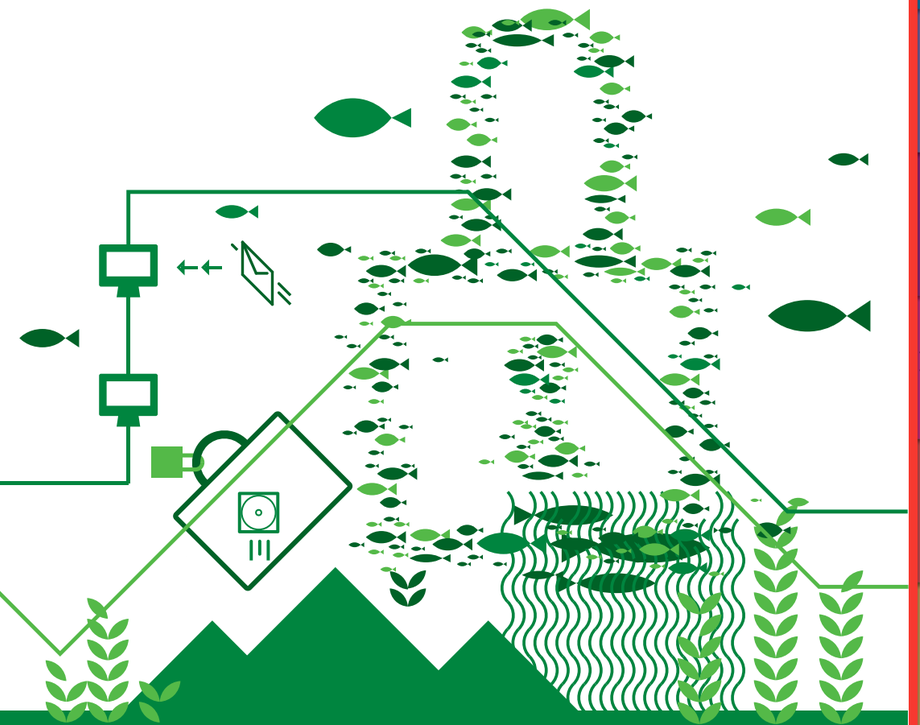
INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS

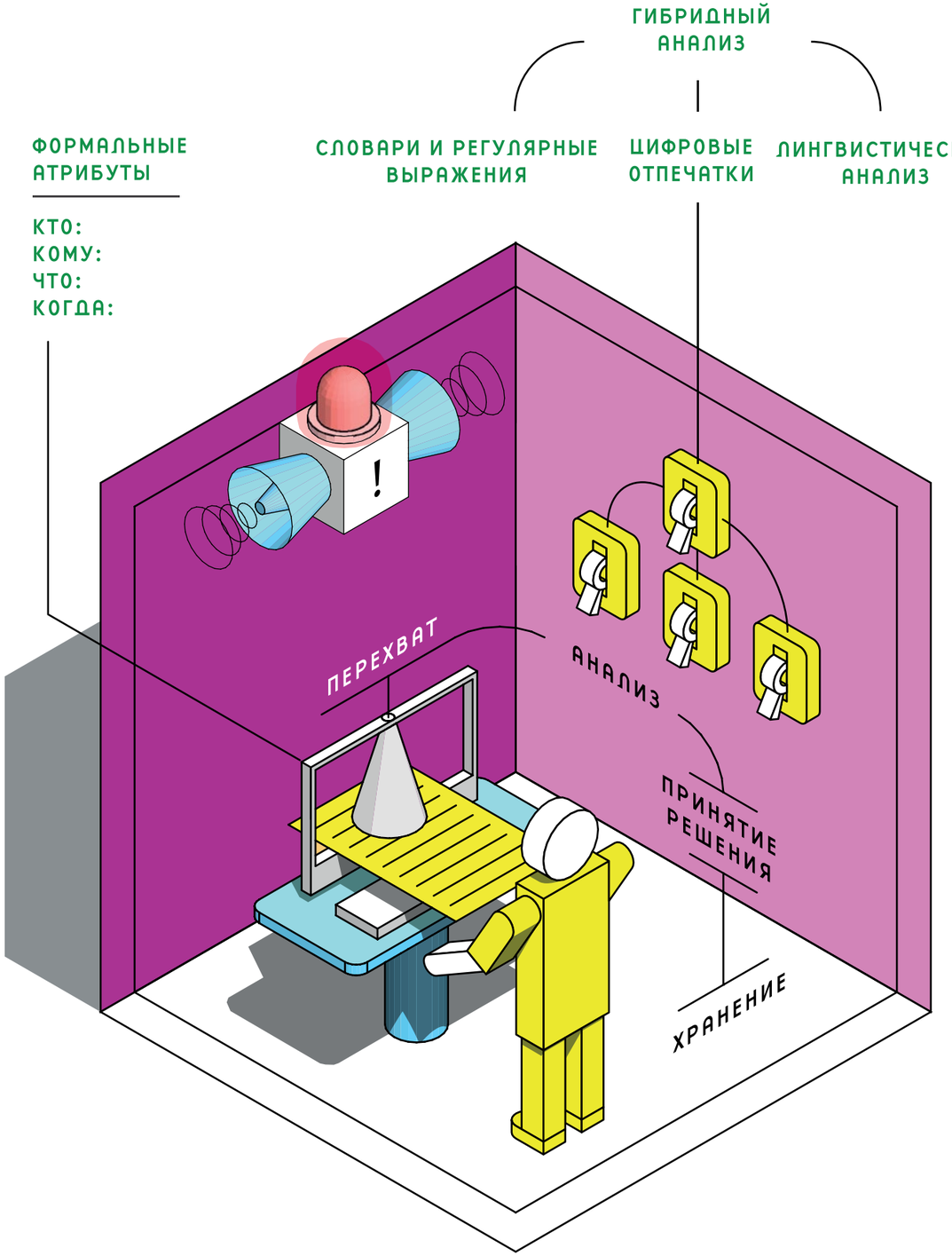
Технологии InfoWatch для анализа и защиты информационных активов компании

Одним из самых ценных активов любой компании в наши дни является информация, и от того, насколько успешно она защищается, часто зависит успех всего бизнеса. В то же время обеспечение высокого уровня информационной безопасности не должно быть связано с ограничением использования современных средств коммуникации и технологий обмена данными, которые сейчас активно внедряются в компаниях любого масштаба. Защита информации от утечки, в результате которой она может стать доступной третьим лицам — задача очень сложная, и простого решения этой проблемы в виде одной программы или одного устройства не существует. Для эффективной защиты необходимо использовать фактически весь арсенал средств информационной безопасности.

Компания InfoWatch разрабатывает высокотехнологичные комплексные решения, которые эффективно защищают корпоративную информацию от утечки или несанкционированного распространения. Технологии InfoWatch позволяют разобрать все документы заказчика, разделив их по категориям, структурировать информационные активы, выявить из большого объема информации конфиденциальные данные. Концепция InfoWatch заключается в том, чтобы контролировать движение информации на всех этапах: начиная от аудита (что и где лежит), выявления контентных маршрутов движения информации (от кого — кому, какая категория данных передается), заканчивая контролем распространения конфиденциальной информации с помощью DLP-системы и настроенных политик информационной безопасности.

Подробнее о технологиях InfoWatch



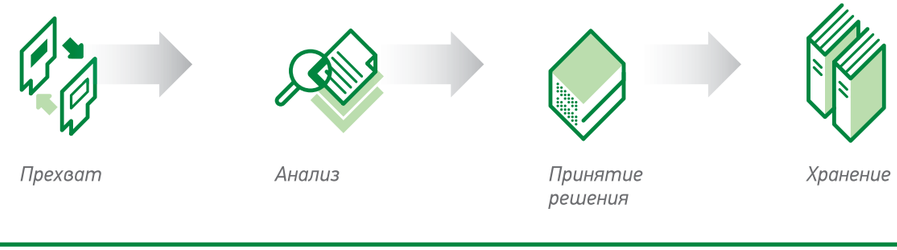


INFOWATCH
BECAUSE YOUR DATA IS YOUR BUSINESS

Основные функции системы защиты данных

InfoWatch Traffic Monitor Enterprise — это современное решение для защиты данных, которое представляет собой набор инструментов и технологий для предотвращения или контроля перемещения конфиденциальной информации за переделы периметра компании. Если в потоке передаваемых данных была выявлена конфиденциальная информация и система классифицировала эту передачу как инцидент, то автоматически срабатывает режим защиты и запускается процедура реагирования на инцидент, например, происходит блокирование передачи информации, отправитель получает предупредительное сообщение или оповещение приходит лицу, ответственному за информационную безопасность. Информация об инциденте вместе с копией перехваченного документа сохраняется в архиве.

РАБОТА СИСТЕМЫ ЗАЩИТЫ ДАННЫХ



Ключевой функцией DLP-системы является автоматическое обнаружение (детектирование) в информационных потоках конфиденциальной информации, которую нужно защищать. Поэтому именно алгоритмы анализа информации являются основой успешной работы DLP-решения и надежной защиты корпоративных данных.

Как из потока данных выявляется конфиденциальная информация?

В решениях InfoWatch используется сразу несколько технологий для выявления конфиденциальных данных из потока передаваемой информации.

Перехваченная информация анализируется сначала по ее внешним признакам — формальным атрибутам, например, для электронного письма определяется «кто» его отправлял, «куда», «когда» и др. Вторым этапом является извлечение содержимого перехваченной информации и его контентный анализ на основе содержащихся в документе или письме слов и выражений. Далее на основании различных методов определяется тематика и степень конфиденциальности перехваченного объекта.

INFOWATCH
BECAUSE YOUR DATA IS YOUR BUSINESS

Методы детектирования конфиденциальной информации

В зависимости от категории информации и каналов ее передачи эффективность применения различных технологий, для выявления данных в общем потоке, разная. В настоящее время существует несколько подходов, и у каждого из них есть свои сильные и слабые стороны.

Так, анализ текстов на основе словарей и регулярных выражений позволяет обнаружить сообщения, содержащие заданные наборы слов или цифр, что наилучшим образом подходит для детектирования объектов, образованных по определенному шаблону. Однако в случае анализа других типов текстовых объектов этот метод бессилен.

Широко используемая технология «цифровых отпечатков» легко справляется с защитой выбранных и предварительно проиндексированных документов, однако она бесполезна в случае анализа неформальной переписки и совершенно беспомощна, если процедура индексации не проводится регулярно.

Использование лингвистического анализа позволяет обеспечить как защиту документов на любом этапе жизненного цикла, так и анализ любой переписки (почтовые сообщения, сообщения в блогах и форумах, ICQ и т.д.) или пересылки документов и его частей, однако требует предварительной настройки и малоэффективен в тех случаях, когда необходимо выполнять выборочную защиту документов.

Эксперты компании InfoWatch разрабатывают и используют в своих решениях гибридный анализ, который является комплексным и объединяет в себе следующие технологии контентного анализа: лингвистический анализ, цифровые отпечатки, анализатор шаблонов. Использование гибридного анализа максимально повышает надежность и точность выявления конфиденциальной информации, что способствует более эффективной ее защите.

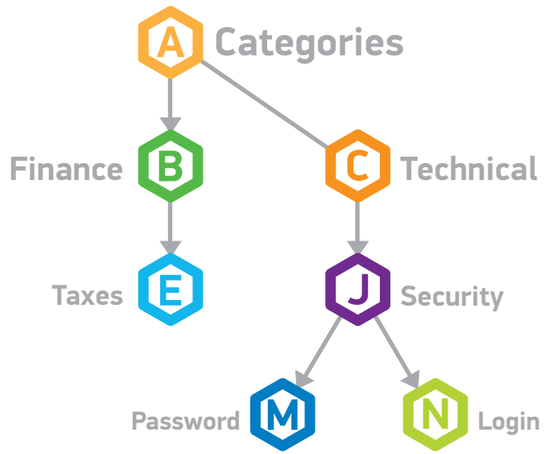
Лингвистический анализ в решениях InfoWatch

Использование метода лингвистического анализа является одним из преимуществ решений InfoWatch по сравнению с конкурентами, так как только этот метод обеспечивает высокий уровень детектирования критической информации на любом этапе жизненного цикла, в том числе сразу после создания, позволяя определить не только формат документа, но и понять его смысл. Этим достигается качественный результат даже при анализе небольших фрагментов текста, которые могут быть вставлены в любой документ или отправлены в неформальной переписке или через систему мгновенного обмена сообщениями (ICQ, jabber и т.д.).

В данном методе автоматическое определение тематики текста выполняется на основе заранее созданной базы контентной фильтрации (БКФ). БКФ не только описывает категории информации, циркулирующей в компании, но и учитывает различные атрибуты её конфиденциальности, в т.ч. специфику деятельности компании, ее требования к безопасности. По результатам проведения лингвистического анализа тексту автоматически присваиваются те или иные категории, соответствующие тематике и содержанию. В анализируемой информации могут встретиться термины (слова и словосочетания) из разных категорий, поэтому она может быть отнесена к одной или нескольким категориям БКФ.

Поэтому важно создать базу, которая обеспечит надежные результаты фильтрации информации по категориям. Основным методом лингвистического анализа с помощью БКФ является поиск в анализируемом фрагменте информации слов и словосочетаний, описывающих конфиденциальные данные и структурированных по категориям.

Создание БКФ



База контентной фильтрации — это база данных, представляющая собой выделенный на основе вероятностных и математических методов иерархически организованный список (дерево) категорий с произвольным количеством вложенных уровней, и содержащая слова и выражения, наличие которых в документе позволяет определить тематику и степень конфиденциальности информации.

Для создания БКФ сначала нужно составить ее структуру — рубрикатор или дерево контентных категорий. Такое дерево представляет собой иерархический список с категориями и подкатегориями.
Затем каждая категория наполняется списком терминов, ключевых слов, словосочетаний и фраз, появление которых в анализируемом фрагменте информации указывает на его принадлежность к определенной контентной категории. После этого для каждого термина/словосочетания устанавливается вес, который этот термин будет иметь при отнесении информации к определенной категории. Решение о том, является ли текст релевантным контентной категории, принимается по результатам сравнения общей суммы веса терминов, найденных в тексте, с порогом релевантности этой категории.

INFOWATCH
BECAUSE YOUR DATA IS YOUR BUSINESS

Для обеспечения качественной категоризации БКФ необходимо поддерживать в актуальном состоянии — редактировать изменяющиеся со временем категории, добавлять и/или удалять термины и словосочетания, изменять их вес и др.

Характеристические и частотные термины
Термины, входящие в БКФ разделяются на частотные и характеристические.

Характеристический термин — термин, который единожды встретившись в анализируемом фрагменте информации, 100% свидетельствует о принадлежности его к определенной категории.

Частотный термин — термин, который, при наличии его в анализируемом фрагменте информации, с определенной долей вероятности свидетельствует о принадлежности этого фрагмента к определенной категории.

Стандартная БКФ

В состав InfoWatch Traffic Monitor Enterprise входит стандартная база контентной фильтрации, содержащая наиболее общие категории и термины, встречающиеся практически в любых областях. Такая БКФ гарантирует детектирование данных по таким тематикам, как «Счета», «Бухгалтерия», «Кредиты», «Тендеры», «HR» и т.д. Для обеспечения эффективного лингвистического анализа стандартная БКФ нуждается в дальнейшей доработке под компанию, в ходе которой должны быть учтены как отраслевые, так и специфические особенности организации.

Отраслевая БКФ

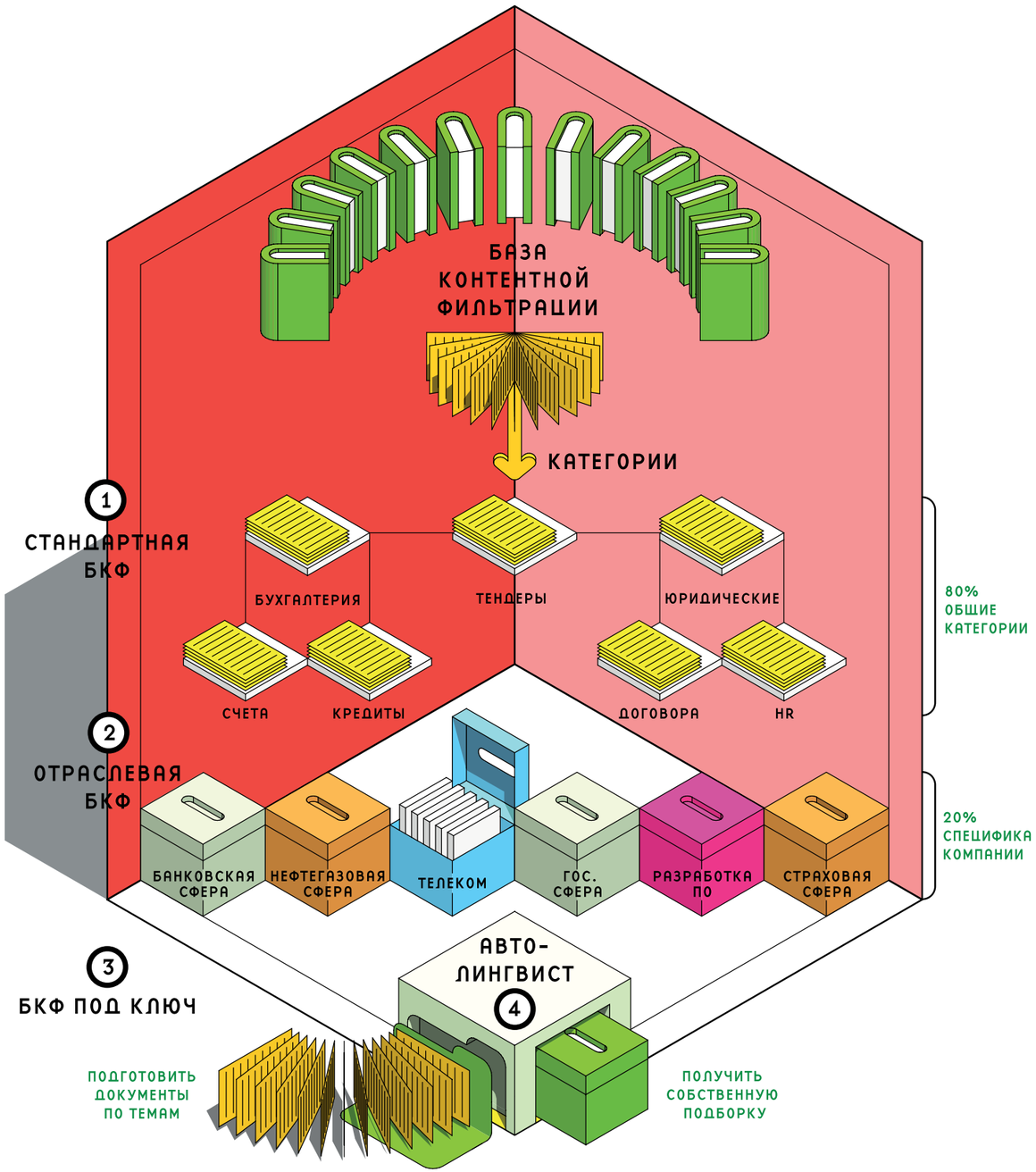
На основе своего многолетнего сотрудничества с компаниями, работающими на различных вертикальных рынках, InfoWatch разработала несколько баз контентной фильтрации, оптимизированных под потребности конкретных сегментов рынка. Сегодня InfoWatch предлагает следующие вертикально-адаптированные БКФ:

- банковская (финансовая)
- страховая
- нефтегазовая
- телеком
- разработка ПО
- государственная (выявление нарушений законодательства РФ)

Оптимизация БКФ под определенный сегмент рынка означает, что такая БКФ содержит наиболее распространенные категории, характерные для данной отрасли. Запуск InfoWatch Traffic Monitor с предустановленной БКФ, оптимизированной под определенный вертикальный рынок, позволяет компании незамедлительно начать пользоваться продуктом и обеспечить высокую точность детектирования конфиденциальной информации.

БКФ «под ключ»

БКФ «под ключ» — это услуга компании InfoWatch по доработке отраслевой БКФ с учетом специфики деятельности конкретной компании. В базе контентной фильтрации, адаптированной под потребности определенной рыночной вертикали, примерно 80% категорий являются общими для всех компаний этого сектора. Оставшиеся 20% составляют категории, характерные для конкретной компании. Дополнение такой отраслевой БКФ категориями, отражающими специфику деятельности данной компании, обеспечивает лучшую категоризацию и более точное детектирование конфиденциальной информации в информационных потоках компании более 85%. Дополнение БКФ специфическими категориями может происходить вручную, либо с использованием специального программного продукта — InfoWatch Автолингвист.



INFOWATCH
BECAUSE YOUR DATA IS YOUR BUSINESS

InfoWatch Автолингвист для создания собственной БКФ

InfoWatch Автолингвист — компонент решения InfoWatch Traffic Monitor, который позволяет автоматизировать процесс создания собственной БКФ или доработки отраслевой БКФ, а также поддерживать ее в актуальном состоянии.

Для создания БКФ с помощью InfoWatch Автолингвист необходимо подготовить репрезентативную коллекцию документов компании и рассортировать ее по отдельным папкам в зависимости от тематики, например, финансовые документы, договора о неразглашении и др. После обработки коллекции документов с помощью InfoWatch Автолингвист эти папки составят структуру рубрикатора. InfoWatch Автолингвист анализирует загруженную в него коллекцию документов и автоматически выделяет термины, на основании которых будет происходить отнесение анализируемой информации к той или иной категории.

Заключительным этапом при создании базы контентной фильтрации является добавление в нее характеристических терминов, которые не могут быть выделены автоматически, но одновременно однозначно свидетельствуют о конфиденциальности документа, такими как, например, название секретного проекта.

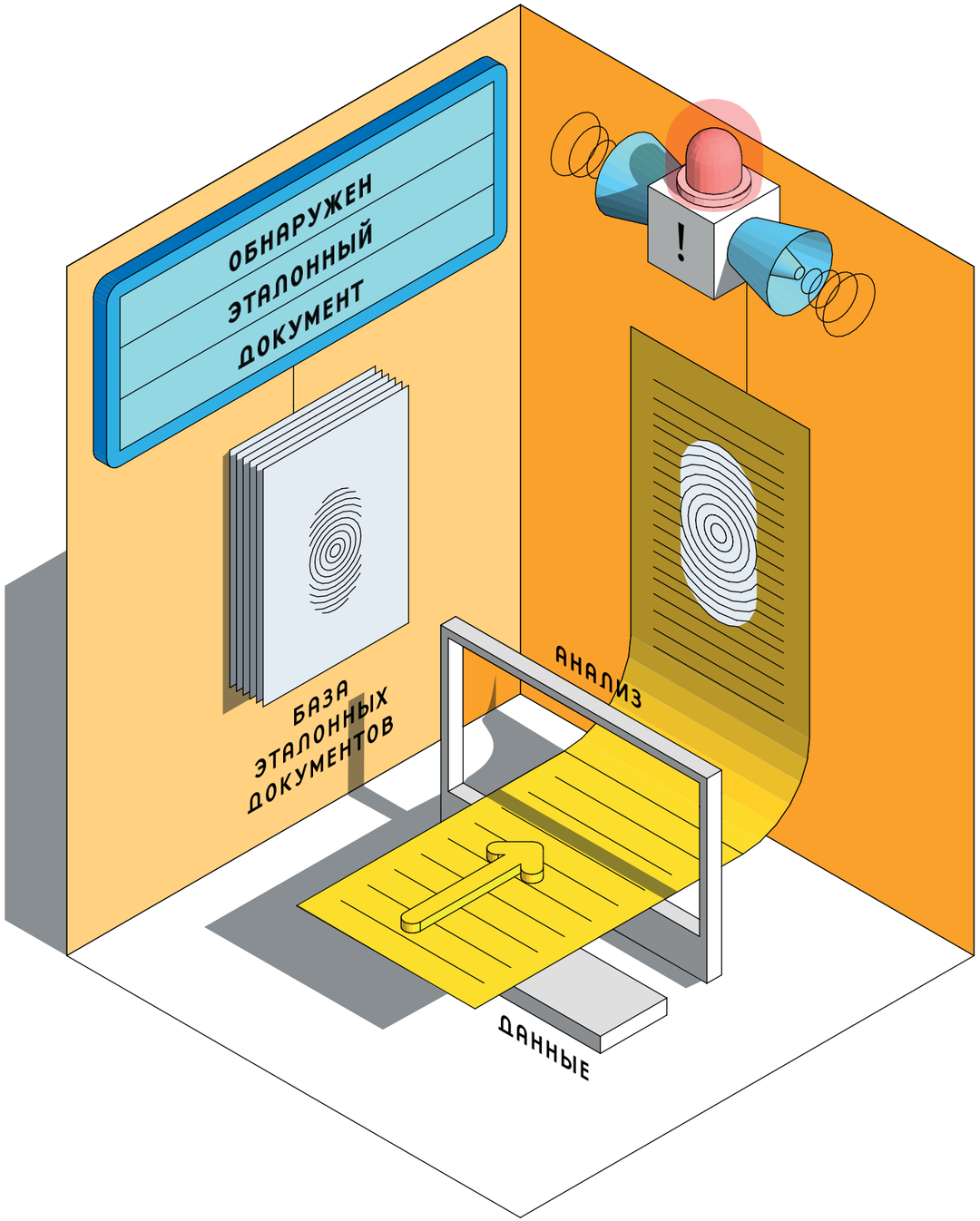
ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ И БАЗА КОНТЕНТНОЙ ФИЛЬТРАЦИИ

В отличие от решений, базирующихся только на технологиях цифровых отпечатков (фингерпринты/fingerprints, шинглы/shingles и т.д.), данная технология позволяет защитить не только редко изменяемые данные, такие как устав компании или реестр ее акционеров, но и постоянно меняющиеся или новые, только что созданные данные, к которым можно отнести различные договоры с партнерами и подрядчиками, описание новой технологии, планы вывода новых продуктов на рынок, анкеты клиентов с указанием условий сотрудничества и персональных данных и многие другие документы.

Преимущества технологии

- проактивная защита, в т.ч. для данных, создаваемых «с нуля» каждый день;
- автоматическая классификация анализируемого текста;
- поддержка всех европейских языков, лингвистическая поддержка для русского, английского, французского, немецкого, испанского, итальянского, украинского, арабского, польского, румынского, латышского языков; автоматическое детектирование языков;
- возможность работы с документами, написанными сразу на нескольких языках;
- поддержка словоизменения (словарная и нечеткая морфология);
- предустановленная база контентной фильтрации общего характера, возможность использования отраслевых и добавления собственных баз.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS



Цифровые отпечатки

Технология *«Цифровые отпечатки»* предназначена для защиты больших по объему документов, содержание которых не изменяется или меняется незначительно. Детектор цифровых отпечатков, используемый в InfoWatch Traffic Monitor, позволяет автоматически обнаруживать в анализируемом тексте цитаты из документов-образцов, содержащих конфиденциальную информацию.

На этапе настройки системы собирается база конфиденциальных документов, для которых затем создаются цифровые отпечатки. Из цифровых отпечатков документов-образцов формируется база эталонных документов. В технологии цифровых отпечатков, все параметры формирования эталонной базы документов-образцов подобраны с таким учетом, чтобы обеспечить оптимальное качество детектирования при минимально возможном объеме эталонной базы. При этом по информации, хранимой в эталонной базе, невозможно восстановить исходный текст документов-образцов. Таким образом, все помещенные в систему конфиденциальные документы будут надежно защищены: даже если злоумышленник получит доступ к эталонной базе цифровых отпечатков InfoWatch Traffic Monitor, утечки данных не произойдет.

Основным отличием технологии «Цифровые отпечатки», применяемой в продукте InfoWatch Traffic Monitor, от отпечатков, используемых в продуктах других компаний, является комплексная обработка анализируемого текста, включающая в т.ч. поддержку лингвистики.

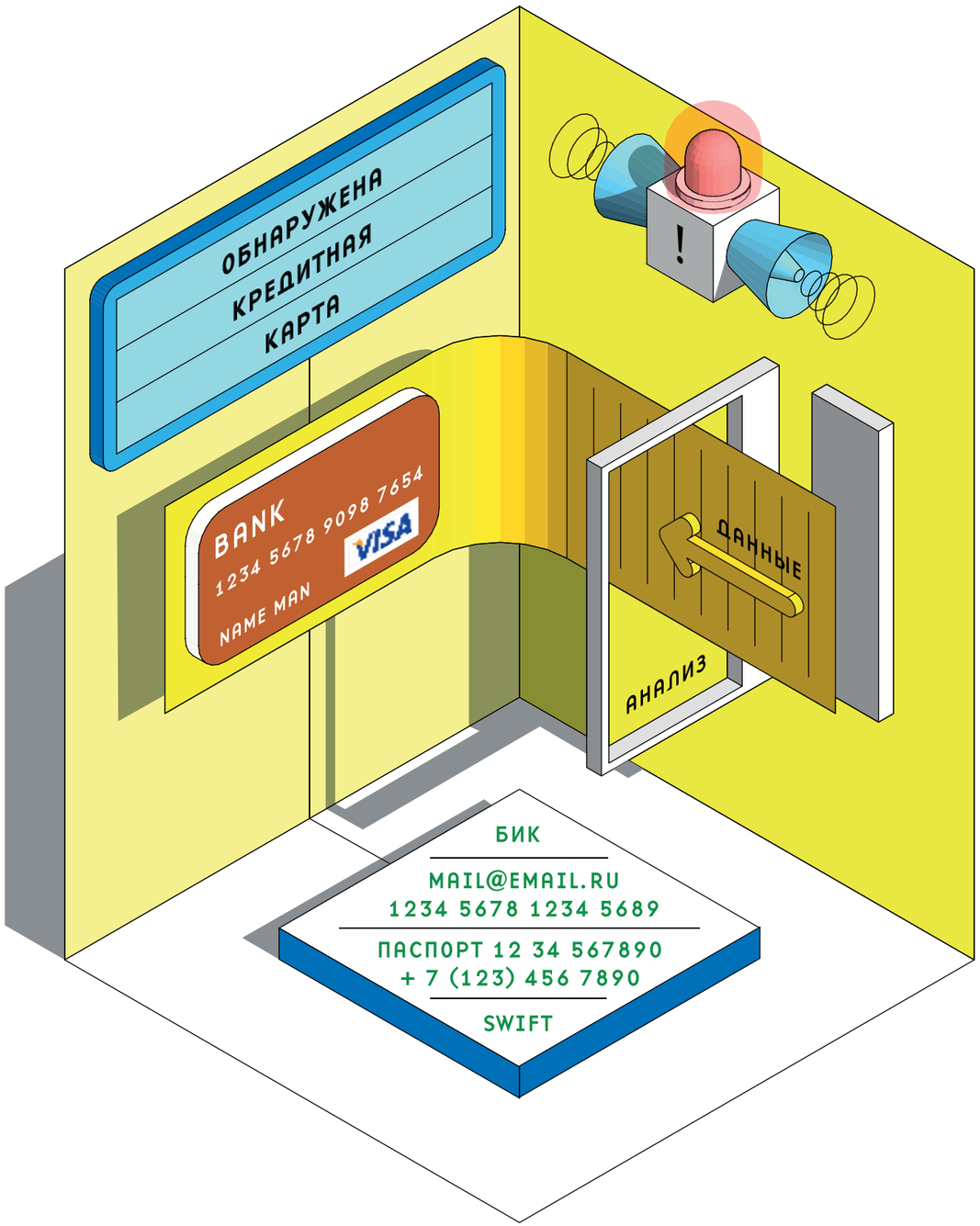
Такой подход существенно повышает качество детектирования конфиденциальной информации и позволяет обеспечить надежность работы метода не только в случае самых простых изменений текста (таких как изменение форматирования, замусоривание текста лишними пробами или пунктуацией и т.д.), но и в более сложных случаях, например:

- различные варианты написания одного и того же слова (например, варианты написания слов в разных регистрах, буквы е/ё в русском языке, варианты написания лигатур в немецком языке и т.д.);
- различные формы одного и того же слова и разные варианты написания сложносоставных слов, технология детектирования опечаток / ошибок и транслита и т.д.

Преимущества технологии

- защита редко изменяемых или неизменяемых (статичных) документов;
- детектирование не только дословных совпадений, но и модифицированных фрагментов текста;
- поддержка лингвистической обработки анализируемого текста (в т.ч. для многоязыковых документов), возможность морфологического анализа;
- автоматическое распознавание схожести документов и обнаружение цитат из документов-образцов.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS



Анализатор шаблонов

Согласно исследованию аналитического центра InfoWatch «Глобальное исследование утечек корпоративной информации и конфиденциальных данных, 2012 год», наиболее уязвимыми по-прежнему остаются персональные данные — на их долю приходится 89,4% всех произошедших за исследуемый период утечек. Это неудивительно: персональные данные довольно просто конвертировать в «живые» деньги, и именно за такими данными охотятся злоумышленники. Прежде всего, это данные о банковских картах, номерах социального страхования, паспортные данные, адреса, телефоны и т.д.

Технология *«Анализатор шаблонов»*, применяемая в InfoWatch Traffic Monitor, предназначена для детектирования алфавитно-цифровых объектов по шаблону данных (маске) и позволяет наиболее эффективно выявлять факты пересылки персональных данных или финансовой информации. Кроме того, данная технология может использоваться как вспомогательный метод для обнаружения фактов несанкционированной пересылки внутренних документов, содержащих формализованные данные, образованные по определенному шаблону (например, договоров или счетов в случае детектирования банковских реквизитов, кодов классификаторов и т.д.).

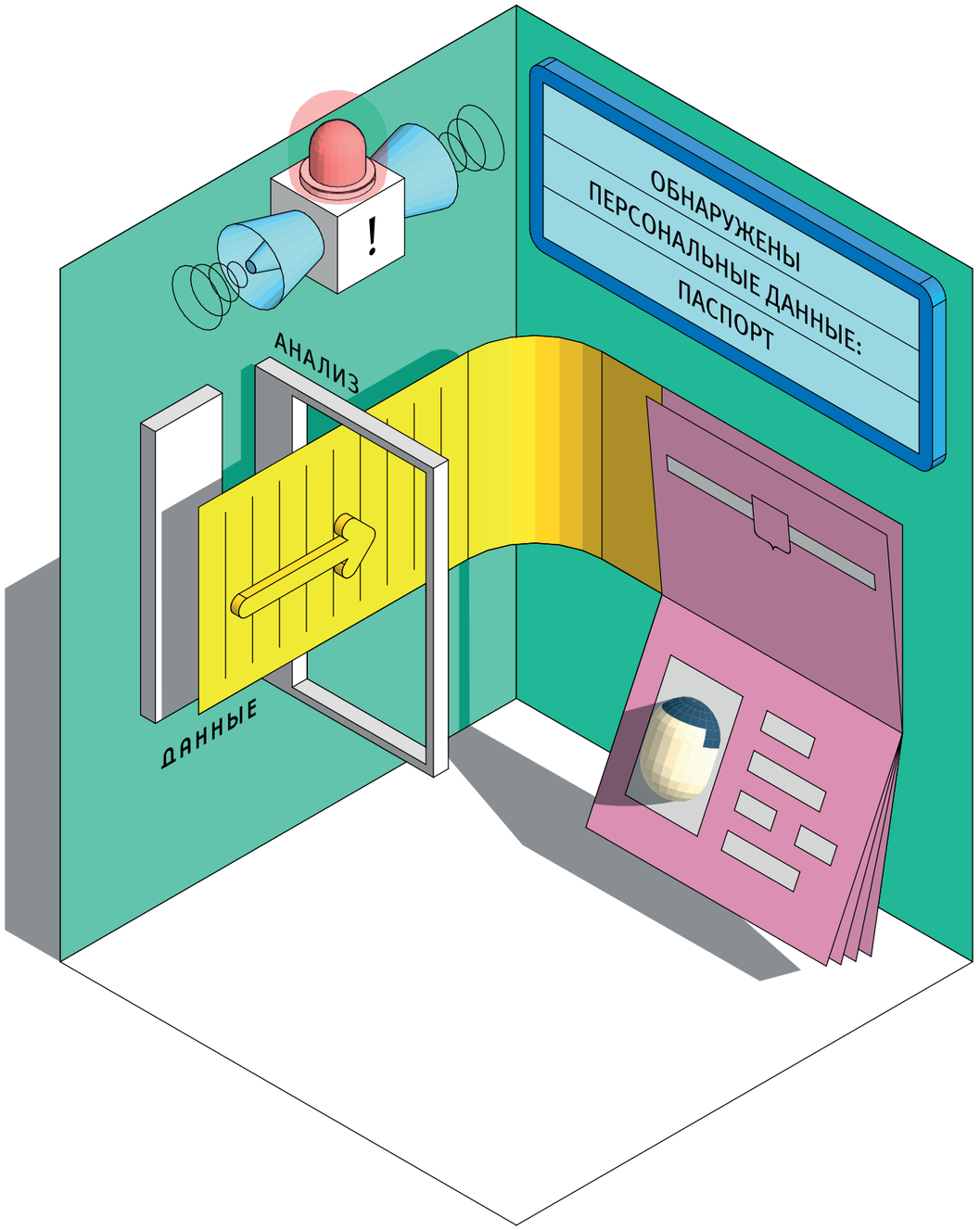
InfoWatch Traffic Monitor поставляется с набором предустановленных шаблонов текстовых объектов, которые можно использовать при создании политик безопасности. В этот список входят:

- номера кредитных карт;
- номера паспортов;
- номера телефонов;
- адреса электронной почты;
- номера карточек пенсионного страхования (РФ);
- индивидуальные номера налогоплательщиков (РФ);
- идентификационные коды банков (РФ и SWIFT);
- международные идентификационные коды ценных бумаг;
- коды общероссийских классификаторов и т.д.

Преимущества технологии

- высокая эффективность в детектировании персональных данных и финансовой информации;
- позволяет детектировать специфичный для каждой организации тип контента;
- практически не требуют актуализации, так как шаблоны основных конфиденциальных данных изменяются крайне редко.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS



Детектор паспортов

Паспорт является основным документом, удостоверяющим личность гражданина. Его главная страница содержит основные сведения о личности гражданина: фамилия, имя, отчество, пол, дата рождения, место рождения. Такая информация, безусловно, подпадает под категорию персональных данных и требует особого обращения и защиты в соответствии с федеральным законом 152-ФЗ «О персональных данных».

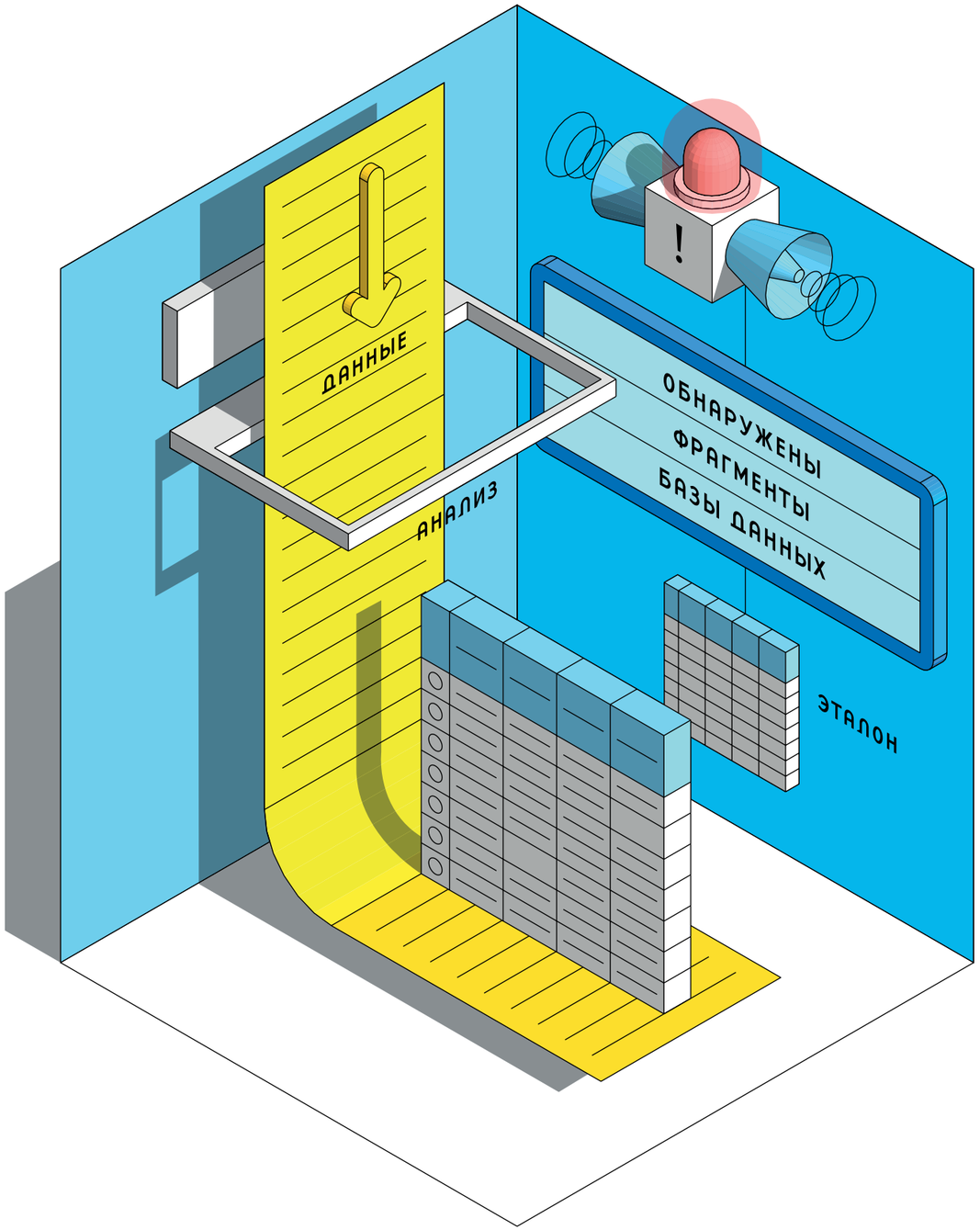
Технология *«Детектор паспортов»* решения InfoWatch Traffic Monitor позволяет определять наличие страниц паспорта гражданина Российской Федерации в потоке перехваченных файлов.

При обнаружении главной страницы паспорта в консоле управления InfoWatch Traffic Monitor в разделе «Отчёты» появится предупреждение об утечке персональных данных. В информации о перехваченных объектах указывается объект «Паспорт РФ».

Преимущества технологии

Основным преимуществом технологии «Детектор паспортов» от InfoWatch является полное отсутствие необходимости в какой-либо предварительной конфигурации. Технология работает, что называется, «из коробки» сразу после запуска системы и не требует никакой дополнительной настройки.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS



Детектор выгрузок из баз данных

Сейчас сложно представить компанию, в которой бы не вёлся учёт клиентов с использованием баз данных. При этом обычно работа с такой системой для сотрудника подразумевает доступ ко всей массе клиентских данных, что, конечно, становится большим соблазном их копирования с целью дальнейшего коммерческого использования.
В решении InfoWatch Traffic Monitor существуют технологии, защищающие базы данных и предотвращающие утечку данных с помощью технологии «Детектор выгрузок из баз данных».

«Детектор выгрузок из баз данных» позволяет фиксировать наличие эталонных выгрузок из баз данных в сетевом трафике и текстовых документах. Технология делает возможным оперативно реагировать на передачу информации, скопированной из базы данных, определять нарушителя, предотвращать выход данных за пределы организации и использовать цифровые доказательства по инциденту в случае необходимости проведения расследования или судебного преследования.

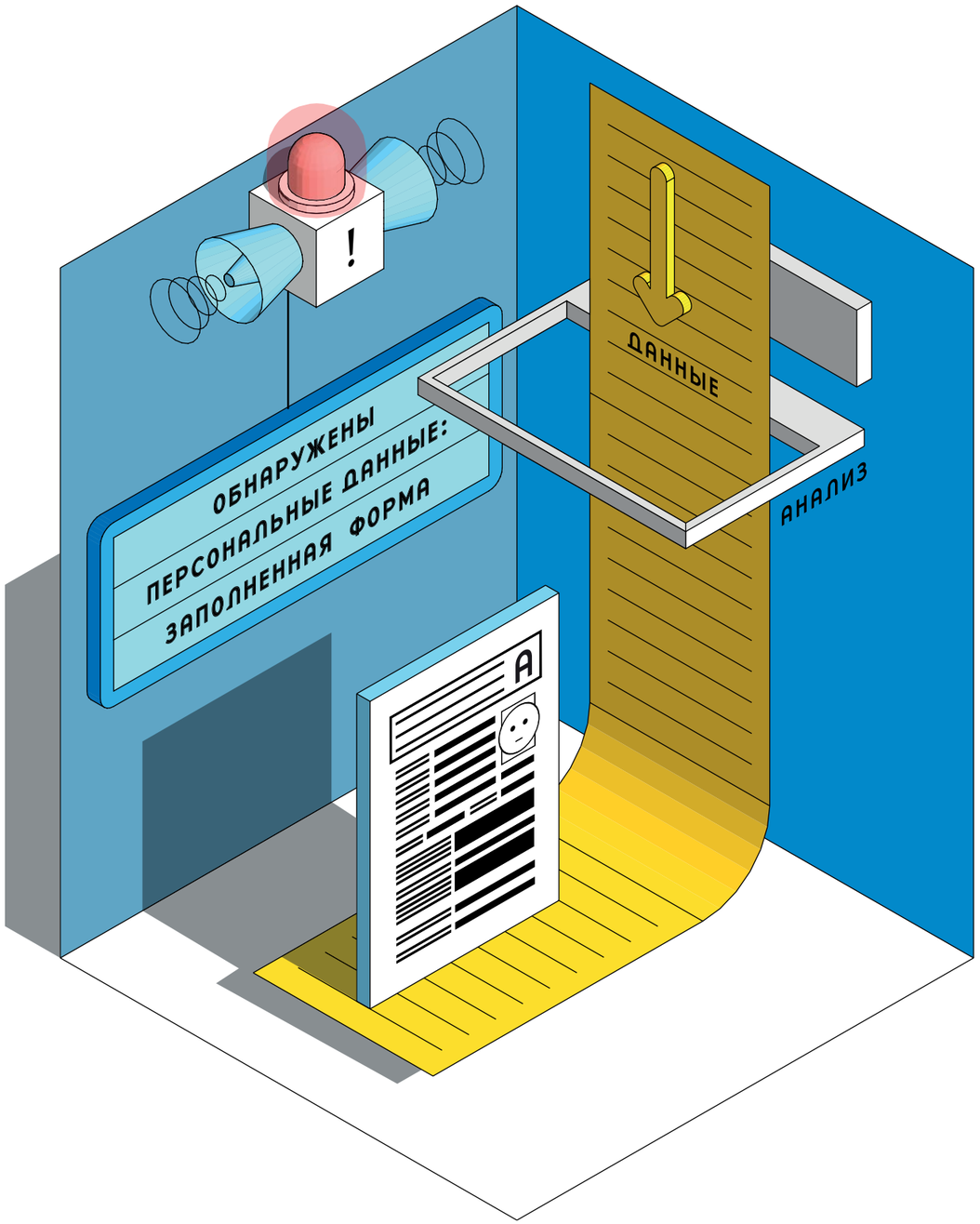
Настройка модуля «Детектор выгрузок из баз данных», как правило, не вызывает больших вопросов. Перед тем как создать эталонную выгрузку, нужно проанализировать структуру базы данных и выбрать те столбцы или их комбинации, в которых хранится конфиденциальная информация. Например, список клиентов сам по себе может не являться тайной, но список клиентов и их прямые контакты уже становится конфиденциальной информацией. Тайну могут составлять как один, так и несколько столбцов, а также сочетание одного столбца и любого из нескольких заданных (например, различное написание названия организации). Далее производится пробная выгрузка выбранных столбцов и сохраняется в текстовом файле, который будет использоваться системой InfoWatch Traffic Monitor в качестве эталонного образца.

При обнаружении признаков эталонной выгрузки в потоке сетевого трафика, в консоли управления InfoWatch Traffic Monitor в разделе «Отчёты» появляется уведомление об утечке конфиденциальной информации. В деталях «Инцидента» о перехваченных объектах содержится имя файла эталона, соответствующего детектированной выгрузке. В секции «Контекст» в поле «Релевантность» можно посмотреть рейтинг максимального из условий, указавших на конфиденциальность информации.

Преимущества технологии

- высокая эффективность в детектировании различных комбинаций и фрагментов персональных данных клиентов, скопированных из корпоративной базы данных;
- гибкая настройка под специфические форматы данных, хранящихся в БД;
- простое создание новых эталонных выгрузок.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS



Детектор заполненных форм

Простая анкета на протяжении долгого времени остаётся лёгким и эффективным способом сбора персональных данных граждан. При этом сам бланк, как правило, не несёт никакой ценности, а заполненный образец может содержать всеобъемлющую информацию о человеке. Поэтому с целью защиты именно заполненных анкет в компании InfoWatch была разработана технология *«Детектор заполненных форм»*. Её применение позволяет отслеживать передачу по сетевым каналам анкет, содержащих персональные данные, и вовремя уведомлять Офицера информационной безопасности об инцидентах утечки конфиденциальных данных.

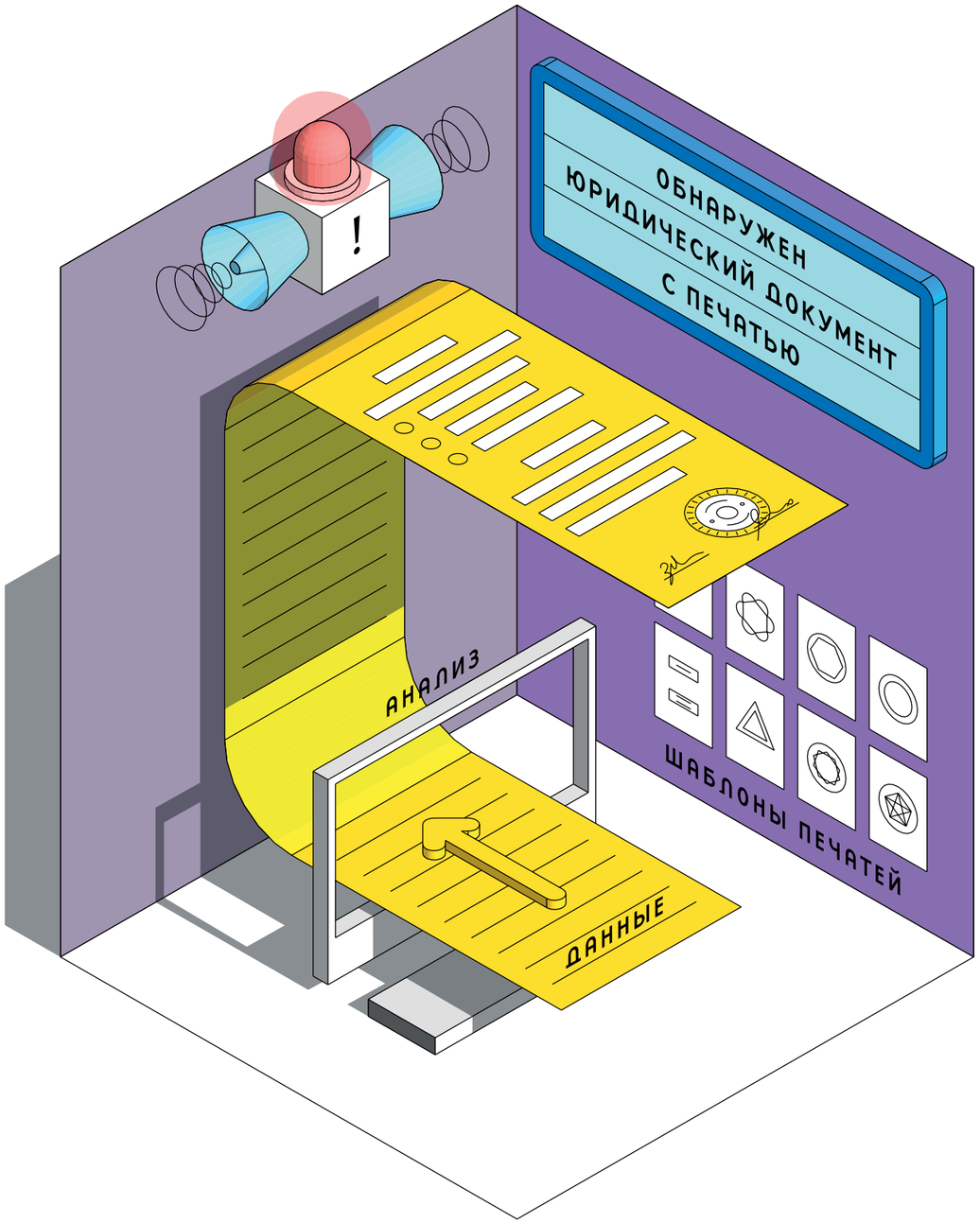
Настройка технологии не требует специальных знаний. Достаточно перевести типовую анкету в текстовый формат в кодировке UTF-8. Для этого необходимо извлечь текст из документа с формой при помощи стандартного экстрактора InfoWatch Traffic Monitor. Получившийся в результате файл следует добавить как эталонный документ через консоль InfoWatch Traffic Monitor.

При обнаружении признаков эталонной формы в потоке сетевого трафика, в консоли управления InfoWatch Traffic Monitor в разделе «Отчёты» появляется уведомление об утечке конфиденциальной информации. В деталях о перехваченных объектах содержится имя файла эталона, соответствующего детектированной эталонной форме. В секции «Контекст» в поле «Релевантность» можно посмотреть количество найденных заполненных полей эталонной формы.

Преимущества технологии

- высокая эффективность детектирования персональных данных в формате анкет;
- низкий показатель ложноположительных срабатываний; детектируются только заполненные формы;
- простое добавление новых эталонных форм.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS



Детектор печатей

Подсознательно всем нам понятно, что любой документ в организации имеет определённую ценность. Но, если у него нет никаких явных признаков того, что в нём содержится конфиденциальная информация, то делать выводы о степени его важности достаточно сложно. В то же время, наличие на документе официальной печати почти со стопроцентной вероятностью говорит о его юридической силе и ценности для компании.

Технология *«Детектор печатей»* в составе решения InfoWatch Traffic Monitor позволяет отслеживать передачу отсканированных документов, содержащих изображения эталонных печатей.

Добавление эталонной печати происходит через интерфейс копирайтного анализа в консоли управления InfoWatch Traffic Monitor. Изображение не должно содержать существенных сторонних элементов кроме самой печати. Для создания эталона достаточно поставить печать на белом листе бумаге и отсканировать его для загрузки в InfoWatch Traffic Monitor. При обнаружении признаков эталонной формы в потоке сетевого трафика, в консоли управления InfoWatch Traffic Monitor в разделе «Отчёты» появляется уведомление об утечке конфиденциальной информации. В деталях о перехваченных объектах содержится имя файла эталона, соответствующего детектированной печати. В секции «Контекст» можно оценить релевантность эталона по отношению к захваченному изображению.

Преимущества технологии

- высокая эффективность детектирования отсканированных изображений документов;
- простое создание эталонных печатей;
- инновационная технология, недоступная в решениях конкурентов[1].

1 По данным из открытых источников

Преимущества технологий анализа данных InfoWatch

На данный момент все существующие DLP решения нацелены больше на защиту статической информации. InfoWatch Traffic Monitor является единственной, представленной на рынке DLP-системой, которая благодаря сочетанию нескольких технологий контентного анализа позволяет эффективно использовать достоинства каждой из них и защищать информацию в течение всего жизненного цикла.

InfoWatch Traffic Monitor, вынося вердикт о конфиденциальности перехваченного сообщения, использует сразу несколько критериев:

- категории, присвоенные сообщению в результате лингвистического анализа;
- найденные в анализируемом тексте цитаты из базы эталонных документов;
- обнаруженные шаблоны текстовых объектов.

Это позволяет не только повысить эффективность решения за счет более точной настройки политик безопасности, но и проводить анализ бизнес-процессов, выявляя контентные маршруты информации как внутри, так и за пределами организации.

Большое внимание при разработке и интеграции технологий контентного анализа было уделено скорости работы решения. Используемые в продуктах InfoWatch технологии позволяют без потери качества минимизировать время, затрачиваемое на обработку текста.

Отличительной чертой технологий InfoWatch является особый уровень внимания к финансовой, государственной, страховой, нефтегазовой и телекоммуникационной отраслям экономики, основанный на многолетнем опыте сотрудничества с ведущими компаниями России, СНГ и стран дальнего зарубежья. Это гарантирует компаниям из данных сегментов обеспечение максимально высокой точности детектирования и защиты конфиденциальной информации, характерной для их отрасли.

INFOWATCH
BECAUSE YOUR DATA
IS YOUR BUSINESS

Контактная информация

+7 (495) 22-900-22
sales@infowatch.ru
www.infowatch.ru