

СИСТЕМА ИНФОРМАЦИОННОГО МОНИТОРИНГА КОНТРАГЕНТОВ

Д. Л. Кузьмин¹ [0000-0003-0011-6292], К. А. Григорян² [0000-0001-6470-1832]

Институт информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета. Казань, ул. Кремлевская, 35

¹to_kdima@mail.ru, ²karigri@yandex.ru

Аннотация

В условиях всевозрастающих информатизации, автоматизации и цифровизации бизнеса возникают новые схемы недобросовестных действий со стороны как юридических, так и физических лиц. В связи с этим остро встает проблема быстрого, эффективного и качественного выявления информации о потенциальном либо действующим контрагенте, решение которой позволит оперативно принять правильные управленческие решения.

В статье предложен один из способов решения данной проблемы – разработка системы информационного мониторинга контрагентов, которая позволит оперативно выявлять и анализировать информацию об их деятельности.

Ключевые слова: разработка системы информационного мониторинга контрагентов, технологии сбора данных из открытых источников, анализ данных с помощью моделей машинного обучения.

ВВЕДЕНИЕ

Быстрое развитие информационных технологий способствуют глобальному росту экономики. Цифровая экономика позволяет представителям бизнеса оперативно получать сведения о своих покупателях, поставщиках и партнёрах. Заключение сделок с поставщиками и заказчиками онлайн, осуществление платежей, распространение рекламы, предоставление некоторых видов услуг с использованием ИТ – всё это увеличивает скорость совершения сделок, товарного и денежного оборота.

В настоящее время значительная часть предприятий, вне зависимости от их размера, сферы деятельности и расположения, использует информационные технологии в своей деятельности. Однако в условиях цифровой экономики бизнес-

сообщество окружают различные фирмы и предприятия, которые не всегда являются надежными. Каждый субъект бизнеса хочет быть уверен в своих контрагентах, обезопасить свою деятельность и минимизировать риски. Для этого необходимо проверять потенциального партнёра перед заключением сделки, особенно когда договора предполагают крупные суммы.

На наш взгляд, для решения названной проблемы необходимо разработать систему, позволяющую провести анализ финансово-хозяйственной деятельности контрагента на текущий момент времени, а также на периодической основе получать оповещения о важных изменениях в его деятельности. Разработка данной системы позволит облегчить работу специалистов, осуществляющих сбор данных о людях, предприятиях, тенденциях и явлениях с помощью ресурсов интернета.

1. ОБЗОР ЛИТЕРАТУРЫ

Информационно-аналитическая система «Спарк»

На сегодняшний день имеется множество различных информационно-аналитических систем (ИАС) мониторинга контрагентов. Одной из наиболее популярных и известных ИАС является система «Спарк», разработанная компанией Интерфакс в 2004 г. Система является профессиональным решением для проверки контрагентов, управления кредитными и налоговыми рисками, маркетинга, инвестиционного анализа и поиска аффилированности. Данное решение занимает лидирующие позиции на рынке и активно используется службами безопасности банков и организаций [8, 14].

Информационно-аналитическая система «Контур.Фокус»

Другой популярной системой проверки контрагентов выступает система Контур.Фокус, разработанная СКБ Контур. Система Контур.Фокус позволяет найти всю необходимую информацию об организации или контрагенте через официальные источники информации, такие как ФНС России, Картотека арбитражных дел, Росстат, ФССП и т. д. [13].

Информационно-аналитическая система «Глобас»

Следующей ИАС проверки контрагентов является система «Глобас» от компании Creditinform. Так же, как ее ближайшие аналоги, она позволяет проверять контрагентов по всем основным источникам: сайтам Арбитражного суда, судебных приставов, налоговой, позволяет установить аффилированность и косвенные

связи, получать актуальные сведения о негативных проявлениях в финансово-хозяйственной деятельности контрагентов [12].

Сравнительный анализ рассмотренных информационно-аналитических систем

Все вышерассмотренные ИАС разнообразны и имеют свои преимущества и недостатки. При выборе ИАС необходимо обращать внимание на то, какие конкретно цели необходимо достичь при взаимодействии с ними, какой масштаб работы будет покрыт и каким бюджетом располагает организация, желающая внедрить их в текущие бизнес-процессы.

Общими недостатками всех систем является:

- отсутствие функционала проверки по социальным сетям;
- отсутствие функционала проверки по сайту Росфинмониторинга (список экстремистов и террористов);
- отсутствие у более дешевых аналогов аналитической обработки полученной информации.

Обзор технологий для сбора и анализа данных из открытых источников

Система информационного мониторинга контрагентов позволяет осуществлять сбор и анализ данных из открытых источников интернета. В настоящее время существуют различные технологии и библиотеки для автоматизации сбора и анализа данных из открытых источников. Рассмотрим некоторые из них.

К технологиям сбора данных можно отнести:

- сбор данных с использованием API,
- сбор данных с помощью семантического разбора веб-страниц,
- сбор данных с помощью эмуляции поведения человека в браузере [17].

Сбор данных через API позволяет осуществлять http-запросы к сайту, где синтаксис запросов и тип возвращаемых данных строго определены на стороне самого сайта.

Сбор данных с помощью семантического разбора веб-страниц представляет собой последовательный семантический анализ информации, размещенных на веб-страницах.

Сбор данных с помощью эмуляции поведения человека в браузере позволяет в точности имитировать взаимодействие человека с веб-приложением.

Одним из популярных в настоящее время способов анализа полученной информации является применение методов машинного обучения и интеллектуального анализа данных. Данные методы позволяют решать задачи для принятия решений в различных сферах деятельности путем обучения алгоритмов на примере решения множества схожих задач [15].

2. ОБЩЕЕ ОПИСАНИЕ ПРОЦЕССА РАБОТЫ ПРИЛОЖЕНИЯ

Чтобы оперативно и качественно разработать программное обеспечение, необходимо понимать логику его работы. В связи с этим нами спроектирована логика работы системы информационного мониторинга контрагентов. Общее описание процесса работы этого приложения представлено на Рис. 1.

Как видно из рисунка, пользователю доступны функции проверки как юридического, так и физического лица. После ввода данных, необходимых для осуществления запроса, проходит проверка наличия либо отсутствия информации о субъекте поиска в базе данных. В случае отсутствия данных в базе формируется запрос на их получение, далее следует проверка готовности парсеров. Если какая-либо задача уже выполняется, то следующая за ней задача ставится в очередь. В случае, когда парсеры находятся в статусе «готов», происходит получение данных с сайтов Судебных приставов, Федеральной налоговой службы, Арбитражного суда, Социальных сетей, реестра банкротов, реестра недобросовестных поставщиков, сайта Росфинмониторинга. Полученная информация анализируется с помощью аналитического модуля системы, заносится в базу данных, и далее результаты выводятся на экран.

При получении запроса от пользователя в том случае, когда субъект поиска уже содержится в базе данных, пользователю на экран выводится актуальная информация из базы по данному контрагенту.

В целях актуализации информации, находящейся в базе данных на сервере, настроена задача на обновление. Формируется запрос на получение данных, проверяется готовность парсеров, осуществляется парсинг данных, и полученные обновления заносятся в базу.

Полученные результаты предоставляют пользователю возможность подписаться на мониторинг контрагента либо отписаться от него.

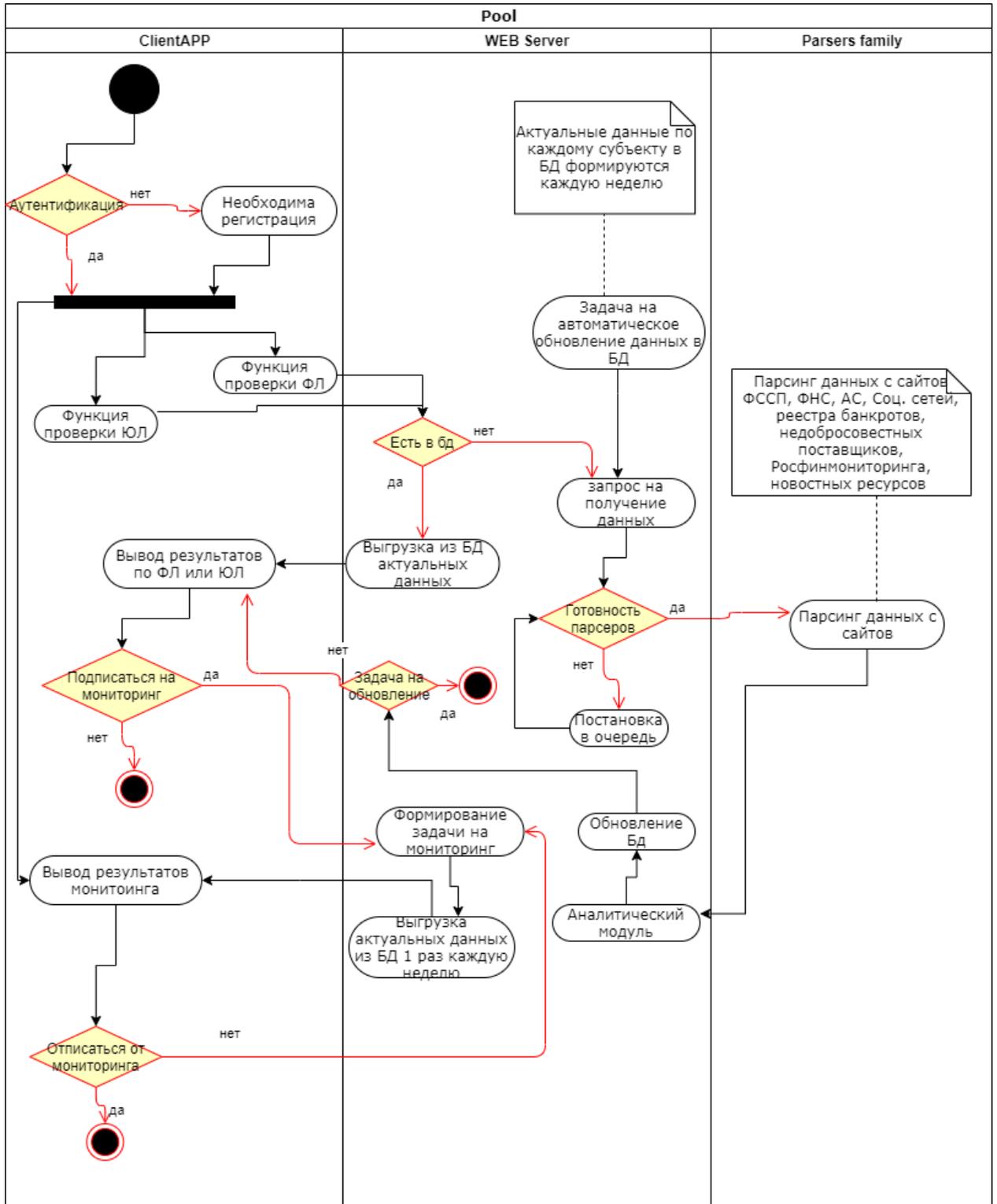


Рис. 1. Описание процесса работы системы информационного мониторинга контрагентов банка.

3. ТЕХНОЛОГИИ СБОРА ДАННЫХ ИЗ ОТКРЫТЫХ ИСТОЧНИКОВ

В данном разделе описаны технологии сбора данных, которые использовались при разработке прототипа системы информационного мониторинга контрагентов.

Как уже отмечалось ранее, источниками данных системы информационного мониторинга контрагентов являются следующие сайты: Федеральная налоговая служба, Служба судебных приставов, Арбитражный суд, Росфинмониторинг, социальная сеть Вконтакте, Федеральный ресурс по банкротству и Реестр недобросовестных поставщиков. В целях сбора данных с вышеуказанных источников, использовались следующие технологии.

1. Технология парсинга данных с использованием модулей, позволяющих эмулировать поведение пользователя в браузере

В настоящее время существует большое количество модулей и библиотек для разных языков программирования, позволяющих эмулировать поведение пользователя в браузере. Одной из популярных библиотек является Selenium WebDriver – программная библиотека для управления браузерами [17]. С использованием данной библиотеки нами осуществлена разработка парсера данных с сайта Росфинмониторинга. В общем виде процесс сбора данных с указанного сайта представлен на Рис. 2.

Росфинмониторинг

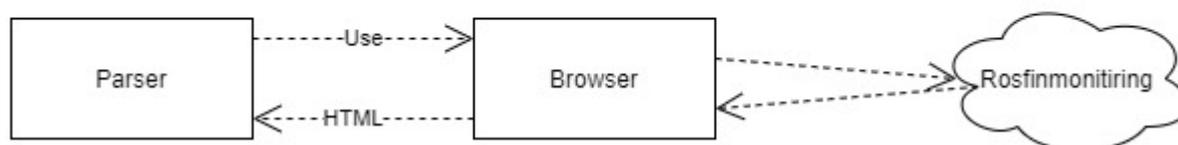


Рис. 2. Схема сбора данных с сайта Росфинмониторинга

Для получения нужных данных необходимо с помощью метода `get()` осуществить запрос к сайту Росфинмониторинга `'https://www.fedsfm.ru/documents/terr-list'`. Далее с помощью метода `find_element_by_name()` нужно обратиться к поисковой строке и, используя метод `send_keys()`, передать в нее ФИО и Дату Рождения искомого субъекта. Метод `find_element_by_name().click()` позволяет нажать

на кнопку поиска информации по списку. Следующее действие – возврат полученной информации с использованием метода `find_element_by_xpath()` и выведение ее на экран с помощью функции `print()`.

Применение названной технологии позволило нам получать всю необходимую информацию с ресурсов Федеральной налоговой службы, Арбитражного суда, Росфинмониторинга, Федерального ресурса по банкротству и Реестра недобросовестных поставщиков.

2. Сбор данных с помощью API

Другой технологией, которая использовалась при разработке прототипа системы, является получение необходимой информации путем осуществления запросов на сайт через API. С использованием API нами реализовано получение данных с сайта социальной сети Вконтакте. Схема получения этих данных представлена на Рис. 3.

Вконтакте



Рис. 3. Схема сбора данных с сайта социальной сети Вконтакте

Для разработки этого парсера на языке программирования Python был установлен модуль `vk`. Далее на сайте социальной сети было зарегистрировано собственное приложение и получен сервисный ключ для осуществления запросов. Следующим действием создадим переменную `session` и запишем в нее выполнение метода `vk.Session()`, где в качестве аргумента будет указан сервисный ключ. Далее в переменную `api` передадим метод `vk.API()`, где в качестве аргумента укажем переменную `session`. Реализация данных методов позволит нам избежать ошибки авторизации при отправке запроса. После этого для получения нужной информации о физическом лице используем метод `api.users.get` с параметрами `user_ids`, где записан `id` страницы искомого пользователя, и `fields`, где указаны необходимые к получению параметры (например, `about`, `education`, `home_town` и т. д.).

Использование технологии сбора данных через API позволило нам получить данные с сайтов Судебных приставов и социальной сети Вконтакте.

Анализ полученных данных с применением технологий машинного обучения

Технологии, описанные выше, позволяют получить из открытых источников большое количество полезной информации о контрагенте, в том числе о его финансовой отчетности с сайта БФО Федеральной налоговой службы. В целях анализа полученной информации, а также оценки риска сотрудничества с контрагентом нами разработан аналитический модуль системы информационного мониторинга контрагентов, который, используя методы машинного обучения, позволяет оценить риск сотрудничества с контрагентом.

Для разработки модели машинного обучения использовалась выборка из 11974 компаний, выгруженных из системы Спарк с уже присвоенной оценкой высокого и низкого риска сотрудничества. К данной выборке были добавлены показатели финансовой отчетности компаний, полученные путем парсинга: внеоборотные и оборотные активы, валюта баланса, уставной капитал, капитал и резервы, долгосрочные и краткосрочные обязательства, а также выручка и чистая прибыль (убыток).

Далее с помощью методов `_get_numeric_data()`, `df1.fillna()`, `df2.astype()` библиотеки Pandas языка Python была проведена подготовка данных к построению моделей: были построены модели случайного леса (RandomForestClassifier) [4], дерева решений (DecisionTreeClassifier) [2] и логистической регрессии (LogisticRegression) [3]. Оценка моделей приведена в Таблице 1.

Таблица 1

Метрики качества построенных моделей

	F-мера	AUC-ROC
Логистическая регрессия	0,76	0,79
Дерево решений	0,86	0,86
Случайный лес	0,9	0,9

Как видно из Таблицы 1, лучшее качество показала модель случайного леса, далее идут модели дерева решений и логистической регрессии. Также для визуальной оценки моделей сравним гос-кривую модели случайного леса с гос-кривыми моделей дерева решений и логистической регрессии. На Рис. 4 представлено сравнение гос-кривых случайного леса и дерева решений.

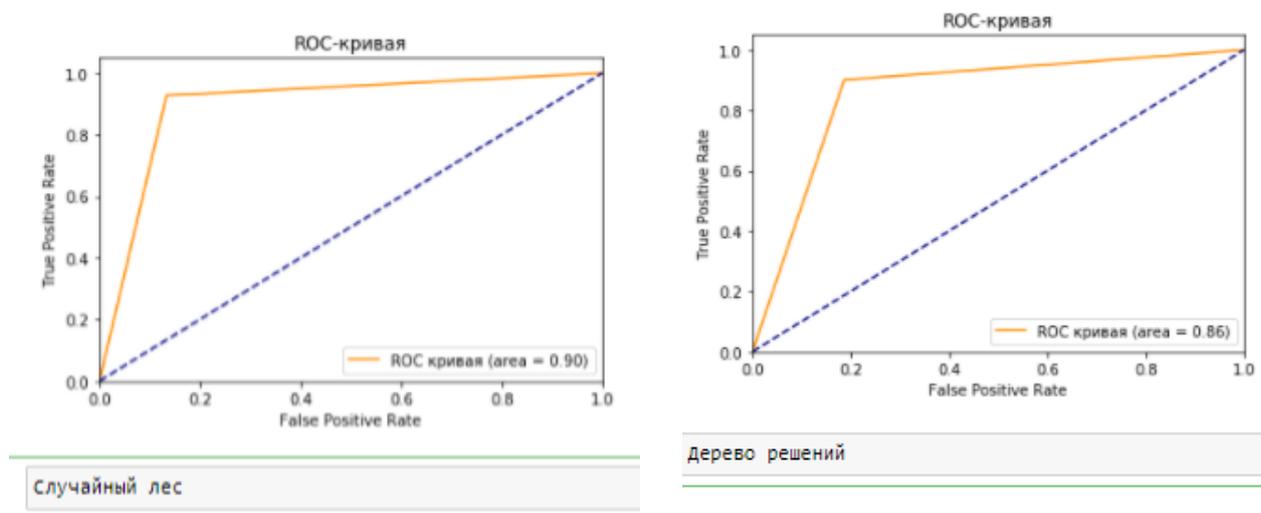


Рис. 4. Сравнение гос-кривых моделей случайного леса и дерева решений

На Рис. 5 представлено сравнение гос-кривых моделей случайного леса и логистической регрессии.

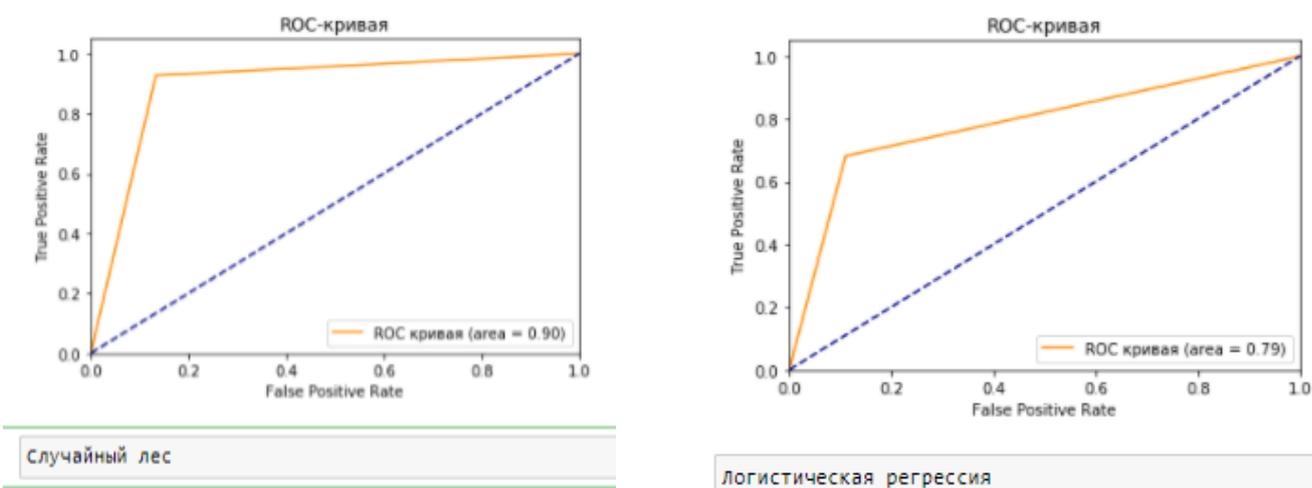


Рис. 5. Сравнение гос-кривых моделей случайного леса и логистической регрессии

На основании проведенных сравнений для реализации аналитического модуля системы информационного мониторинга контрагентов было принято решение использовать модель случайного леса.

ЗАКЛЮЧЕНИЕ

Результатом наших исследований является разработка прототипа системы информационного мониторинга контрагентов, позволяющего оперативно собирать данные о контрагенте из открытых источников, а также анализировать полученную информацию на предмет риска сотрудничества с точностью до 90%.

Дальнейшее исследование технологий сбора и анализа данных из открытых источников может открыть немало новых интересных подходов в области проверки и мониторинга контрагентов.

СПИСОК ЛИТЕРАТУРЫ

1. *Готфрид И.А.* Сравнительный анализ информационно аналитических систем, применяемых налоговыми органами // Электронный научный журнал «Студенческий». 2019. № 24(68), часть 3 С. 8–9. URL: https://sibac.info/archive/journal/student/24%2868_3%29.pdf

2. Документация Decision Trees. URL: <https://scikit-learn.org/stable/modules/tree.html>

3. Документация LogisticRegression. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

4. Документация RandomForestClassifier. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

5. Документация scikit-learn. URL: <https://scikit-learn.org>

6. Документация Selenium. URL: <https://www.selenium.dev/>

7. Интерфейс программирования приложений Вконтакте.

URL: https://vk.com/dev/first_guide

8. *Колтайс А.С.* Информационно-аналитическая система «Спарк» при обеспечении экономической безопасности предприятия // Всероссийская научно-практическая онлайн-конференция «Экономическая безопасность: финансовые, правовые и it-аспекты». 2017. С. 178–182.

9. *Кораблев А.Ю.* Машинное обучение в бизнесе // Азимут научных исследований: экономика и управление. 2018. С. 68–72.

10. Лобин М.А. Машинное обучение в экономике // Вестник УлГТУ. 2019. №3. С. 68–71.

11. Митчелл Р. Скрапинг сайтов с помощью Python. М.: ДМК Пресс, 2016. 280 с.

12. Официальный сайт «Глобас-и». URL: <https://globas.credinform.ru/ru-RU>

13. Официальный сайт «Контур.Фокус». URL: <https://focus.kontur.ru>

14. Официальный сайт «Спарк». URL: <https://www.spark-interfax.ru/> (дата обращения 01.02.2021).

15. Свободная энциклопедия Википедия. URL: <https://ru.wikipedia.org/>

16. Стригалева М.А. Проверка и мониторинг контрагентов: практические аспекты обеспечения финансовой безопасности экономических субъектов // Материалы международной научно-практической конференции «Проблемы обеспечения финансовой безопасности и эффективности экономических систем в XXI в.». 2017. URL: http://www.spbume.ru/file/pages/1197/1-imp_201718_probl_fin_b.pdf#page=349

17. Суханов А.А. Анализ сбора социальных данных из сети Интернет// International Scientific Review. 2017. № 1. URL: <https://scientific-conference.com/images/PDF/2017/32/analiz-sposobov-sbora-sotsialnykh.pdf>

SYSTEM OF INFORMATION MONITORING OF CONTRACTORS

D. L. Kuzmin¹ [0000-0003 – 0011- 6292], Karen Grigoryan²[0000-0001-6470-1832]

Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University, ul. Kremlyovskaya, 35, Kazan, 420008

¹to_kdima@mail.ru, ²karigri@yandex.ru

Abstract

In the context of ever-increasing informatization, automation and digitalization of business, new schemes of unfair actions by both legal entities and individuals are emerging. In this regard, there is an acute problem of quick, effective and high-quality

identification of information about a potential or current counterparty, which will allow you to quickly make the right management decisions.

The article describes one of the ways to solve this problem – the development of a system of information monitoring of counterparties, which will allow you to quickly identify and analyze information about their activities.

Keywords: *development of a system for information monitoring of counterparties, technologies for collecting data from open sources, data analysis using machine learning models.*

REFERENCES

1. *Gottfried I. A.* Comparative analysis of information and analytical systems used by tax authorities // Electronic scientific journal "Studentskiy". 2019. No. 24 (68), part 3. P. 8–9. URL: https://sibac.info/archive/journal/student/24%2868_3%29.pdf
2. Decision Trees documentation. URL: <https://scikit-learn.org/stable/modules/tree.html>
3. LogisticRegression documentation. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
4. Randomforestclassifier documentation. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
5. scikit-learn documentation. URL: <https://scikit-learn.org>
6. Selenium documentation. URL: <https://www.selenium.dev/>
7. Vkontakte application programming interface.
URL: https://vk.com/dev/first_guide
8. *Koltais A.S.* Information and analytical system "Spark" in ensuring the economic security of the enterprise // Materials for the first All-Russian scientific and practical online conference "Economic security: financial, legal and IT aspects". 2017. P. 178–182
9. *Korablev A.Yu.* Machine learning in business // Azimut of scientific research: economics and management. 2018. P. 68–72.
10. *Lobin M.A.* Machine learning in economics // Vestnik UISTU. 2019. No. 3. P. 68–71.
11. *Mitchell R.* Scraping sites using Python. Moscow: DMK Press, 2016. 280 s.
12. Official website of Globas-i. URL: <https://globas.credinform.ru/ru-RU>

13. The official website of "Contour.Focus". URL: <https://focus.kontur.ru>
14. The official website of "Spark". URL: <https://www.spark-interfax.ru/> (accessed 01.02.2021)
15. Free encyclopedia Wikipedia. URL: <https://ru.wikipedia.org/>
16. *Strigaleva M.A.* Verification and monitoring of counterparties: practical aspects of ensuring financial security of economic entities // Materials of the International scientific and practical Conference "Problems of ensuring financial security and efficiency of economic systems in the XXI century". 2017. URL: http://www.spbume.ru/file/pages/1197/1-imp_201718_probl_fin_b.pdf#page=349
17. *Sukhanov A.A.* Analysis of the collection of social data from the Internet // International Scientific Review. 2017. No. 1.
URL: <https://scientific-conference.com/images/PDF/2017/32/analiz-sposobov-sborasotsialnykh.pdf>

СВЕДЕНИЯ ОБ АВТОРАХ



Дмитрий Леонидович КУЗЬМИН – магистрант Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, специальность «Программная инженерия».

Dmitry Leonidovich KUZMIN – Master Student in Computer Science at Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University.

Email: to_kdima@mail.ru

ORCID: 0000-0003 – 0011- 6292



ГРИГОРЯН Карен Альбертович – кандидат экономических наук, доцент, Казанский (Приволжский) федеральный университет, г. Казань.

Karen Albertovich GRIGORIAN – Candidate of Economics, Associate Professor, Kazan (Volga region) Federal University, Kazan.

Email: karigri@yandex.ru

ORCID: 0000-0001-6470-1832

Материал поступил в редакцию 17 июня 2021 года