

РАЗРАБОТКА МЕТОДИКИ И НАСТРОЙКА ЛИНГВИСТИЧЕСКИХ РЕСУРСОВ ДЛЯ ОБРАБОТКИ ИНФОРМАЦИИ ИЗ НЕСТРУКТУРИРОВАННЫХ ДАННЫХ ДЛЯ МОНИТОРИНГА ДЕЯТЕЛЬНОСТИ КРЕДИТНЫХ ОРГАНИЗАЦИЙ

DEVELOPMENT OF METHODOLOGY AND CONFIGURATION OF LINGUISTIC RESOURCES FOR PROCESSING INFORMATION FROM UNSTRUCTURED DATA FOR MONITORING THE ACTIVITIES OF CREDIT INSTITUTIONS

**A. Vorobeva
E. Odinkova
A. Romanov**

Summary. Analysis and comparison of reports of suspicious transactions from credit institutions allow Federal Financial Monitoring Service to determine how effectively countering money laundering/terrorist financing measures are applied in banks, as well as the participation of banks themselves in so-called «schemes».

Additional sources of information about the activities of credit institutions can be unstructured data from the Internet, presented in the form of a text that may contain a variety of information useful for countering money laundering/terrorist financing. The sources of such data may be the official websites of organizations that host a variety of information about their activities, as well as the websites of their competitors, specialized business, and client forums.

Keywords: credit institution, information resource, information technology, information, computational linguistics, unstructured data.

Воробьева Алла Викторовна

кандидат технических наук, доцент,
Московский государственный университет технологий
и управления имени К.Г. Разумовского
vorobevaav@mgtm.ru

Одиноква Елена Владимировна

Кандидат педагогических наук, доцент,
Московский государственный университет технологий
и управления имени К.Г. Разумовского
eodinokova@mgtm.ru

Романов Александр Алексеевич

Московский государственный университет технологий
и управления имени К.Г. Разумовского
aromanov@mgtm.ru

Аннотация. Анализ и сравнение сообщений о сомнительных операциях от кредитных организаций позволяют Росфинмониторингу определить, насколько эффективно применяются меры ПОД/ФТ в банках, а также участие самих банков в так называемых «схемах».

Дополнительными источниками информации о деятельности кредитных организаций могут служить неструктурированные данные из сети «Интернет», представленные в форме текста, который может содержать разнообразную информацию, полезную для целей противодействия отмыванию доходов/финансирования терроризма. Источниками таких данных могут являться официальные сайты организаций, где размещается разнообразная информация об их деятельности, а также сайты их конкурентов, специализированные бизнес — и клиентские форумы.

Ключевые слова: кредитная организация, информационный ресурс, информационные технологии, информация, компьютерная лингвистика, неструктурированные данные.

Контроль над профессиональной банковской деятельностью возложен на Центральный Банк Российской Федерации в соответствии с Федеральным законом № 86-ФЗ «О Центральном банке Российской Федерации (Банке России)», в то время как проблематика ПОД/ФТ в России регулируется Федеральным законом № 115-ФЗ «О противодействии легализации (отмыванию) доходов, полученных преступным путем, и финансированию терроризма». Для реализации данного закона в РФ создана и функционирует Федеральная служба по финансовому мониторингу (Росфинмониторинг). [3]

Основная задача Федерального закона № 115-ФЗ — создать систему противодействия отмыванию доходов, полученных преступным путем, и финансирования терроризма. [5]

Действие Федерального закона № 115-ФЗ распространяется на кредитные организации, которые осуществляют операции с денежными средствами или иным имуществом.

В частности, обязательному контролю подлежат операции с денежными средствами при участии физических или юридических лиц, включенных в перечень лиц, в отношении которых имеются сведения об их причастности к экстремистской деятельности или терроризму. Росфинмониторинг формирует перечень таких организаций и физических лиц. [7]

Федеральный закон № 115-ФЗ предписывает кредитным организациям, разрабатывать правила внутреннего контроля и осуществлять процедуры по выявлению

сомнительных, необычных сделок, а также операций в пользу лиц и организаций, в отношении которых имеются сведения об их причастности к преступлениям в сфере ОД/ФТ и сообщать о них уполномоченному органу, которым является Росфинмониторинг. В законе приводятся правила идентификации клиента и подлежащие обязательному контролю операции. [9]

Своевременное выявление операций, имеющих признак сомнительных, направленных на отмыwanie преступных доходов и финансирование терроризма, это необходимая процедура, помогающая сохранить репутацию банка, его стабильную и непрерывную работу. [11] Под сомнительными операциями понимают операции, имеющие запутанный и необычный характер, не несущие экономического смысла и исполнения правовых законных целей.

Важнейшими источниками информации для Росфинмониторинга при выполнении возложенных на него обязанностей по ПОД/ФТ являются сообщения об операциях (сделках) с денежными средствами и иным имуществом, а также данные из Единого государственного реестра юридических лиц (ЕГРЮЛ) и Книги государственной регистрации кредитных организаций, ведомственного классификатора Банка России (КГРКО).

Основную часть знаний аналитики получают в результате сравнения, анализа и синтеза информации из разрозненных фактов, размещенных в текстах. [4]

Стоит отметить, что неструктурированные данные могут быть представлены в виде текстовых документов, аудио— или видеофайлов. В данной работе мы рассматриваем обработку неструктурированных данных, представленных в текстовом формате, так как значительная доля информации, доступная в электронном виде, представлена текстами на естественном языке. Заключение в них полезная информация не структурирована, а значит, ее невозможно обработать и проанализировать с помощью наиболее широко распространенных методов, используемых для анализа структурированных данных. Такие методы не позволяют обрабатывать неструктурированные данные с целью выделения из них объектов, связей, атрибутов объектов, фактов и т. д.

Информация, полученная в ходе обработки неструктурированных данных — динамична и зачастую носит в себе субъективный и эмоционально окрашенный характер. Постоянно увеличивающийся объем неструктурированных данных, и их неоднозначность негативно сказываются на скорости и качестве их обработки вручную. Ручная обработка такого объема различной по структуре и формату информации затруднительна и имеет крайне высокую трудоемкость. Большие затраты на техническое обслуживание систем хранения

и обработки данных стимулируют спрос на специализированные программные решения, для задач обработки неструктурированных данных, основанные на информационно — поисковых методах и компьютерной лингвистике. Применение данных методов позволяет обрабатывать, хранить, фильтровать, сортировать, осуществлять поиск по большим массивам информации, выделять различные классы сущностей, упомянутых в тексте и их связь.

В настоящее время в сфере финансового мониторинга стоит задача повышения качества учета и использования информации из открытых источников в ее деятельности. В распоряжении аналитических подразделений имеется инструментарий, который позволяет обрабатывать поступающий поток новостей в зависимости от тематики. В свете задачи повышения эффективности обработки неструктурированных данных в целях мониторинга деятельности кредитных организаций имеет смысл дополнить систему обработки таких данных, структурируя потоки документов не по тематике, а по встречающимся в них объектам. Для этого требуется описать объекты и внести их в НСИ (нормативно — справочная информация) системы, а также необходимо описать методику решения данной задачи.

Шаг 1. Путем интервьюирования сотрудников службы, анализа предметной области, относящихся к ней нормативно-правовых актов и выявляемых системой КФП «Новости» упоминаний кредитных организаций формируется структура досье, содержащего всю необходимую информацию о кредитной организации, ее профессиональной деятельности и связанных с ней юридических и физических лицах.

Шаг 2. В процессе анализа упоминаний кредитных организаций в СМИ создается описание фактов для их выявления из текстовых массивов, характерных для данной предметной области и соответствующих структуре досье.

Шаг 3. Производится подбор кредитных организаций, с разными характеристиками, они составят обучающее множество, на котором будет производиться настройка лингвистического инструментария.

Шаг 4. Создается описание кредитных организаций в формате XML и связанных с ними физических и юридических лиц. Для каждого объекта описываются его атрибуты и синонимы.

Шаг 5. Описания объектов загружаются в модуль КФП «Новости». «Мониторинг (Настройка объектов)» и производится отладка.

Шаг 6. Формируются группы объектов в КФП «Новости». «Мониторинг. Объекты и группы».

Шаг 7. Настройка создания отчетов для выбранных групп объектов. Система с заданной периодичностью автоматических обновляет и рассылает отчеты обо всех упоминаемых объектах. Данный отчет позволяет оперативно отслеживать значимые события, связанные с объектами, оценивать информационный фон вокруг кредитных организаций и связанных с ними физических и юридических лиц.

Таким образом, в методике описывается подход к работе с неструктурированными данными, результат обработки которых является наиболее важным фактором, влияющим на производительность труда. Использование информационно-поисковых систем и методов ком-

пьютерной лингвистики дает возможность в короткие сроки получить срез по событиям, связанным с интересующими объектами из СМИ и повысить качество дальнейшего анализа данной информации.

Внедрение разработанной методики позволит повысить качество и скорость обработки информации, полученной из открытых источников. Совершенствование технологий обработки неструктурированных данных может быть полезно в практической деятельности Росфинмониторинга в части оперативного реагирования на резонансные (значимые, сигнальные и т. д.) события в информационном фоне кредитных организаций и связанных с ними физических и юридических лиц.

ЛИТЕРАТУРА

1. Анистратенко О.В. Подходы к оценке несостоятельности кредитных организаций // Финансовый бизнес. 2023. № 4 (238). С. 188–193
2. Богданов В.В., Гринева Н.В. Разработка методики оценки надзорным органом достаточности величины ожидаемых кредитных убытков, рассчитанных коммерческими банками // Проблемы экономики и юридической практики. 2022. Т. 18. № 3. С. 57–66.
3. Гордиенко М.С., Попков В.А. Разработка системы индикаторов финансовых рисков на основе логистической регрессии и методики Corporatometrics // Финансы и управление. 2022. № 1. С. 9–30.
4. Демильханова Б.А. Операции банка России с ценными бумагами и оценка факторного влияния на их доходность // Вестник Чеченского государственного университета им. А.А. Кадырова. 2022. № 1 (45). С. 72–79.
5. Крутов Р.А. Возможности и ограничения применения нефинансовых показателей как исходных данных модели кредитного скоринга юридических лиц // Финансовые рынки и банки. 2022. № 12. С. 141–144.
6. Магомаева Л.Р., Ойбаев Р.С.С., Разина О.М. Разработка адаптивной модели оценки уровня риска для розничных банков // Свидетельство о регистрации программы для ЭВМ 2022660371, 02.06.2022. Заявка № 2022619675 от 23.05.2022
7. Мешкова Е.И., Данилова Е.Н. Скоринговые модели как интеллектуальная собственность банка // Финансовые рынки и банки. 2022. № 4. С. 57–60
8. Назаров Д.М. Использование карт Кохонена для определения банков с высоким риском банкротства // Экономические исследования и разработки. 2022. № 11-1. С. 37–41
9. Павленко Е.Ю. Моделирование антиципационных методов противодействия киберугрозам для крупномасштабных систем с адаптивной сетевой топологией // НИР: грант № 22-21-20008. Российский научный фонд. 2022
10. Серов Е.Р., Васильев С.А. Применение квантовых технологий в банковском бизнесе // Экономика и управление. 2023. Т. 29. № 3. С. 248–255.
11. Татьяна Ю.С. Использование финансовой информации для оценки экономической безопасности коммерческих банков // Скиф. Вопросы студенческой науки. 2022. № 1 (65). С. 84–88.
12. Торопова И.В., Амеличева Д.В., Чирков М.А., Тиханов Н.Е. Оценка надежности коммерческих банков на основе эконометрических моделей бинарного выбора // Банковское дело. 2023. № 8. С. 27–3

© Воробьева Алла Викторовна (vorobeavaav@mgutm.ru); Одиноква Елена Владимировна (eodinkova@mgutm.ru);

Романов Александр Алексеевич (aromanov@mgutm.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»