Список цитируемых источников

1. *Бакунова, О. М.* Программный комплекс оценки антропогенной нагрузки на территориальные образования / О. М. Бакунова, О. Н. Образцова // Доклады БГУИР. — 2018. — № 1 (111). — С. 37—42.
2. UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (Aarhus Convention) [Электронный ресурс]. — Режим доступа: https://www.unece.org/env/pp/treatytext.html . — Дата доступа: 23.02.2018.
Национальная система мониторинга окружающей среды в Республике Беларусь: результаты наблюдений, 2016 год / [Электронный ресурс]. — Электрон. текстовые, граф. дан. (21 Мб). — Минск, Респ. центр по гидрометеорологии, контролю радиоактивного загрязнения и мониторингу окружающей среды. — 2017. — 1 элек- трон. опт. диск (CD-ROM): цв.; 12 см. — Систем. требования: Pentium II и выше; Windows XP.

УДК 004.6

А. М. Бакунов, И. Л. Калитеня, А. Ф. Палуйко, Е. Н. Александрович
Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники, Минск

ПРОБЛЕМЫ И ПРЕИМУЩЕСТВА ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ

Введение. Сейчас сложно найти крупную компанию, которая бы так или иначе не использовала технологии Big Data. Принимая во внимание перспективность этого направления, Big Data активно изучается и используется в различных сферах. Прорабатываются и анализируются риски использования новых технологий и разрабатываются способы решения таких проблем, чему и посвящена данная статья.

Основная часть. Благодаря информационным технологиям и современным решениям, стали рождаться огромные массивы данных и возможность их обрабатывать. С появление Big Data реальностью стала возможность решить давнюю цель и идею бизнеса — узнать всё о клиентах, конкурентах и тенденциях рынка. По данным исследователей Forrester, 100% компаний, которые используют для принятия решений аналитику данных, внедряют у себя и обработку Big Data.

Среди главных преимуществ больших данных для бизнеса, по информации, полученной в результате опроса исследовательской компании “The Economist Intelligence Unit” и консалтинговой компании “Accenture”, можно выделить: 1) поиск новых источников дохода (56%), 2) улучшение опыта клиентов (51%), 3) новые продукты и услуги (50%), 4) приток новых клиентов и сохранение лояльности старых (47%) [1].

Сейчас сложно найти крупную компанию, которая бы так или иначе не использовала технологии Big Data. Принимая во внимание перспективность этого направления, Big Data активно изучается и используется в различных сферах. Технология помогает управлять рисками, бороться с мошенничеством, сегментировать и оценивать клиентскую кредитную способность, управлять персоналом, прогнозировать очереди, рассчитывать бонусы для сотрудников и т. д.

К сожалению, существует множество проблем, которые препятствуют компаниям, специализирующимся на сборе данных, обеспечивать достойную защиту своим ценным накоплениям. Тем не менее у каждой проблемы есть свое решение.

Традиционных механизмов безопасности, таких как брандмауэры и антивирусное программное обеспечение, устанавливаемое на компьютерах, недостаточно для эффективной защиты больших данных. Основная проблема состоит в том, что такие способы создавались для защиты небольших объемов статической информации — файлов, сохраненных на жестких дисках, а не большого информационного потока, прибывающего из облака. Меры безопасности должны быть достаточно гибкими и оперативными, что позволит обеспечить бесперебойность получения данных и безопасность многочисленных «точек входа».

Существенным риском для больших данных является их утрата (частичная или полная). Причины могут быть различны: от активности злоумышленников до чрезвычайной ситуации. Единственный способ защититься — резервирование данных. Очевидно однократное резервирование. Если оценка риска велика и сильно влияет на бизнес, то рекомендовано двукратное и трехкратное резервирование.

В некоторых случаях несколько примитивных ошибок могут испортить долгую кропотливую работу. Большие данные не являются исключением, а учитывая, что объемы больших данных способны достигать огромных размеров, ошибки весьма вероятны (как в содержании и структуре самих данных, так и в инструментах работы с ними).

Для снижения риска ошибок больших данных рекомендуется: проводить периодические ревизии данных; контролировать ключевые параметры данных; вести журнал выявленных ошибок и их устранения; разрабатывать инструменты и алгоритмы устранения ошибок и некорректных состояний данных; оценивать результативность инструментов; применять специальные средства тестирования данных и инструментов, которые разрабатываются самостоятельно; использовать инструменты последовательно, подконтрольно и пошагово с постоянным контролем обрабатываемых данных в целом или по выборкам [2].

Все сводится к тому, что нужно фокусироваться на безопасности ресурсов и приложений, а не устройств, изолировать критически важные устройства и серверы, внедрять средства управления

информацией и событиями информационной безопасности в режиме реального времени, а также обеспечивать баланс реактивной защиты.

Эксперты по облачным технологиям считают, что самым разумным проводником в вопросах улучшения безопасности Big Data является антивирусная индустрия. На протяжении десятилетий антивирусное программное обеспечение ведет борьбу с различными видами угроз. Есть множество поставщиков антивирусного программного обеспечения, предлагающих самые разные решения [3]. И все они могут оказаться полезными, когда речь заходит о неприятных цифровых ошибках или серьезных угрозах.

Также высоко оценивается открытость антивирусной индустрии в отношении данных. Вместо блокировки своих секретов безопасности для получения конкурентного преимущества производители антивирусного программного обеспечения (в том числе неправительственные организации, государственные учреждения и даже частные предприятия) свободно обмениваются друг с другом данными об угрозах. Лидеры отрасли могут сотрудничать, чтобы бороться с новыми и опасными вредоносными программами во всем мире, обеспечивая максимальную безопасность Big Data.

Заключение. Компаниям необходимо разрабатывать процессный подход к анализу и обработке данных, а также автоматизировать процессы, касающиеся обеспечения безопасности больших данных в рамках устоявшихся практик. Автоматизация может включать в себя в том числе элементы машинного обучения — искусственный интеллект с помощью которого возможно извлекать из добавляемых в кластер данных признаки «конфиденциальности», выявлять паттерны, не характерные для нормальной работы с данными, составлять профили пользователей и фиксировать отклонения в работе пользователей от их нормального профиля поведения, т. е. выявлять мотивы пользователей при работе с данными. Именно для повышения эффективности принимаемых решений и снижения рисков неправильных решений компании обращаются к Big Data. Но даже видя реальные риски, разумно использовать обработку больших данных, ведь технологии развиваются, появляются способы защиты информации.

Список цитируемых источников

1. Data Science for Business. What You Need to Know about Data Mining and Data-Analytic Thinking. // O'Reilly Media. — 2013. — С. 414.
2. Shiwen Mao, Min Chen, Victor C.M. Leung, Yin Zhang, Big Data: Related Technologies, Challenges and Future Prospects. — 2014. — С. 89.
3. Безопасность больших данных [Электронный ресурс]. — Режим доступа: http://rtbinsight.ru/articles/big-data-security.html. — Дата доступа: 22.02.2018.

УДК 504.064.2.001.19

О. М. Бакунова, А. М. Бакунов, М. А. Калугина, О. Н. Образцова

Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники, Минск

ПРИНЦИПЫ ПОСТРОЕНИЯ БАЗЫ ЗНАНИЙ ПРОГРАММНОГО КОМПЛЕКСА ОЦЕНКИ АНТРОПОГЕННОЙ НАГРУЗКИ

Введение. Задачи мониторинга состояния окружающей среды и оценки антропогенной нагрузки в настоящее время весьма актуальны в связи с увеличением промышленного, транспортного, сельскохозяйственного, радиационного и рекреационного загрязнения. Для компьютерного моделирования данной предметной области предлагается использовать метод конечных предикатов, который в первую очередь позволяет решить задачу приведения множества неоднородных показателей в единую форму.

Основная часть. Предлагаемый подход к построению базы знаний диагностической системы использует представление знаний в виде конечного предиката, определенного на множестве характеристик. При решении задач распознавания образов, связанных с поиском импликативных закономерностей, необходимо столкнуться с проблемой проверки полноты системы запретов, который можно рассматривать как обобщение известной NP-полной задачи о выполнимости КНФ Булевой функции: как проблемы о выполнимости КНФ конечного предиката, который на языке матриц формулируется: пусть K — Булева матрица, разбитая по столбцам. Требуется выяснить, есть ли хотя бы одно покрытие для него, т. е. существует ли подмножество столбцов, взятых ровно по одному из каждого раздела, которые вместе содержали бы хотя бы одну единицу в каждой строке матрицы. Экспериментально установлено, что для классической задачи осуществимости существует так называемый критический интервал значений параметров, в котором лежат действительно сложные индивидуальные задачи. Поэтому имеет смысл определить закономерности между размерами исходной матрицы и ее целесообразности. В связи с этим были рассчитаны математические ожидания некоторых случайных величин, одной из которых, например, является среднее число матриц E заданного размера, не имеющих покрытия. На основе метода конечных предикатов,