiya-modeley-mashinnogo-obucheniya-dlya-avtomatizirovannogo-opredeleniya-tonalnosti-russkoyazychnykh-tekstov> (accessed: 18.01.2025).
- Statyev V.Yu., Dokuchaev V.A., & Maklachkova V.V. [Information security in the "big data" space]. T-Comm. 2022. No. 4 (in Russ.). Available at: <https://cyberleninka.ru/article/n/informatsionnaya-bezopasnost-na-prostranstve-bolshih-dannyh> (accessed: 20.01.2025).
- Magomadov V.S. [LOW-CODE and NO-CODE platforms as a way to make programming more accessible to the general public]. MNIZH. 2021. No. 6-1 (in Russ.). Available at: <https://cyberleninka.ru/article/n/platformy-low-code-i-no-code-kak-sposob-sdelat-programmirovaniye-bolee-dostupnym-dlya-shirokoy-obschestvennosti> (accessed: 17.01.2025).

Open-source Text Data Intelligence System Based on the Domestic Visual Programming Environment PolyAnalyst

K. S. Korneev, Junior Researcher, Military Innovation Technopolis«ERA», Anapa, Russia

M. S. Spirin, Head of Department, Military Innovation Technopolis «ERA», Anapa, Russia

E. M. Stepanov, Senior Operator, Military Innovation Technopolis«ERA», Anapa, Russia

D. S. Derunov, Senior Operator, Military Innovation Technopolis«ERA», Anapa, Russia

The article is devoted to the application of natural language processing (NLP) technologies to enhance open-source intelligence (OSINT) analysis in government agencies responsible for ensuring national security. With the ever-increasing data flow from various online sources, efficient information processing and analysis is critical for timely detection and prevention of threats. To solve this task, advanced machine learning and artificial intelligence algorithms are implemented using low-code principles. This allows significant acceleration of the development and implementation of analytical solutions, minimizing the need for traditional programming and involving specialists with less programming experience, but having expertise in data analysis and OSINT.

The central part of the system under consideration is the PolyAnalyst visual programming platform. Its architecture, based on low-code principles, provides users with an intuitive interface for building complex analytical pipelines. The system includes specialized sets of functional nodes designed to perform various stages of data processing. The ETL (Extract, Transform, Load) module provides loading, cleaning and transforming data from various sources. The DataMining module enables the application of machine learning algorithms for pattern discovery and prediction. A key element is the TextMining module, using NLP algorithms to analyze textual information, extract keywords, determine sentiment, and identify entities of interest for OSINT analysis. Finally, the BI (Business Intelligence) module provides analysis result visualization and builds interactive reports for easy perception and decision making. Altogether, these modules enable the creation of powerful and flexible analytical systems that can process large volumes of OSINT data efficiently and provide valuable information for strategic national security decision-making.

Keywords: Open data intelligence, OSINT, natural language processing, NLP, machine learning, artificial intelligence, low-code, data analysis, PolyAnalyst.

Получено: 19.08.24

Образец цитирования

Система анализа на основе открытых источников текстовых данных, на базе отечественной среды визуального программирования – PolyAnalyst / К. С. Корнеев, М. С. Спирин, Е. М. Степанов, Д. С. Дерунов // Интеллектуальные системы в производстве. 2025. Т. 23, № 2. С. 104–111. DOI: 10.22213/2410-9304-2025-2-104-111.

For Citation

Korneev K.S., Spirin M.S., Stepanov E.M., Derunov D.S. [Open-source Text Data Intelligence System Based on the Domestic Visual Programming Environment PolyAnalyst]. *Intellektual'nye sistemy v proizvodstve*. 2025, vol. 23, no. 2, pp. 104-111. DOI: 10.22213/2410-9304-2025-2-104-111.