С. В. Филябин

ТЕХНОЛОГИЯ АВТОМАТИЗАЦИИ МОНИТОРИНГА И КОНТРОЛЯ ЛЕГАЛЬНОСТИ ФИНАНСОВЫХ ОПЕРАЦИЙ СОВРЕМЕННЫХ КРЕДИТНЫХ ОРГАНИЗАЦИЙ

ВВЕДЕНИЕ

В последние годы существенно возросла необходимость контроля финансовых операций в кредитных организациях, предоставляющих различные сервисы: кредитование, переводы денежных средств, оплата услуг, управление счетами через Интернет. Наличие и многообразие сервисов дает большие преимущества при осуществлении финансовых операций, но минусом является отсутствие продуманных и четких механизмов мониторинга информационных потоков с целью недопущения легализации доходов, полученных преступным путем, финансирования экстремистской деятельности, контроля оттока средств за пределы РФ.

Вопросы обеспечения информационной безопасности для современного банка являются жизненно важными по ряду причин:

- основным операционным материалом банка являются деньги. С точки зрения информационной безопасности это существенно повышает риски;
- автоматизированные банковские системы (АБС) в большинстве случаев гетерогенны, то есть представляют собой большой набор бэк-офисных и фронт-офисных программных систем от различных производителей. Управление этими системами осложняется территориальной распределенностью, наличием филиалов и офисов. Очевидно, что низкая степень интеграции и незаконное манипулирование информационными потоками между подсистемами могут привести к серьезным убыткам;
- банк — это точка пересечения внутренних корпоративных сетей, публичных сетей (Интернет) и коммерческих финансовых сетей (Western Union, Visa, Master Card, SWIFT). Наличие удаленного доступа к информационной системе банка посредством этих сетей может привести к несанкционированным финансовым операциям;
- банк хранит конфиденциальную информацию своих клиентов, которая является коммерческой тайной организаций. Возможна утеч-

ка персональных данных о клиентах, физических лицах, и в результате — проведение мошеннических операций.

Основной упор при защите банковских информационных систем необходимо делать на превентивные меры, анализ рисков, построение процессов обеспечения непрерывности деятельности и управление инцидентами.

Требования к защите автоматизированных информационных систем современных банков:

- комплексность — учет всех направлений деятельности банка и возможных информационных атак по каждой из подсистем, создание корпоративной политики безопасности;
- интегрируемость — создание единых интерфейсов обмена информационными сообщениями между подсистемами, использование современных интеграционных решений и платформ;
- легитимность — соответствие положениям и указаниями центрального банка (ЦБ) РФ и других регулирующих органов, международным стандартам ISO 13569 «Banking and related financial services — Information security guidelines», Payment Card Industry Data Security Standard (PCI DSS) и другим;
- управляемость — возможность построения эффективной системы бизнес-процессов банка на основе АБС, оперативный контроль информационных потоков;
- масштабируемость — возможность вносить изменения в архитектуру АБС с целью оптимизации работы набора подсистем;
- отказоустойчивость — возможность быстрого и полного восстановления работоспособности АБС при сбоях.

1. ЗАДАЧА ПРОЕКТА

Подробнее остановимся на легитимности осуществляемых операций. Центральный банк РФ в 2003 году установил порядок представления кредитными организациями в уполномоченный орган сведений об операциях с денежными средствами или иным имуществом, подлежащих обязательному контролю, а также иных операциях с денежными средствами или иным имуществом, связанных с легализацией (отмыванием) доходов, полученных преступным путем (Положение 207-П). В положении определяются основные критерии подозрительности и принципы анализа финансовых операций, вводится перечень официальных справочников экстремистских организаций и физических лиц, справочник государств, не входящих в группу

стран, противодействующих экстремизму, определяется формат электронной отчетности и правила предоставления отчетности.

В случаях, когда службе безопасности и бизнес-подразделениям банков необходимы дополнительные отчеты по легальности движения средств, вырабатываются внутренние регламенты анализа операций, данных и документов, определяются внутренние критерии их подозрительности.

Основной задачей разрабатываемого проекта является создание настраиваемой системы мониторинга и анализа финансовых операций по критериям подозрительности с учетом законодательства РФ (207-П) и внутренних регламентов банков. Основными требованиями к системе являются возможность интеграции с многоплатформенными АБС, масштабируемость, настраиваемый набор критериев подозрительности, наличие инструментальных средств для доработки и получения отчетов по подозрительным операциям.

2. ИНТЕГРАЦИОННАЯ ПЛАТФОРМА

Информационная структура любой финансовой организации — это хранилище данных, то есть совокупность специализированных баз данных (БД) для различных подсистем бизнеса. Каждая база данных имеет собственную структуру.

Технология хранилищ данных позволяет эффективно решать следующие задачи:

- интеграция бизнес-данных всех филиалов и подразделений;
- автоматизация технологий управления;
- автоматизация работ по выпуску консолидированной отчетности;
- аудит филиалов и дочерних предприятий;
- обеспечение архивными данными для оценки развития бизнеса;
- построение единого информационного пространства распределенной организации;
- предоставляет механизм объединения БД бизнес-подсистем в единое целое, используя интерфейс составных запросов к нескольким БД.

На современном рынке программного обеспечения существует большой выбор интеграционных платформ. Для реализуемого проекта был выбран набор программных продуктов IBM WebSphere. Он позволяет создать платформу для электронного бизнеса, основанную на использовании возможностей Интернет. Программная платформа WebSphere базируется на

широко распространенных стандартах, таких как Java, XML, J2EE. Это позволяет легко интегрировать разнотипные ИТ-среды, оперативно адаптироваться к изменению задач бизнеса, упростить доступ внешних пользователей к ресурсам системы, что приводит к увеличению производительности, уменьшению издержек и к более быстрому выходу на рынок с новыми продуктами и услугами.

Платформа WebSphere формирует фундамент законченной оболочки бизнес-решений. Большое количество различных продуктов обеспечивает решение многих проблем, с которыми сталкиваются предприятия любых размеров. IBM WebSphere предоставляет единую точку доступа к приложениям, контенту, процессам и пользователям в рамках интегрируемых информационных систем.

3. АРХИТЕКТУРА СИСТЕМЫ

С учетом описанных ранее ограничений и недостатков современных банковских информационных систем, требований законодательства и бизнеса по контролю легальности операций, а также с учетом технологических возможностей платформы IBM WebSphere, оптимальной представляется архитектура АБС, показанная на рис.1. В данном случае подсистема анализа финансовых операций представляет собой встраиваемый в информационную структуру организации компонент, использующий механизм составных запросов-сообщений через IBM WebSphere и предоставляющий интерфейс к своим внутренним функциям проверки.

Модуль анализа операций состоит из следующих блоков: блок настройки, блок анализа структурированной информации, блок анализа неструктурированной информации, блок анализа повторяющихся операций, блок отчетности. Схема модуля показана на рис. 2.

Блок настройки позволяет указывать, какие сущности (суммы, даты, наименования, текстовые поля) необходимо контролировать, в каких БД и таблицах, позволяет указывать тип проверки: «структурированная-неструктурированная-повторяющиеся операции», и в зависимости от типа проверки определять критерии подозрительности.

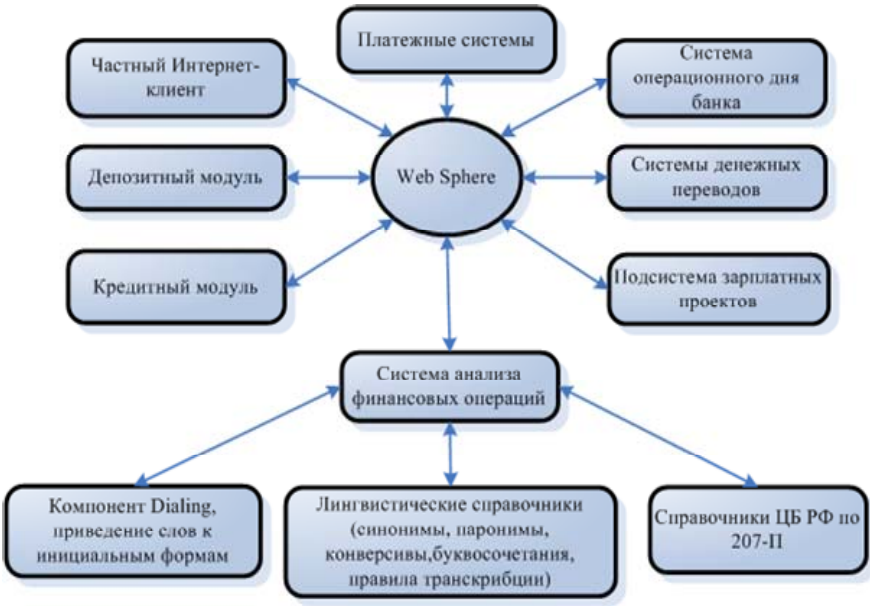

Рис. 1. Архитектура АБС. Интеграция системы анализа финансовых операций

Блок анализа структурированной информации анализирует значения конкретных финансовых реквизитов (суммы, даты, количество и лимиты операций). Например, если указать контролируемый реквизит — сумму документа, критерий подозрительности — превышение определенной суммы, то все документы системы, где сумма превышает заданную, будут считаться подозрительными.

Блок анализа неструктурированной информации использует следующие предметно-ориентированные компоненты и справочники для анализа текстов на естественном языке: компонент приведения слов к инициальным формам Dialing, компонент поиска совпадений, официальный справочник лиц-экстремистов, официальный справочник экстремистских организаций, справочник фраз, считаемых подозрительными (пример фразы: «содействует экстремизму»), справочники синонимов, антонимов, паронимов, конверсивов.

Система Dialing выбрана в качестве компоненты приведения слов к инициальными формам по следующим причинам:

- Dialing является разработкой, над которой в период с 1999 по 2002 год работала группа ведущих российских учёных-лингвистов и программистов;
- система объединяет в себе достоинства других известных систем, таких как ФРАП (система французко-русского автоматического перевода) [5], ПОЛИТЕКСТ (система анализа политических текстов) [6], Микрокосмос [7] и других;
- Dialing включает в себя модули работы с русским и английским языками;
- базовая функциональность системы реализована в рамках технологии COM, что даёт возможность использования её во внешних приложениях.

Инициальной или начальной формой считается: для имен существительных — именительный падеж, единственное число; глаголов — форма инфинитива; для прилагательных — именительный падеж, единственное число, мужской род.

При поиске подозрительных наименований используются следующие критерии:

- инициальная форма наименования присутствует в справочниках экстремистов;
- установлено, что анализируемый объект имел и/или имеет финансовые взаимодействия с экстремистами;
- наименования присутствует в массиве прошлых наименований подозрительного объекта.

При поиске подозрительных текстовых фраз в базах клиентов и документах системы используются следующие критерии:

- фраза подозрительна, если совпали нескольких подряд идущих букв в проверяемой и подозрительной фразах:
- фраза подозрительна, если все слова подозрительной фразы входят в проверяемую фразу:
- фраза подозрительна, если ее подфразой является подозрительная фраза
- фразы подозрительна, если совпали 50% букв проверяемой и подозрительной фразы без учета их последовательности и отдаленности друг от друга;
- фраза подозрительна, если часть проверяемой фразы имеет синоним, подстановка которого делает фразу подозрительной.

Опишем некоторые классы поисковых алгоритмов. Классы можно условно разделить на алгоритмы прямого поиска и алгоритмы, требующие

предварительной обработки документов, создания вспомогательного индексного файла, для ускорения и упрощения поиска.

Алгоритмы прямого поиска — поиск происходит при помощи последовательного просмотра документов.

Достоинства метода: неограниченные возможности по приближенному и нечеткому поиску. Прямой поиск работает непосредственно по оригинальным документам без искажений, любое индексирование всегда связано с упрощением и нормализацией терминов, т. е. с потерей информации. Несмотря на неоптимальность метода, последние несколько десятилетий прямой поиск интенсивно развивается. Выдвинуто множество идей, сокращающих время поиска в несколько раз. Алгоритмы прямого поиска часто разбиваются на отдельные составляющие, которые при группировке дают хорошие результаты. Существующие алгоритмы непрерывно оптимизируются.

Прямой просмотр всех текстов — медленный процесс, но алгоритмы данного класса с успехом применяются в Интернете. Примером может послужить поисковая система Fast Search использующая чип, реализующий логику прямого поиска упрощенных регулярных выражений и кластер из 256 чипов на одной плате. Такая архитектура позволяет системе обслуживать большое количество запросов в единицу времени.

Так же существует множество программ, объединяющих разнотипные классы алгоритмов, например, индексный поиск с дальнейшим прямым поиском внутри блока.

Класс алгоритмов с использованием инвертированных файлов

Инвертированный файл — специальная структура данных. В литературе аналог такой структуры называется «конкордансом» — алфавитно-упорядоченный исчерпывающий список слов из одного текста или принадлежащих одному автору (например, конкордансы произведений Владимира Даля, словарь-конкорданс публицистики Ф. М. Достоевского).

В терминах программирования баз данных инвертированный файл получают при построении индекса таблицы по ключевому полю. После получения упорядоченного по алфавиту списка слов, где для каждого слова перечислены все адреса-позиции, в которых это слово встречается, алгоритм отыскивает нужное слово и возвращает информацию по позициям и встречаемости данного слова в исследуемом тексте.

В инвертированном файле можно хранить информацию не только об адресе, но и другие атрибуты: номер слова, различные гипертекстовые теги

слова. В классической теории информационного поиска (Information Retrieval) в инвертированном файле хранят номер документа и число употреблений этого слова в нем.

Для оптимизации использования дискового пространства используются алгоритмы сжатия информации инвертированных файлов: LZW, алгоритм Хафмана и другие. Необходимо учитывать, что при архивировании файла возрастает нагрузка на процессор при его упаковке-распаковке.

В нашем случае при поиске подозрительных операций и документов можно использовать справочник конкордансов, составленный по досье клиентов. В качестве слов данного справочника будут выступать слова и фразы, считающиеся подозрительными, а адресами будут идентификационные номера клиентов, в досье которых встречается данное слово или фраза. Там же будет храниться и частота встречаемости данной фразы в досье. Справочник будет регулярно дополняться новыми данными в фоновом режиме.

Математические модели

При повышенных требованиях к качеству поиска и большом объеме информации, возможно построение математической модели поиска — на основании модели выводится формула, позволяющая системе принять решение: какой документ считать найденным и как его ранжировать. Модели традиционного информационного поиска можно разделить на три вида: теоретико-множественные (булевская, нечетких множеств, расширенная булевская), алгебраические (векторная, обобщенная векторная, латентно-семантическая, нейросетевая) и вероятностные.

Булевские модели — простейший пример, если слово встречается в документе, то результат функции: true, иначе false.

Ранжирование в векторной модели основано на естественном статистическом наблюдении: чем больше локальная частота термина в документе и больше «редкость» — обратная встречаемость в документах термина, — тем выше вес данного документа по отношению к термину.

Вероятностная модель: вероятность оказаться релевантным для каждого следующего документа рассчитывается на основании соотношения встречаемости терминов в релевантном наборе и в остальной, нерелевантной части коллекции. Вероятностные модели располагают документы в порядке убывания «вероятности оказаться релевантным».

Лингвистические алгоритмы

Лингвистическими считаются методы, в основе своей использующие различные лингвистические правила, справочники и словари (морфологические, синтаксические, семантические), созданные учеными-лингвистами, например, словари синонимов, антонимов и др. Основная масса языков требует должного уровня лингвистической обработки по причине наличия в языке падежей, склонений, множественного и единственного числа, времен глаголов и др. Список задач класса лингвистических алгоритмов:

- автоматическое определение языка документа;
- токенизация (графематический анализ): выделение слов, границ предложений;
- исключение неинформативных слов;
- лемматизация : приведение слов к инициальным формам;
- разделение сложных слов для некоторых языков.

Блок анализа повторяющихся операций согласно критериям, указанным в блоке настроек, например, количество документов с одинаковой корреспонденцией и их общая сумма за определенный период, количество зачислений на один счет, на счета одного клиента и их общая сумма, количество списаний с одного счета, со счетов одного клиента и их общая сумма, модуль анализа помечает подозрительные операции и отправляет их на дополнительный контроль.

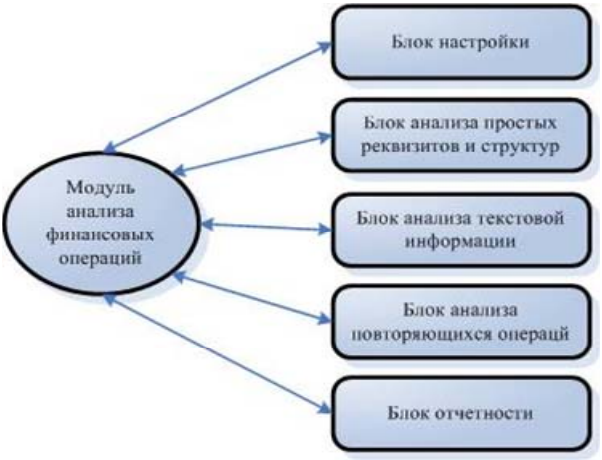

Рис. 2. Общая схема построения модуля анализа операций

Блок отчетности использует данные, собранные в результате проверок всех типов (операции и причины подозрительности), и отображает их в установленном виде для последующего анализа уполномоченным сотрудником. Позволяет выгружать файлы в нужном формате для отправки в ЦБ.

ЗАКЛЮЧЕНИЕ

Хочется отметить, что предлагаемое решение имеет несколько путей развития и доработки: в частности, планируется определение оптимального набора контекстных поисковых алгоритмов различных классов (прямой поиск, индексные, лингвистические), подключение иностранных словарей для анализа текстовой информации на других языках, использование справочников буквосочетаний для разных языков с целью идентификации наименований, применение правил транскрипции, расширение критериев подозрительности и возможностей настройки анализатора.

СПИСОК ЛИТЕРАТУРЫ

1. Построение хранилищ данных // Банковские технологии 7-8/2002. — М.: ПрофиПресс, 2002. — С. 56–60.
2. Лукацкий А. Информационная безопасность банка. — http://www.cisco.com/global/RU/news/media/0209.shtml
3. Анализ уровня информационной безопасности банка. — http://kiev-security.org.ua/box/9/7.shtml
4. Селезнев К. Обработка текстов на естественном языке.— http://www.osp.ru/os/2003/12/048.htm
5. Леонтьева Н.Н. Система французско-русского автоматического перевода (ФРАП): лингвистические решения, состав, реализация // МП и ПЛ. Проблемы создания системы автом. перевода / Сб. научн. тр. МГПИИЯ им. М. Тореза. — М., 1987. — Вып. 271. — С. 6–25.
6. Леонтьева Н.Н. ПОЛИТекст: информационный анализ политических текстов // Сб. НТИ. — 1995. — Сер. 2, № 4.
7. Raskin V., Nirenburg S. Lexical Semantics of Adjectives // Recent Papers from the Mikrokosmos and Corelli Projects. — New Mexico State Univ., 1996. — Vol. 2.