УДК 004.056 Вестник УрФО № 4(46) / 2022, с. 21–25

Зырянова Т. Ю.

DOI: 10.14529/secur220402

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЕ МЕТОДЫ В СИСТЕМАХ МЕНЕДЖМЕНТА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В сфере технологий обеспечения безопасности информации огромное значение имеют технологии анализа больших информационных массивов данных. Информационно-аналитические системы безопасности – одно из современных и активно развивающихся направлений в сфере теории, методологии и практики защиты информации. Статья посвящена обзору основных понятий, положений и категорий, связанных с информационно-аналитическими системами и описанию основных подходов к выполнению интеллектуального анализа данных с применением современных информационных технологий. Обоснована необходимость применения информационно-аналитических систем при разработке систем менеджмента информационной безопасности, приведены их основные особенности.

Ключевые слова: *информационно-аналитическая система, система менеджмента информационной безопасности, база данных, хранилище данных, анализ данных.*

Zyryanova T. Yu.

INFORMATION-ANALYTICAL METHODS IN INFORMATION SECURITY MANAGEMENT SYSTEMS

In the field of information security technologies, technologies for analyzing large information data arrays are of great importance. Information-analytical security systems is one of the modern and actively developing areas in the field of theory, methodology and practice of information security. The article is devoted to an overview of the basic concepts, provisions and categories associated with information and analytical systems and a description of the main approaches to the implementation of data mining using modern information technologies. The necessity of using information-analytical systems in the development of information security management systems is substantiated, their main features are given.

Keywords: *information-analytical system, information security management systems, database, data warehouse, data analysis.*

Понятие системы менеджмента информационной безопасности (СМИБ) введено Международным стандартом ИСО/МЭК 27000 «Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология». Согласно стандарту СМИБ – это часть общей системы менеджмента, основанная на подходе бизнес-рисков по созданию, внедрению, функционированию, мониторингу, анализу, поддержке и улучшению информационной безопасности. Основные функции СМИБ должны состоять в оценке степени критичности ситуации, связанной с нарушением информационной безопасности, оценке уровня риска нарушения информационной безопасности и в поддержке принятия решений относительно действий в данной ситуации. Принятие решений в такой системе затруднено по ряду причин: не всегда возможно сформировать полное множество угроз информационной безопасности, количественно оценить степень критичности возникшей ситуации, построить прогноз ее развития. Для решения таких слабоструктурированных задач, характеризующихся высокой степенью неопределенности исходных данных может быть эффективно применение подходов, характерных для современных информационно-аналитических систем.

Существуют различные определения понятия информационно-аналитической системы (ИАС), которые во многом характеризуют различные сферы применения ИАС, такие как экономика, управление, бизнес и многие другие. Приведем некоторые, наиболее общие, из них.

ИАС – компьютерная система, позволяющая получать информацию, создавать ее, проводить ее обработку и анализ.

ИАС – автоматизированная система, позволяющая экспертам быстро анализировать большие объемы знаний.

Прообразом современных ИАС систем стали развивающиеся с конца 70-х годов XX века системы поддержки принятия решений (СППР) – интерактивные автоматизированные системы, помогающие лицу, принимающему решения, использовать данные и математические модели для решения структурированных проблем. Принципиальное отличие современных ИАС от СППР состоит в том, что, как правило, в качестве исходных данных в ИАС используются большие объемы изначально неструктурированной информации, которые необходимо преобразовывать и анализировать для получения значимых результатов.

В соответствии со своим назначением ИАС должны решать задачи:

1. Поиска, сбора, передачи и хранения информации;

2. Обобщения и сортировки информации;

3. Анализа информации, формирования выводов и заключений.

Подробному описанию информационно-аналитических систем безопасности (ИАСБ) посвящена, например, работа [1].

Для решения первой и второй задач в ИАС применятся технология хранилищ данных. Для решения третьей задачи применяется целый комплекс различных технологий анализа. Это технология произвольных информационных запросов SQL, технология гиперкубического представления и многомерного анализа данных OLAP, технологии интеллектуального анализа данных Data Mining.

Основоположниками теории хранилищ данных стали Уильям Х. Инмон (Билл Инмон) в 1991 году и Ральф Кимбалл в 1998 году.

Билл Инмон сформулировал определение хранилища данных как «предметно-ориентированные, интегрированные, стабильные, поддерживающие хронологию наборы данных, организованные для целей поддержки управления, призванные выступать в роли единственного источника истины» [2].

Определение Ральфа Кимбалла более простое, короткое и обобщенное: «хранилища данных – место, где люди могут получать доступ к своим данным» [3].

Существуют принципиальные различия между хранилищами данных и ставшими уже привычными традиционными базами данных. Они состоят в следующем.

1. Базы данных применяются, как правило, для обеспечения повседневной работы и не предназначены для решения аналитических задач. Хранилища данных – это структуры данных, предназначенные непосредственно для поддержки принятия решений на основе анализа помещенной в них информации.

2. Информация в базах данных терпит постоянные изменения в режиме реального времени в процессе работы пользователей. Для информации в хранилищах данных характерна относительная стабильность. Добавление данных в хранилища производится в определенные моменты времени.

3. В базах данных хранятся, как правило, внутренние данные той или иной информационной системы. Хранилища данных используются большие массивы данных из внешних источников.

4. В базах данных набор запросов пользователей, с которыми они могут обратиться, четко регламентирован и известен уже на этапе проектирования базы данных. Для хранилищ данных характерны нерегламентированные запросы аналитиков.

На этапе становления технологии баз данных применялись различные подходы к структуризации информации. В результате ее эволюции появилась модель данных, получившая название реляционной (от слова relationship – отношение), впервые описанная в статье Эдгара Кодда «Реляционная модель данных для больших коллективных банков данных» [4]. Отличительной особенностью этой модели от предшествующих стало представление данных об объектах реального мира с учетом связей между ними в форме самого естественного их представления – прямоугольных таблиц. С целью реализации реляционной модели данных был разработан стандарт высокоуровневого языка программирования реляционных баз данных, получившего название SQL – Structured Query Language – Структурированный язык запросов – единый интегрированный язык, содержащий все средства для работы с реляционными базами данных, такие как:

1. Операторы формулирования запросов;

2. Средства определения схемы базы данных и манипулирования схемой;

3. Операторы для определения ограничений и триггеров;

4. Средства определения представлений;

5. Средства авторизации доступа к отношениям и их атрибутам;

6. Средства управления транзакциями.

Именно язык запросов SQL используется для сбора исходных данных в современных информационно-аналитических системах.

Основной технологией, применяемой в информационно-аналитических системах является OLAP – On-Line Analytical Processing – Оперативная аналитическая обработка данных. Ее основоположником также стал Эдгар Кодд, который в 1993 году дал этой технологии следующее определение: «OLAP – динамический анализ, включающий в себя возможность выявления новых или непредвиденных отношений между переменными, способность работать с большими объемами данных, создавать неограниченное число измерений (частей консолидации) и определять условия и выражения пересечения переменных» [5]. Вообще говоря, OLAP – это не отдельно взятый продукт и даже не конкретная технология, а скорее концепция, основанная на математической модели представления данных в виде многомерных кубов.

Массив данных в терминологии OLAP называется кубом. Кубы OLAP не обязательно имеют одинаковое число элементов по всем осям, могут быть многомерными. Ввиду своей многомерности сам куб для анализа непригоден. Из него извлекаются двумерные таблицы по интересующим аналитика параметрам (операция разрезания куба), из которых затем создаются многоуровневые объединения, называемые иерархиями. Исходные данные берутся из нижнего уровня иерархий и суммируются для получения данных на более высоком уровне.

Для поиска закономерностей и извлечения знаний в ИАС широкое применение нашла технология Data Mining, получившая свое название от английского слова mining – дословно – добыча полезных ископаемых – процесс, требующий просеивания большого количества сырого материала для поиска ценностей.

Согласно определению, данному Григорием Пятецким-Шапиро: «Data Mining – процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности» [6].

Также известно определение Data Mining, данное Вячеславом Анатольевичем Дюком: «Поиск в больших объемах данных неочевидных, объективных и полезных на практике закономерностей» [7]. Слово «неочевидных» в данном определении означает, что закономерности не обнаруживаются стандартными методами обработки информации, например, статистическими, или экспериментальным путем.

Согласно В. А. Дюку выделяют пять видов

закономерностей в данных, которые могут быть обнаружены применением технологии Data Mining:

1. Ассоциация – взаимосвязь событий друг с другом;

2. Последовательность – наличие цепочки событий, связанных друг с другом последовательно во времени;

3. Классификация – наличие признаков, характеризующих группу объектов;

4. Кластеризация – аналог классификации, при котором изначально не задано ни количество групп объектов, ни признаки, по которым объекты будут распределяться на группы;

5. Прогнозирование – наличие исторической информации, содержащей шаблоны в поведении процесса, которые можно использовать для предсказания параметров процесса в будущем.

Методы, применяемые в ИАС для поиска таких закономерностей разнообразны. Это и классические статистические методы, и методы интеллектуального анализа данных, основанные на применении принципов работы человеческого мышления.

Для решения задач поиска ассоциаций могут применяться методы корреляционно-регрессионного анализа. Корреляционный анализ позволяет оценить существенность влияния одного объекта, процесса или явления на другой, регрессионный анализ применяется для установления вида зависимости между объектами, процессами, явлениями.

Наиболее эффективным статистическим методом прогнозирования является метод анализа временных рядов. Временным рядом называется последовательность измерений значений переменной (процесса) за определенный период времени через одинаковые промежутки.

Термин «кластерный анализ» впервые ввел математик Р. Трион в 1939 году [8]. Этот метод включает в себя набор различных алгоритмов классификации и предназначен для того, чтобы организовать наблюдаемые данные в группы (кластеры), содержащие в себе объекты с похожими свойствами. Кластеризация относится к так называемой разведочной добыче данных, так как помогает выделить информацию о связях объектов, событий или явлений, которые не видны визуально, в огромных объемах данных. Принципиальное отличие методов добычи данных от классического статистического анализа состоит в том, что аналитик не знает заранее, какую информацию он ищет.

Интеллектуальные методы анализа данных применяются в тех случаях, когда предполагается, что из имеющихся данных можно будет извлечь знания для принятия решения в условиях неопределенности. В отличие от классического разведочного анализа данных аналитика не интересует конкретный вид зависимостей между параметрами задачи. Главной целью не является выяснение природы участвующих в задаче функций или конкретной формы зависимостей. Основное внимание уделяется поиску решений. Оно достигается путем совмещения методов классического разведочного анализа и интеллектуальных методов. Data Mining часто называют смесью статистики и искусственного интеллекта. К методам Data Mining относятся, например, теория нечетких множеств и нечеткой логики, теория искусственных нейронных сетей, эволюционное программирование, генетические алгоритмы.

Необходимость создания информационно-аналитических систем в СМИБ обусловлена в первую очередь, следующими предпосылками:

1. В настоящее время происходит активная интеллектуализация информационных технологий в целом;

2. Как следствие постоянно реализуются новые интеллектуальные системы защиты информации;

3. Проблемой остается недостаточная эффективность систем обеспечения безопасности, связанных с получением в реальном времени аналитических оценок и прогнозированием состояния, направления развития и уровней угроз безопасности информации.

При разработке ИАС, предназначенных для решения проблем управления информационной безопасностью не следует упускать из внимания следующие их особенности.

1. Возможность не только анализа оперативных данных, относящихся к обеспечению безопасности объекта, но и решения задач прогнозирования уровня безопасности с использованием математических моделей.

2. Вовлечение в процесс анализа разноплановых данных из внешних источников, непосредственно не используемых в системе управления предприятием.

3. Ориентирование на хранение всей совокупности документов.

4. Поддержка интеллектуальных функций

(автоматическая классификация документов, ассоциативный поиск, поиск по образцу и т.д.) с распространением на различные формы представления информации.

Как минимум, инструментальные средства таких ИАС должны включать:

1. Средства статистического анализа данных (традиционные отчеты, диаграммы);

2. Средства динамического анализа данных (динамические системы поддержки принятия решений);

3. Средства моделирования и прогнозирования;

4. Средства визуализации связей и отношений между объектами.

В статье проанализированы наиболее актуальные направления развития технологий информационно-аналитических систем в их применении к проблемам обеспечения информационной безопасности. В настоящее время наблюдается повышенный спрос на разработку интеллектуальных систем защиты информации, способных оказывать поддержку принятия решений во множестве ситуаций – от фиксации нетипичного поведения пользователей до выявления различных аномалий в процессах передачи, хранения и обработки информации. Описанные в статье методы могут стать эффективной математической и методологической основой для разработки таких систем.

Литература

1. Васильев В. И. Интеллектуальные системы защиты информации. – М.: Машиностроение, 2017.
2. William H. Inmon. Building the Data Warehouse. – Wiley Publishing, Inc., 2006.
3. Kimball R. The Data Warehouses Lifecycle Toolkit. – Wiley Publishing, Inc., 2008.
4. Codd E. F. A Relational Model of Data for Large Shared Data Banks // Communications of the ACM, 1970. – С. 377-387.
5. Codd E. F., Codd S. B., Salley C. T. Providing OLAP (on-line analytical processing) to user-analysts. – E. F. Codd & Associates, 1993.
6. Piatetsky–Shapiro G. Advances In Knowledge Discovery and Data Min ing, 1996.
7. Дюк В. А., Самойленко А. П. Data Mining. – СПб: Питер, 2001.
8. Tryon R. C. Cluster analysis. — London: Ann Arbor Edwards Bros, 1939.

References

1. Vasil'ev V. I. Intellektual'nye sistemy zashhity informacii. – M.: Mashinostroenie, 2017.
2. William H. Inmon. Building the Data Warehouse. – Wiley Publishing, Inc., 2006.
3. Kimball R. The Data Warehouses Lifecycle Toolkit. – Wiley Publishing, Inc., 2008.
4. Codd E. F. A Relational Model of Data for Large Shared Data Banks // Communications of the ACM, 1970. – С. 377-387.
5. Codd E. F., Codd S. B., Salley C. T. Providing OLAP (on-line analytical processing) to user-analysts. – E. F. Codd & Associates, 1993.
6. Piatetsky–Shapiro G. Advances In Knowledge Discovery and Data Min ing, 1996.
7. Djuk V. A., Samojlenko A. P. Data Mining. – SPb: Piter, 2001.
8. Tryon R. C. Cluster analysis. — London: Ann Arbor Edwards Bros, 1939.

ЗЫРЯНОВА Татьяна Юрьевна, кандидат технических наук, доцент кафедры «Информационные технологии и защита информации» Уральского государственного университета путей сообщения. 620034, г. Екатеринбург, ул. Колмогорова, д. 66. E-mail: tzyryanova@usurt.ru

ZYRYANOVA Tatiana Yuryevna, candidate of technical sciences, associate professor of Department «Information Technology and Information Security» of the Ural State University of Railway Transport. 620034, Yekaterinburg, st. Kolmogorova, 66. E-mail: tzyryanova@usurt.ru