ЭКОНОМИКА И УПРАВЛЕНИЕ

УДК 330.44

Л.В. Калягина, П.Е. Разумов

КАТЕГОРИЯ «ДАННЫЕ»: ПОНЯТИЕ, СУЩНОСТЬ, ПОДХОДЫ К АНАЛИЗУ

В работе дается несколько подходов к определению понятия «данные» и близких по смысловому содержанию к нему терминов. Рассматривается история использования, понимания термина «анализ данных» и его синонимов. Предлагается последовательность процедур, позволяющая получить информацию о структуре данных.

***Ключевые слова**: данные, анализ, структура, информация, знания, принятие решений, статистические процедуры обработки данных, эксперимент, исследователь.*

L.V. Kalyagina, P.E. Razumov

CATEGORY "DATA": CONCEPT, ESSENCE, APPROACHES TO THE ANALYSIS

Some approaches to the definition of the "data" concept and the terms close to it in the semantic contents are given in the article. The history of use, understanding of the "data analysis" term and its synonyms are considered. The procedure sequence allowing to receive the information on the data structure is offered.

Key words: data, analysis, structure, information, knowledge, decision-making, statistical procedures of data processing, experiment, researcher.

Данные – это воспринимаемые человеком факты, события, сообщения, измеряемые характеристики, регистрируемые сигналы. Специфика данных состоит в том, что они, с одной стороны, существуют независимо от наблюдателя, а с другой – становятся собственно «данными» лишь тогда, когда существует целенаправленно собирающий их субъект. Из всего множества происходящих событий, из богатства свойств реальных объектов исследователь выделяет только вполне конкретные данные – ту малую часть огромного, потенциально существующего материала, которая, на его взгляд, необходима для решения поставленной перед ним задачи. Таким образом, данные оказываются тем основанием, на котором затем возводится все здание заключений, выводов и решений. Они вторичны по отношению к цели исследования и предметной области, но первичны к методам их обработки и анализа, извлекающим из данных только ту информацию, которая потенциально доступна в рамках отобранного материала. Если сбор данных произведен неверно, если они не отражают существенных взаимосвязей предметной области, то и анализ оказывается бесполезным [1].

Да́нные (в пер. «калька» от лат. «факт», от англ. data, от слав. «дати» (греч. *διδόναι*), от рус. «дать, давать») – представление фактов и идей в формализованном виде, пригодном для передачи и обработки в некотором информационном процессе [11].

Синонимами термина «данные» являются слова: информация, сведения. Антонимами – параметры, код.

К определению значения термина «данные» существует несколько подходов:

1. Совокупность *сведений*, информация.

2. Совокупность *свойств*, способностей как условие для достижения какой-либо цели или выполнения какой-либо работы, например: «У него неплохие данные для бокса».

3. Пассивная *часть программного обеспечения*, совокупность значений определенных ячеек памяти, преобразование которых осуществляет код.

Термин «данные» в широком смысле означает фактический материал, являющийся основой для обсуждения или принятия решений; в статистике – это информация, пригодная для анализа и интерпретации.

К базовым понятиям, которые используются в экономической информатике, относятся: данные, информация и знания. Хотя эти понятия используются как синонимы, между ними существуют принципиальные различия.

Считается, что термин «данные» происходит от слова *data* – факт, а «информация» (*informatio*) означает разъяснение, изложение, т.е. сведения или сообщение.

Остановимся на следующих определениях [11]:

Данные – это совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки. Преобразование и обработка данных позволяют получить информацию.

Информация – это результат преобразования и анализа данных. Отличие информации от данных состоит в том, что данные – это фиксированные сведения о событиях и явлениях, которые хранятся на определенных носителях, а информация появляется в результате обработки данных при решении конкретных задач. Например, в базах данных хранятся различные данные, а по определенному запросу система управления базой данных выдает требуемую информацию.

Существуют и другие определения информации, например: информация – это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний.

Знания – это зафиксированная и проверенная практикой обработанная информация, которая использовалась и может многократно использоваться для принятия решений.

Знания – это вид информации, которая хранится в базе знаний и отображает знания специалиста в конкретной предметной области. Знания – это интеллектуальный капитал.

Формальные знания могут быть в виде документов (стандартов, нормативов), регламентирующих принятие решений или учебников, инструкций с описанием решения задач. Неформальные знания – это знания и опыт специалистов в определенной предметной области.

Необходимо отметить, что универсальных определений этих понятий (данных, информации, знаний) нет, они трактуются по-разному.

Принятие решений осуществляется на основе полученной информации и имеющихся знаний.

Принятие решений – это выбор наилучшего в некотором смысле варианта решения из множества допустимых на основании имеющейся информации.

Взаимосвязь данных, информации и знаний в процессе принятия решений представлена на рисунке 1.

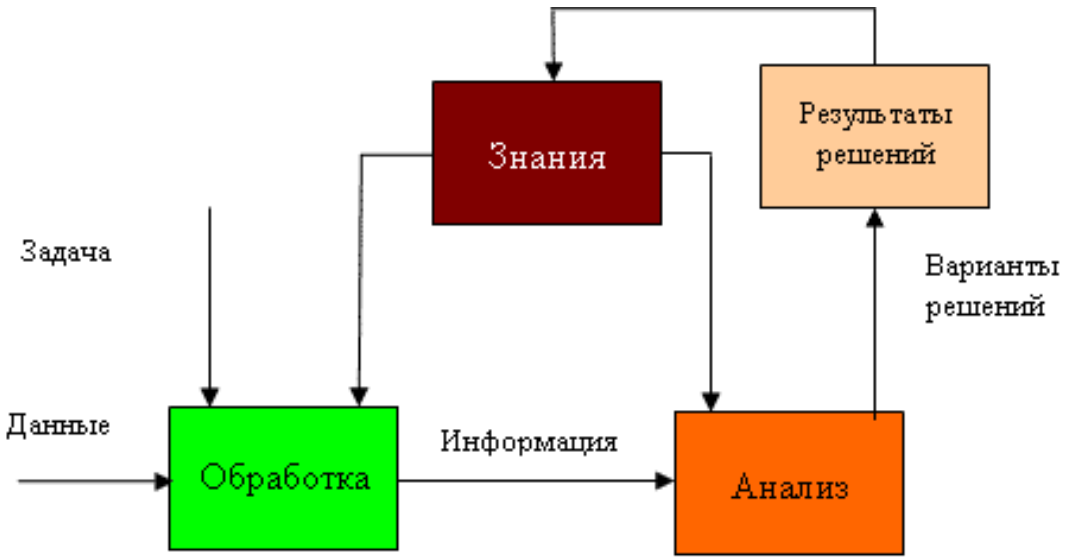

Рис. 1. Взаимосвязь понятий

Для решения поставленной задачи фиксированные данные обрабатываются на основании имеющихся знаний, далее с их помощью полученная информация анализируется. На основании анализа предлагают-

ся все допустимые решения, а в результате выбора принимается одно наилучшее. Результаты решения пополняют знания.

В зависимости от сферы использования информация может быть различной: научной, технической, управляющей, экономической и т.д.

В итоге можно сказать, что изначально *данные* величины, т.е. величины, заданные заранее, вместе с условием задачи. Противоположность – переменные величины. Данные – это зарегистрированные сигналы. Данные могут рассматриваться как записанные наблюдения, которые не используются, а пока хранятся.

Если данные ориентированы на их понимание человеком непосредственно при их восприятии или после их некоторого преобразования, то они содержат в себе информацию. Возможна ситуация, когда данные не содержат информацию, на настоящее время доступную человеку. Человек способен извлекать информацию не из всех доступных для него данных. Шифрование информации делает ее недоступной для всех, кто не имеет ключа (кода) расшифровывания. Шифротекст содержит информацию, но она недоступна.

В экономике под *данными* представляется результат измерения (наблюдения, регистрации, описания и т.п.) *свойств* исследуемых объектов. То есть в данном случае данные характеризуют природу и *структуру* реально анализируемой информации. При экономических исследованиях в анализе данных наиболее часто применяется статистический подход к интерпретации исходной информации, который подразумевает необходимость счета таких статистических характеристик, как среднее, дисперсия, ковариация и т.д.

При статистическом подходе к интерпретации исходной информации термины *данные*, *наблюдения*, *реализация* являются синонимами. Наблюдения служат реализацией некоторой *случайной величины* и они *поставляют данные* для изучаемой проблемы.

В свою очередь, *случайная величина* – это величина, которая принимает в результате опыта одно из множества значений, причём появление того или иного значения этой величины до её измерения нельзя точно предсказать. Неопределенность связана с действием случайных причин, которые не могут быть учтены заранее. Поэтому до компьютеризации населения понятие «анализ данных» в литературе сопоставлялся с математической статистикой либо частью статистики, либо, наоборот, считалось более широким понятием, чем статистика [1]. В отечественной литературе термин «анализ данных» был синонимичен термину «прикладная статистика», подчеркивающему практическую направленность соответствующих методов обработки данных. При массовом распространении персональных компьютеров термин «анализ данных» стал использоваться как синоним к моделированию данных.

В современной трактовке «анализ данных» – область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных; процесс исследования, фильтрации, преобразования и моделирования данных с целью извлечения полезной информации и принятия решений [8].

При анализе данных обычно пытаются получить информацию, позволяющую вскрыть структуру той основы, на которой формируются данные. При этом модель структуры данных, как правило, определена неполно. Проникновение в *структуру* данных, ее информационное раскрытие составляет *цель анализа данных.*

В контексте статьи будем считать, что анализ данных – это совокупность методов и средств извлечения из определенным образом организованных данных информации для принятия решения. Под *определенной организацией* понимается форма, в которой представлены данные.

При интенсивно развивающихся информационных технологиях в настоящее время появилось множество разновидностей в подходах к анализу данных. Перечислим основные из них:

- **Интеллектуальный анализ** данных – это особый метод анализа данных, который фокусируется на моделировании и открытии данных, а не на их описании [2, 3, 7, 9].
- **Бизнес-аналитика** охватывает анализ данных, который полагается на агрегацию [6].
- **В статистическом смысле** некоторые разделяют анализ данных на описательную статистику, исследовательский анализ данных и проверку статистических гипотез [1, 4, 5].
- **Исследовательский анализ** данных занимается открытием новых характеристик данных, а проверка статистических гипотез – подтверждением или опровержением существующих гипотез [1, 3, 5].
- **Прогнозный анализ** фокусируется на применении статистических или структурных моделей для предсказания или классификации, а анализ текста применяет статистические, лингвистические и структур-

ные методы для извлечения и классификации информации из текстовых источников, принадлежащих к неструктурированным данным [1, 2, 10].

❖ Data Mining (рус. *добыча данных, интеллектуальный анализ данных, глубинный анализ данных*) – собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Термин введён Григорием Пятецким-Шапиро в 1989 году.

Английское словосочетание «Data Mining» пока не имеет устоявшегося перевода на русский язык. Поэтому при переводе используются следующие словосочетания: *просев информации*, *добыча данных*, *извлечение данных*, а также *интеллектуальный анализ данных*. Более полным и точным является словосочетание *обнаружение знаний в базах данных* (англ. knowledge discovering in databases, KDD) [2, 4, 10].

Основу методов Data Mining составляют всевозможные методы классификации, моделирования и прогнозирования, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов, эволюционного программирования, ассоциативной памяти, нечеткой логики. К методам Data Mining нередко относят *статистические методы* (дескриптивный анализ, корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов, анализ выживаемости, анализ связей). Такие методы, однако, предполагают некоторые априорные представления об анализируемых данных, что несколько расходится с целями Data Mining (обнаружение ранее неизвестных нетривиальных и практически полезных знаний).

Одно из важнейших назначений методов Data Mining состоит в наглядном представлении результатов вычислений (визуализация), что позволяет использовать инструментарий Data Mining людьми, не имеющими специальной математической подготовки. В то же время применение статистических методов анализа данных требует хорошего владения теорией вероятностей и математической статистикой.

❖ **Интеграция данных** – это предшественник анализа данных, а сам анализ данных считается связанным с визуализацией данных и распространением данных. Интеграция данных включает объединение данных, находящихся в различных источниках, и предоставление данных пользователям в унифицированном виде. Этот процесс становится существенным как в коммерческих задачах (когда двум похожим компаниям необходимо объединить их базы данных), так и в научных (комбинирование результатов исследования из различных биоинформационных репозиториев – для примера). Роль интеграции данных возрастает, когда увеличивается объём и необходимость совместного использования данных. Это стало фокусом обширной теоретической работы, а многочисленные проблемы остаются нерешенными [3, 9].

Разнообразные задачи анализа данных занимают центральное место при проведении экспериментальных исследований в любой области знаний. Вместе с тем в ряде случаев попытки использования широко известных статистических процедур обработки данных заканчиваются неудачно, так как возможности статистических методов в большинстве случаев не оправдывают требования исследователей. Это обусловлено прежде всего формализованным подходом к применению методов, относительно поверхностным знанием предпосылок их применения. С другой стороны, исследование детерминированных объектов выдвигает задачи анализа данных при неполном знании механизма явлений, происходящих в изучаемом объекте.

Характеризуя с этих позиций весь арсенал статистических методов, можно видеть, что практически ни один раздел статистики не дает прямого решения задачи анализа данных в такой ее постановке [1]. Так, раздел описательных статистик позволяет с помощью моментов в сжатой форме представить данные, но не затрагивает вопрос о структуре данных. Раздел статистических выводов дает процедуры для проверки статистических гипотез, но в рамках заданной модели структуры данных. Методы идентификации и многомерной статистики позволяют выбрать наиболее пригодную модель для описания выборочных данных. Вопрос о пригодности выбранной модели для отображения общей структуры данных нередко остается открытым.

Отсутствие полностью определенной заранее модели приводит к организации процессов анализа данных в виде последовательностей процедур, позволяющей получить информацию о структуре данных. Именно грамотная организация работы человека, который не является специалистом по анализу данных, может помочь в решении поставленной задачи. Основные этапы решения задачи анализа данных показаны в левой части рисунка 2.

В правой части каждый из них разбит на более мелкие стадии.

Этап 1 Постановка задач	1.1. Определение цели исследования	1.1.
	1.2. Определение состава данных	1.2.
	1.3. Сбор данных	1.3.
	1.4. Выбор средств анализа данных	1.4.
Этап 2 Ввод данных в обработку	2.1. Ввод данных в память компьютера	2.1
	2.2. Работа с архивом данных	2.2.
	2.3. Формирование задания обработки	2.3.
Этап 3 Качественный анализ	3.1. Определение простейших характеристик данных	3.1.
	3.2. Визуализация данных	3.2.
	3.3. Анализ структуры данных	3.3.
Этап 4 Количественное описание данных	4.1. Выбор модели данных	4.1.
	4.2. Выполнение обработки	4.2.
Этап 5 Интерпретация результатов	5.1. Анализ результатов	5.1.
	5.2. Принятие решений	5.2.

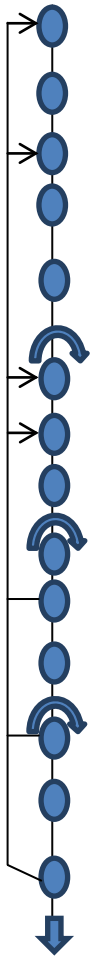

Рис. 2. Основные типы решения задачи анализа данных и их взаимосвязи

При построении таких последовательностей необходимо уделять большое значение сбору данных, приемам накопления однородных данных, схемам классификации и группировки однородных данных, планам для выделения источников различного типа неоднородностей. Перечисленные приемы и методы развиваются в различных аспектах теории решающих функций, методов обработки данных, планирования эксперимента. Неформализованным в анализе данных остается выбор критерия для отбора в некотором смысле наилучшей последовательности процедур, раскрывающих структуру данных. Следует отметить, что никакие методы и программы обработки данных не помогут дилетанту извлечь из данных (и хороших, и плохих) информацию. Определяющим фактором всегда является профессионализм исследователей.

Литература

1. *Александров В.В., Алексеев А.И., Горский Н.Д.* Анализ данных на ЭВМ (на примере системы СИТО). – М.: Финансы и статистика, 1990. – 192 с.
2. *Дюк В., Калягина Л.В.* Современные технологии обнаружения знаний в базах данных // Вестник КрасГАУ. – 2004. – № 4.
3. *Ежов А.А., Шумский С.А.* Избранные лекции по нейрокомпьютингу. – М.: Изд-во МИФИ, 1998.
4. *Журавлёв Ю.И., Рязанов В.В., Сенько О.В.* Распознавание. Математические методы. Программная система. Практические применения. – М.: Фазис, 2006. – 176 с.

5. *Зиновьев А.Ю.* Визуализация многомерных данных. – Красноярск: Изд-во КГТУ, 2000. – 180 с.
6. *Паклин Н.Б., Орешков В.И.* Бизнес-аналитика: от данных к знаниям (+ CD). – СПб.: Питер, 2009. – 624 с.
7. *Ситник В.Ф., Краснюк М.Т.* Інтелектуальний аналіз даних (дейтамайнінг): навч. посібник. – Киев: КНЕУ, 2007. – 376 с.
8. Социология: энцикл. / сост. *А.А. Грицанов, В.Л. Абушенко, Г.М. Евелькин* [и др.]. – Минск: Книжный Дом, 2003. – 1312 с.
9. *Черняк Л.* Интеграция данных: синтаксис и семантика // Открытые системы. – 2009. – № 10.
10. *Ian H.* Witten, Eibe Frank and Mark A. Hall Data Mining: Practical Machine Learning Tools and Techniques. – 3rd Edition. – Morgan Kaufmann, 2011. – P. 664.
11. URL: http://ru.wikipedia.org/ (википедия).

УДК 631.158:633.1:631.531.02

Р.Р. Субхангулов

К ВОПРОСУ ОБЕСПЕЧЕНИЯ ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ СЕЛЬСКОХОЗЯЙСТВЕННОГО ПРЕДПРИЯТИЯ ПРИ ПРОИЗВОДСТВЕ ПРОДУКЦИИ

В статье рассматривается вопрос обеспечения сохранности зерна на уборке зерновых культур до места хранения, выявлены причины, влияющие на хищение главного сельскохозяйственного продукта. Результаты работы представлены в виде предложений по учету зерна.

***Ключевые слова**: питание, хищение зерна, учет зерна, сопроводительный документ, движение зерна, продовольственная безопасность.*

R.R. Subkhangulov

TO THE ISSUE OF THE AGRICULTURAL ENTERPRISE ECONOMIC SAFETY IN THE PRODUCT MANUFACTURING

The issue of providing the grain safety in grain harvesting to its storage place is considered in the article, the reasons influencing the theft of the main and agricultural product are revealed. The results are presented in the form of proposals of grain accounting.

Key words: nutrition, grain theft, grain accounting, accompanying document, grain movement, food security.

Введение. Производство продуктов питания «является самым первым условием жизни непосредственных производителей и всякого производства вообще» [1], а уровень обеспечения населения продовольствием рассматривается как важнейший фактор и определяющий критерий уровня социальной жизни любой страны.

Основу производства продуктов питания составляет зерновое производство. Производимые из зерна продукты по своим потребительским свойствам и доступности обеспечивают до 35 % калорийности пищевого рациона, в том числе от 40 до 50 % суточной потребности организма человека в белках и углеводах [2].

Зерновое хозяйство Российской Федерации традиционно является стратегической и одновременно многоцелевой, многофункциональной и системообразующей отраслью в экономике страны вообще и агропромышленного комплекса в частности. Уровень его развития характеризует надежность хлебофуражного снабжения, экономическую и социально-политическую стабильность в стране, её продовольственную безопасность.

Легального определения понятия «экономическая безопасность», которое было бы приведено в нормативно-правовых актах РФ, не существует, в том числе нет и понятия «экономическая безопасность» предприятия, а особенно сельскохозяйственного.