

**А. Н. Осокин, А. Н. Мальчуков**

# ТЕОРИЯ ИНФОРМАЦИИ

**УЧЕБНОЕ ПОСОБИЕ ДЛЯ СПО**

*Рекомендовано Учебно-методическим отделом среднего профессионального образования  
в качестве учебного пособия для студентов образовательных учреждений среднего  
профессионального образования*

**Книга доступна в электронной библиотеке [biblio-online.ru](http://biblio-online.ru),  
а также в мобильном приложении «Юрайт.Библиотека»**

**Москва ■ Юрайт ■ 2019**

УДК 621.391(075.32)

ББК 32.811я723

О75

**Авторы:**

**Осокин Александр Николаевич** — кандидат технических наук, доцент кафедры вычислительной техники Института кибернетики Томского политехнического университета;

**Мальчуков Андрей Николаевич** — кандидат технических наук, доцент кафедры вычислительной техники Института кибернетики Томского политехнического университета.

**Рецензенты:**

*Ехлаков Ю. П.* — доктор технических наук, профессор Томского государственного университета систем управления и радиоэлектроники;

*Сарайкин А. В.* — доктор технических наук, профессор заведующий кафедрой автоматизации обработки информации Томского государственного университета систем управления и радиоэлектроники.

**Осокин, А. Н.**

О75 Теория информации : учеб. пособие для СПО / А. Н. Осокин, А. Н. Мальчуков. — М. : Издательство Юрайт, 2019. — 205 с. — (Серия : Профессиональное образование).

ISBN 978-5-534-11417-1

В пособии рассмотрены этапы обращения информации в информационных системах, методы и модели измерения количества информации, датчики, описание сигналов (спектральное и вейвлет-представление сигналов, спектры гармонического сигнала и прямоугольного импульса, теоремы об изменении спектра сигнала при преобразованиях сигнала), переход от аналоговой формы сигнала к цифровой, модуляция и демодуляция сигналов, согласование характеристик сигнала и канала связи, виды физических линий связи, разделение (мультиплексирование) линий связи, обобщенная информационная модель канала по Шеннону, теоретические модели каналов связи и их пропускная способность, теоремы Шеннона о кодировании для канала связи без помех и с помехами, сжатие данных, методы повышения помехоустойчивости, помехоустойчивое кодирование, представление информации. Каждая глава содержит контрольные вопросы.

Соответствует актуальным требованиям Федерального государственного образовательного стандарта среднего профессионального образования и профессиональным требованиям.

*Предназначено для студентов, изучающих основы информатики и вычислительной техники.*

УДК 621.391(075.32)

ББК 32.811я723



Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав. Правовую поддержку издательства обеспечивает юридическая компания «Дельфи».

ISBN 978-5-534-11417-1

© Осокин А. Н., Мальчуков А. Н., 2014

© ФГАОУ ВО НИ ТПУ, 2014

© ООО «Издательство Юрайт», 2019

# ОГЛАВЛЕНИЕ

<b>ВВЕДЕНИЕ .....</b>	<b>7</b>
<b>ТЕМА 1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ ТЕОРИИ ИНФОРМАЦИИ .....</b>	<b>9</b>
1.1. О понятии <i>информация</i> . Сигналы, данные, информация, знания .....	10
1.2. Этапы обращения информации в автоматизированных системах .....	11
1.3. Методы и модели оценки количества информации.....	14
1.3.1. Структурные меры информации .....	14
1.3.2. Статистическая мера информации.....	17
1.3.3. Сравнение статистической меры информации с аддитивной мерой Хартли .....	19
1.3.4. Количество алгоритмической информации .....	19
Контрольные вопросы и задания к теме .....	20
<b>ТЕМА 2. ДАТЧИКИ .....</b>	<b>21</b>
2.1. Общие сведения о датчиках .....	21
2.2. Физические датчики .....	21
2.3. Химические сенсоры.....	26
2.4. Биологические сенсоры .....	28
2.5. RFID .....	29
2.5.1. Состав системы RFID .....	29
2.5.2. Активные и пассивные метки.....	30
2.5.3. Достоинства радиочастотных меток.....	31
2.5.4. Недостатки радиочастотных меток .....	32
2.5.5. Примеры использования RFID .....	34
2.5.6. Стандартизация в области RFID .....	35
2.6. «Умные датчики» .....	35
Контрольные вопросы и задания к теме .....	36
<b>ТЕМА 3. ОПИСАНИЕ СИГНАЛОВ .....</b>	<b>37</b>
3.1. Временная и спектральная формы описания сигналов .....	37
3.2. Спектры некоторых сигналов .....	41
3.3. Некоторые свойства преобразований Фурье (теоремы о спектрах) .....	45
3.3.1. Теорема о спектре сигнала, представленного суммой сигналов .....	46
3.3.2. Теорема о спектре сигнала, сдвинутого во времени.....	46
3.3.3. Теорема о спектре сигнала при изменении масштаба времени сигнала .....	47
3.3.4. Спектр сигнала при его дифференцировании и интегрировании.....	47
3.4. Использование вейвлет-функций для описания сигналов .....	49
Контрольные вопросы и задания к теме .....	50
<b>ТЕМА 4. ДИСКРЕТИЗАЦИЯ И КВАНТОВАНИЕ СИГНАЛОВ.....</b>	<b>52</b>
4.1. Основные понятия и определения .....	52
4.2. Квантование по уровню.....	55

4.3. Дискретизация по времени .....	57
Контрольные вопросы к теме .....	60
<b>ТЕМА 5. МОДУЛЯЦИЯ СИГНАЛОВ.....</b>	<b>62</b>
5.1. Классификация видов модуляции .....	62
5.2. Аналоговая модуляция .....	67
5.2.1. Амплитудная модуляция .....	67
5.2.2. Спектр АМ-колебаний .....	69
5.2.3. Демодуляция АМ-сигналов .....	73
5.2.4. Угловая модуляция.....	75
5.3. Импульсная модуляция .....	76
Контрольные вопросы и задания к теме .....	79
<b>ТЕМА 6. ОБЩИЕ СВЕДЕНИЯ О ПЕРЕДАЧЕ ИНФОРМАЦИИ .....</b>	<b>81</b>
6.1. Основные задачи, решаемые при передаче информации.....	81
6.2. Режимы передачи данных .....	81
6.3. Согласование характеристик сигнала и канала связи .....	84
Контрольные вопросы и задания к теме .....	86
<b>ТЕМА 7. ВИДЫ ФИЗИЧЕСКИХ ЛИНИЙ СВЯЗИ.....</b>	<b>87</b>
7.1. Акустические линии связи .....	87
7.2. Электрические линии связи .....	88
7.3. Радиолинии .....	93
7.4. Виды спутниковых линий связи .....	97
7.5. Глобальные системы ориентации .....	101
7.6. Стандарты беспроводной связи (радиоинтерфейсы).....	107
7.7. Оптические линии связи.....	111
7.7.1. Волоконно-оптические линии связи (ВОЛС).....	111
7.7.2. Оптические линии связи без использования волоконно-оптического кабеля с большой дальностью связи.....	119
7.7.3. Локальные бескабельные оптические линии связи .....	121
7.8. Концепция структурированных кабельных систем.....	122
Контрольные вопросы и задания к теме .....	124
<b>ТЕМА 8. РАЗДЕЛЕНИЕ ЛИНИЙ СВЯЗИ (МУЛЬТИПЛЕКСИРОВАНИЕ).....</b>	<b>126</b>
8.1. Постановка задачи.....	126
8.2. Частотное разделение .....	127
8.3. Временное разделение .....	128
8.4. Кодовое разделение.....	130
8.5. Фазовое разделение.....	133
8.6. Разделение по форме.....	134
8.7. Комбинированные методы разделения .....	135
Контрольные вопросы и задания к теме .....	135

<b>ТЕМА 9. ПЕРЕДАЧА ИНФОРМАЦИИ ПО КАНАЛУ. ТЕОРЕТИЧЕСКИЕ МОДЕЛИ КАНАЛОВ СВЯЗИ. ТЕОРЕМЫ ШЕННОНА О КОДИРОВАНИИ ДЛЯ КАНАЛОВ СВЯЗИ.....</b>	<b>136</b>
9.1. Обобщённая информационная модель канала .....	136
9.2. Пропускная способность канала.....	137
9.3. Дискретный канал без помех .....	138
9.4. Теорема Шеннона о кодировании для дискретного канала без помех .....	139
Контрольные вопросы и задания к теме .....	139
<b>ТЕМА 10. СЖАТИЕ ДАННЫХ.....</b>	<b>140</b>
10.1. Основные понятия.....	140
10.2. Характеристики алгоритмов сжатия данных .....	141
10.3. Алгоритмы сжатия без потерь .....	143
10.4. Статистические алгоритмы сжатия .....	144
10.5. Алгоритм Хаффмана.....	145
10.6. Алгоритм арифметического кодирования .....	153
10.7. Алгоритмы сжатия, использующие исключение повторов .....	153
10.8. Алгоритмы KWE .....	155
10.9. Словарные и словарно-статистические алгоритмы сжатия .....	155
10.10. Алгоритмы сжатия с потерями .....	156
10.10.1. Алгоритмы сжатия растровых статических изображений .....	156
10.10.2. Сжатие изображений с помощью алгоритма JPEG.....	157
Контрольные вопросы и задания к теме .....	165
<b>ТЕМА 11. ПЕРЕДАЧА ИНФОРМАЦИИ ПО КАНАЛУ С ПОМЕХАМИ .....</b>	<b>167</b>
11.1. Дискретный канал с помехами .....	167
11.2. Пропускная способность дискретного канала с помехами.....	167
11.3. Непрерывный канал с помехами .....	168
11.4. Методы повышения достоверности передачи и приема .....	168
Контрольные вопросы и задания к теме .....	170
<b>ТЕМА 12. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ.....</b>	<b>172</b>
12.1. Классификация помехоустойчивых кодов .....	172
12.2. Систематические помехоустойчивые коды.....	175
12.3. Связь корректирующей способности кода с кодовым расстоянием..	176
12.4. Код Хэмминга.....	179
12.5. Общие сведения о циклических кодах.....	185
12.6. Арифметика по модулю два.....	185
12.7. Двоичные циклические коды .....	186
12.8. Кодирование .....	187
12.9. Декодирование.....	188
12.10. Краткая характеристика современных помехоустойчивых кодов...	190
Контрольные вопросы и задания к теме .....	191

<b>ТЕМА 13. ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ.....</b>	<b>193</b>
13.1. Каналы получения информации человеком .....	193
13.2. Визуализаторы .....	193
13.3. Основные цветные системы (модели) и их использование в вычислительной технике .....	196
13.3.1. Цветовые модели .....	196
13.3.2. Система RGB.....	197
13.3.3. Цветовая модель CMYK .....	197
13.4. Способы формирования цветных изображений в визуализаторах, использующих модель RGB .....	199
13.5. Другие технические средства представления информации, используемые в современных информационных системах.....	202
Контрольные вопросы и задания к теме .....	203
<b>СПИСОК ЛИТЕРАТУРЫ.....</b>	<b>205</b>

## ВВЕДЕНИЕ

Деятельность людей связана с обработкой материалов, энергии и информации. Особенностью современного этапа развития общества является переход от индустриального к информационному, постиндустриальному обществу. Процесс, обеспечивающий этот переход, называют информатизацией. Неизбежность информатизации обусловлена резким возрастанием роли и значения информации. Для нормального функционирования организации любого масштаба уже недостаточно только традиционных для индустриального общества ресурсов (материальных, природных, трудовых, финансовых, энергетических), необходимо знать, как наиболее эффективно эти ресурсы использовать, иметь информацию о технологиях. Поэтому существенным ресурсом стала информация. Информационные ресурсы в настоящее время рассматриваются как отдельная экономическая категория, важнейший стратегический ресурс общества.

В общем случае роль может ограничиваться эмоциональным воздействием на человека, однако наиболее часто она используется для выработки управляющих воздействий в автоматических (чисто технических) и автоматизированных (человеко-машинных) системах. В подобных системах принято выделять отдельные этапы обращения информации: сбор информации, передачу информации (в том числе и хранение), обработку информации; отображение информации.

В рамках дисциплины «Теория информации» изучаются математические вопросы сбора информации с помощью датчиков, преобразование сигналов для передачи данных по каналам связи, основы теории представления информации.

В пособии имеется 13 тем.

В теме 1 «Основные понятия и определения теории информации» рассматриваются понятия *сигнал*, *данные*, *информация*, *знания* и соотношение между этими понятиями, измерение информации структурными, статистической и алгоритмической мерами, области применения этих мер.

В теме 2 «Датчики» приводятся общие сведения о датчиках, рассматриваются физические датчики, химические сенсоры, биологические сенсоры, датчики систем радиочастотной идентификации (RFID), кратко излагается информация про «умные» датчики.

В теме 3 «Описание сигналов» рассматриваются спектральная и временная формы описания сигналов, вейвлет-представление сигналов, определение практической ширины спектра прямоугольного импульса,

изменение спектра при различных преобразованиях сигналов (теорема о спектре суммы сигналов; теорема о спектре сигнала, сдвинутого во времени; теорема о спектре сигнала при изменении масштаба времени; изменение спектра при интегрировании и дифференцировании сигнала).

В теме 4 «Дискретизация и квантование» сформулирована цель перехода от аналоговой формы представления сигнала к цифровой, приведены формулы для выбора частоты дискретизации по времени в зависимости от предполагаемого варианта восстановления исходного сигнала, числа двоичных разрядов при квантовании по уровню для представления амплитуды сигнала при различных вариантах задания ошибки.

В теме 5 «Модуляция сигналов» рассмотрены основные виды непрерывной модуляции сигналов, спектр АМ-сигнала, демодуляция АМ-сигнала, манипуляция сигналами, основные виды импульсной модуляции.

В теме 6 «Общие сведения о передаче сигналов» рассмотрены понятия «линия связи», «канал связи», как передаются данные при последовательной и параллельной передаче данных; режимы передачи данных (симплекс, полудуплекс, дуплекс); основные задачи, решаемые при организации передачи данных; согласование характеристик передаваемого сигнала и используемого канала.

В теме 7 «Виды физических линий связи» рассмотрены акустические, электрические, электромагнитные и оптические линии связи, их преимущества и недостатки, концепция структурированных кабельных систем.

В теме 8 «Разделение линий связи» приведены математические основы частотного, временного и кодового разделения, их достоинства и недостатки, приведены структурные схемы соответствующих многоканальных систем, рассмотрено разделение по форме и фазовое разделение.

В теме 9 «Передача информации по каналу» рассмотрена обобщенная информационная модель канала, теоретические модели каналов связи, дискретный канал без помех, пропускная способность дискретного канала без помех (без вывода), теорема Шеннона о кодировании для дискретного канала без помех (без доказательства).

В теме 10 «Сжатие данных» рассмотрены основные алгоритмы сжатия без потерь, алгоритмы сжатия с потерями, идея сжатия видеопоследовательностей.

В теме 11 «Передача информации по каналу с помехами» сформулировано понятие *дискретный канал с помехами*; приведены теорема Шеннона о кодировании для дискретного канала с помехами (без доказательства), выражения для подсчета пропускной способности двоичного симметричного канала с помехами, пропускной способности непре-



рывного канала с помехами; рассмотрены методы повышения достоверности передачи и приема.

В теме 12 «Помехоустойчивое кодирование» приведена классификация помехоустойчивых кодов; сформулированы определения теории помехоустойчивого кодирования; детально (с примерами по кодированию и декодированию) рассмотрены код Хэмминга, исправляющий однократную ошибку, и код Боуза-Чоудхури-Хоквингема; кратко изложены современные помехоустойчивые коды.

В теме 13 «Представление информации» рассмотрены теоретические основы представления информации: каналы получения информации человеком, принцип работы основных типов визуализаторов, основные системы цветообразования, достоинства, недостатки, методы формирования цветных изображений при использовании системы RGB, области применения систем распознавания и синтеза речи.

Содержание пособия соответствует утвержденной программе дисциплины.

# ТЕМА 1

## ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ ТЕОРИИ ИНФОРМАЦИИ

### 1.1. О понятии *информация*. Сигналы, данные, информация, знания

Несмотря на широкое применение термина «информация», его строгого научного определения до настоящего времени не существует, поэтому вместо определения обычно используют понятие об информации. Понятия, в отличие от определений, не даются однозначно, а вводятся с использованием других определений: сигнал, данные.

Мы живем в материальном мире. Всё, что нас окружает и с чем мы сталкиваемся ежедневно, относится либо к физическим телам, либо к физическим полям. Физические объекты находятся в состоянии непрерывного движения и изменения, которое сопровождается обменом энергией.

Все виды энергообмена сопровождаются появлением **сигналов**, т. е. все сигналы имеют в своей основе материальную энергетическую природу. При взаимодействии сигналов с физическими телами в последних возникают определенные изменения свойств. Это явление называется **регистрацией** сигналов. Зарегистрированные сигналы называют **данными**.

Данные несут в себе информацию о событиях, произошедших в материальном мире, поскольку они являются регистрацией сигналов, возникших в результате этих событий. Однако данные не тождественны информации.

Наиболее распространены следующие определения информации.

**«Информация – это данные и методы, их обрабатывающие».** Развитием этого определения является определение, данное в учебнике «Информатика. Базовый курс» под редакцией С.В. Симоновича:

**Информация – это продукт взаимодействия данных и адекватных им методов преобразования данных.**

Информация возникает именно в тот момент, когда данные начинают обрабатываться. Причём в зависимости от конкретного способа обработки из исходных данных может быть получена совершенно разная информация. Одни и те же пять цифр могут быть как размером зарплаты, так и кодом запуска стратегических ракет.

Как отмечалось, в физическом мире информация представляется сигналами, однако ценность её возникает в результате воздействия этих

сигналов на принимающую сторону, где они получают некоторую субъективную оценку. Таким образом, информация (не существующая в реальном мире как материальный объект) всегда создается пользователем, который либо интерпретирует данные, вкладывая в них некоторый смысл, либо оценивает полученное сообщение.

Информация может представляться не данными, а алгоритмами. Например, в стеганографии для передачи информации собственные данные нередко вообще не формируются – достаточно алгоритма, наподобие «взять первую букву текста А, вторую букву текста Б, ...».

Информация несет в себе закодированные знания, т. е. описание того, как решать какую-либо задачу. Существует много определений понятия «знание». Определение понятия **знание**, приведенное в Европейской схеме по управлению знаниями [European Guide to good Practice in Knowledge Management, Part1], сформулировано следующим образом: **«Знание – это комбинация данных и информации, к которым добавлено мнение, мастерство и опыт эксперта, что в результате дает ценный актив, который может быть использован для оказания помощи в принятии решений»** Человек может извлекать из информации знания, если он объединяет её с теми знаниями, которые у него уже есть. Если требующихся знаний для извлечения новых знаний из информации недостаточно, то для **данного** человека информация не будет увеличивать его знания.

## 1.2. Этапы обращения информации в автоматизированных системах

**Последовательность действий**, выполняемых с информацией, называют **информационным процессом**.

Системы, реализующие информационные процессы, называют **информационными системами**. Различают автоматические (чисто технические) и автоматизированные (человеко-машинные) системы.

В подобных системах можно выделить отдельные этапы обращения информации, каждый из которых характеризуется определенными действиями.

Основными этапами обращения информации в системах являются:

- сбор информации;
- подготовка (преобразование) информации;
- передача информации;
- обработка информации;
- хранение информации;
- представление информации.

Так как материальным носителем информации является сигнал, то реально это будут этапы обращения и преобразования сигналов. Этапы обращения информации в автоматизированных системах приведены на рис. 1.1.

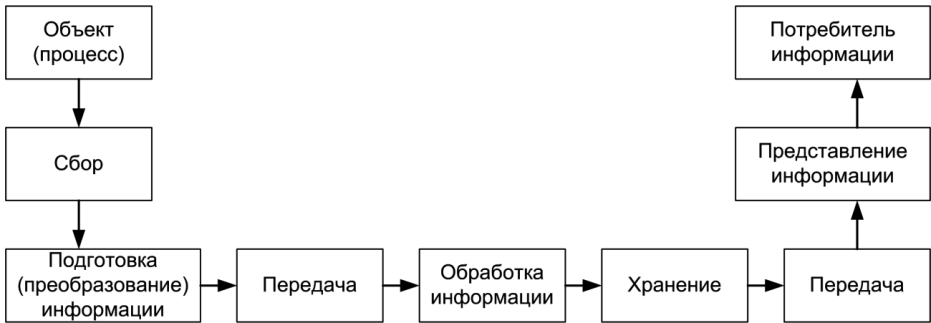


Рис. 1.1. Этапы обращения информации в автоматизированных системах

На *этапе сбора* с помощью датчиков формируются сигналы (чаще всего электрические), отражающие состояние объекта (например: температуру, освещенность, влажность, наличие определенных химических веществ и т. д. в помещении), процесса (температуру, давление).

На *этапе подготовки* осуществляется её первичное преобразование. На этом этапе проводятся такие операции, как нормализация сигналов, снимаемых с датчиков, аналого-цифровое преобразование, модуляция. В результате подготовки формируется сигнал в форме, удобной для передачи, хранения и обработки по конкретной линии связи. Современная промышленность выпускает для систем автоматизации технологических процессов так называемые **устройства согласования с объектом**, которые выполняют перечисленные операции. Иногда этап подготовки включается в этап сбора информации.

На *этапе передачи* информация пересылается от отправителя к получателю. Передача осуществляется по линиям связи различной физической природы (акустическим, электрическим, электромагнитным, оптическим). Основные задачи, решаемые при передаче, следующие: обеспечить эффективное использование канала с помощью эффективного кодирования, помехоустойчивости передаваемых данных, защищенности от несанкционированного доступа к информации.

На *этапе обработки* информации выявляются её общие и существенные взаимозависимости, представляющие интерес для системы. В общем случае под обработкой информации понимается любое её преобразование, проводимое по законам логики, математики, а также неформальным правилам, основанным на «здравом смысле», интуиции, обобщенном опыте. Результатом обработки является тоже информация,

но либо представленная в иных формах (например, упорядоченная по каким-то признакам), либо содержащая ответы на поставленные вопросы (например, решение некоторой задачи). Если процесс обработки формализуем, он может выполняться техническими средствами.

Важной проблемой при обработке информации является обеспечение контроля правильности функционирования устройств обработки информации (УОИ). С точки зрения выявления и исправления ошибок в процессе работы УОИ все устройства и процессы можно разделить на два класса:

- устройства и процессы, в которых информация лишь передается во времени и пространстве, количество нулей и единиц и их положение в двоичном слове не меняется (хранение информации во всех устройствах хранения данных, передача данных по шинам и т. д.);
- устройства, в которых происходит изменение информации, при этом данные на выходе в принципе должны отличаться от данных, поступивших на вход.

Для обеспечения помехоустойчивости работы устройств передачи данных используется 5 методов, которые рассмотрены далее, в теме 11.

Для контроля правильности функционирования устройств, в которых происходит изменение информации, пока реальным и универсальным является лишь дублирование таких устройств и сравнение результатов их работы. Если ставится задача только обнаружения ошибки в работе, то можно задублировать устройства и сравнивать результаты работы двух идентичных устройств. Несовпадение результатов работы рассматривается как ошибка (хотя вероятность того, что ошибка появилась в контролируемом, а не контролирующем устройстве, равна всего 50 %).

Если ставится задача исправления ошибок, то можно использовать трехкратное резервирование с выработкой результата обработки путем голосования с помощью мажоритарных элементов. Эти элементы вырабатывают выходные данные «по большинству». Если из трех устройств одно стало работать неправильно, – это не скажется на результате. Только ошибка в работе двух из трех устройств проявится в результате.

Рассмотренный метод обнаружения, и особенно исправления ошибок, очень дорог. Поэтому для контроля устройств, в которых происходит изменение информации, разработаны более экономичные эвристические схемы контроля отдельных устройств, универсальные методы контроля с использованием систем остаточных классов, контроль с использованием вычетов, контроль с использованием систем счисления с иррациональным основанием (коды Фибоначи, коды «Золотой пропорции», коды Стахова).

На *этапе хранения* информацию записывают в устройство хранения данных. При разработке устройств хранения данных решается задача надежного хранения записанных данных, при этом используются те же методы, что и при передаче информации (ряд авторов называют хранение передачей во времени в отличие от передачи в пространстве при передаче от отправителя к получателю). Решение задач извлечения хранимой информации (поиска информации) связано с разработкой классификационных признаков и схем размещения хранимой информации, систематизацией, правилами доступа к ней, порядком её пополнения и обновления, т. е. всем тем, что определяет возможность целенаправленного поиска и оперативного извлечения хранимой информации.

*Представление* информации требуется тогда, когда в работе системы принимает участие человек. Оно заключается в предоставлении человеку информации с помощью устройств, способных воздействовать на органы чувств человека.

### **1.3. Методы и модели оценки количества информации**

Для теоретической информатики информация играет такую же роль, как и вещество в физике. Для характеристик информации используются единицы измерения, что позволяет некоторой порции информации приписывать числа – количественные характеристики информации.

На сегодняшний день наиболее известны следующие способы измерения количества информации (меры количества информации):

- структурные (объёмные);
- энтропийный;
- алгоритмический.

#### **1.3.1. Структурные меры информации**

Ряд авторов называют их объёмными. При их использовании учитывается только количество информационных элементов, содержащихся в информации, связей между ними или комбинаций из них. Под информационными элементами понимаются неделимые части информации (кванты) в дискретных моделях реальных информационных комплексов, а также элементы алфавитов в числовых системах. В структурной мере различаются геометрическая, комбинаторная и аддитивная меры информации.

Определение **количества информации геометрической мерой** сводится к измерению площади или объёма геометрической модели данного информационного комплекса в количестве дискретных единиц. Например, при рассмотрении характеристик визуализаторов указывают

количество пикселей, которое может формироваться в строке (разрешение по горизонтали), количество строк, которое может формироваться по вертикали (разрешение по вертикали), глубину цвета пикселя (количество цветовых оттенков, которое может принимать пиксель). Геометрической мерой определяется максимально возможное количество информации, которое можно представить данной структурой.

**Количество информации в комбинаторной мере** вычисляется как количество комбинаций элементов информационных комплексов. Для подсчета можно использовать формулы комбинаторики (число сочетаний, перестановок, размещений).

Формулы комбинаторики неудобны из-за больших получаемых величин. В 1928 г. Хартли ввел двоичную логарифмическую меру, позволяющую вычислять количество информации в двоичных единицах – битах [бит].

**Количество информации по Хартли** определяется по формуле

$$I_x = l \cdot \log_2 h \text{ [бит]},$$

где  $h$  – основание системы счисления (количество состояний, которое может принимать элемент, хранящий данное число);

$l$  – число элементов.

Бит по Хартли: количество разрядов равно 1, основание – 2,  $\log_2 2 = 1$  бит.

Бит является минимальной единицей количества данных, которое может хранить элемент с двумя устойчивыми состояниями или передаваться по линии связи (есть импульс – единица, нет импульса – ноль). Были предложены и используются более крупные единицы количества данных.

**Байт** (англ.) – единица измерения количества информации, **обычно** равная восьми битам. К сожалению, длина байта строго не фиксирована. Она может быть: 7 бит + 1 контрольный бит;

8 бит без контрольного бита;

8 бит + 1 контрольный бит.

В ряде современных цифровых сигнальных процессоров используется байт длиной 16 бит и больше.

Для того чтобы подчеркнуть, что имеется в виду восьмибитный байт, в описании сетевых протоколов используется термин **октет** (octet) – совокупность восьми соседних бит, не связанных между собой. **Октет – это цепочка из восьми бит в длинной цепочке бит.**

К сожалению, общепринятых сокращенных названий бита и байта нет. Буквой В может обозначаться и бит и байт, что далеко не одно и то же. Например, в характеристиках линии связи указано, что скорость передачи данных равна 1000 В/с. Реально это будет обозначать либо

1000 бит/с, либо 1000 байт/с = 8000 бит/с. Автор настаивает: требуется указывать **полное название** единицы измерения количества информации.

Для байта кратные приставки для образования производных единиц применяются не как обычно.

Во-первых, уменьшительные приставки не используются совсем. Единица измерения информации, равная половине байта, называется специальным словом **ниббл**.

Во-вторых, существует два вида увеличительных приставок: десятичные (согласно международной системе СИ) и двоичные. Значения приставок приведены в табл. 1.1.

Таблица 1.1

Десятичная приставка (СИ)				Двоичная приставка		
Множитель		Обозначение		Множитель		
значение (вес)	название	англо-язычное	русско-язычное	значение (вес)	обозначение	название
$10^3$	кило	<i>k</i>	к	$2^{10}$	<i>KiB</i>	киби
$10^6$	мега	<i>M</i>	М	$2^{20}$	<i>MiB</i>	меби
$10^9$	гига	<i>G</i>	Г	$2^{30}$	<i>GiB</i>	гиби
$10^{12}$	тера	<i>T</i>	Т	$2^{40}$	<i>TiB</i>	теби
$10^{15}$	пета	<i>P</i>	П	$2^{50}$	<i>PiB</i>	пеби
$10^{18}$	экса	<i>E</i>	Е	$2^{60}$	<i>EiB</i>	экзби
$10^{21}$	зетта	<i>Z</i>	З	$2^{70}$	<i>ZiB</i>	зеби
$10^{24}$	йотта	<i>Y</i>	Й	$2^{80}$	<i>YiB</i>	йоби

Для указания емкости оперативной памяти компьютеров используют обозначения десятичных приставок, а веса – двоичных приставок. При указании емкости накопителей на жестких магнитных дисках (НЖМД), накопителей на магнитной ленте (для создания хранилищ данных они до сих пор успешно используются) применяются веса и названия десятичных приставок.

Законность использования десятичных укрупняющих приставок для указания ёмкости НЖМД отстояли в американском суде производители НЖМД. Операционная система компьютера измеряет ёмкость с использованием двоичных укрупняющих приставок, поэтому она укажет ёмкость меньшую. Например:  $1G = 1000000000$ ,  $1Gi = 1073741824$ .

При больших укрупняющих приставках разница ещё более значительна.

При передаче по линии связи, как правило, используют десятичные укрупняющие приставки.

В настоящее время уже появляются хранилища данных объемом в экзбайты.



### 1.3.2. Статистическая мера информации

В данной мере информация рассматривается как сообщение об исходе случайных событий, реализации случайных величин, функций, а количество информации ставится в зависимость от априорных (доопытных) вероятностей этих событий, величин, функций. Когда приходит сообщение о часто встречающихся событиях, вероятность появления которых стремится к единице, то такое сообщение малоинформативно. Если же приходит сообщение о наступлении события, вероятность которого мала, то сообщение информативно. Главным свойством случайных событий является отсутствие полной уверенности в их наступлении, создающее известную неопределенность. Для практики важно уметь численно оценивать степень неопределенности самых разнообразных событий, чтобы иметь возможность сравнивать их с этой стороны.

Клодом Элвудом Шенноном было предложено принять в качестве меры неопределенности события (опыта) с возможными исходами  $x_1, x_2, \dots, x_n$  величину  $H$ , называемую энтропией:

$$H = -p(x_1) \cdot \log p(x_1) - p(x_2) \cdot \log p(x_2) - \dots \\ \dots - p(x_n) \cdot \log p(x_n) = -\sum_{i=1}^n p(x_i) \cdot \log p(x_i),$$

где  $p(x_1), p(x_2), \dots, p(x_n)$  – вероятность отдельных исходов.

Поскольку строгое определение вероятности будет даваться в курсах, изучаемых студентами данной специальности позднее, введём понятие вероятности.

Рассмотрим следующий опыт. Пусть мы бросаем монету. При проведении бросаний какое-то количество раз она будет падать гербом вверх, какое-то количество раз – цифрой вверх. Если мы бросим монету 10 раз, то в 3 случаях может выпасть герб, а в 7 – цифра. При бросании монеты 100 раз, герб может выпасть 40 раз, цифра – 60 раз.

Если бросать монету 1000 раз, то, скорее всего, герб выпадет 499 раз, а цифра 501 раз.

Вероятность наступления некоторого события равна отношению числа равновероятных исходов, благоприятных для данного события, к общему числу равновероятных событий, т. е.

$$p_i = \lim_{N \rightarrow \infty} \frac{n_i}{N},$$

где  $n_i$  – число положительных исходов опыта (например, выпадений герба);

$N$  – общее число испытаний.

Заметим, что выбор основания системы логарифмов несущественен, т. к. (в силу известной формулы  $\log_b k = \log_b a \cdot \log_a k$ ) переход от одной системы логарифмов к другой сводится лишь к умножению на модуль перехода  $\log_b a$ .

Рассмотрим некоторые свойства энтропии, используемые в данном курсе:

1. Энтропия всегда неотрицательна, т. к. значения вероятностей выражаются дробными величинами, а логарифмы величин, меньших единицы, – отрицательными числами, так что под знаком суммы получается отрицательная величина, а перед знаком суммы стоит знак минус, а минус умножить на минус дает плюс.

2. Энтропия равна нулю в том и только в том случае, когда вероятность одного события равна единице, а остальных – нулю. Это соответствует случаю, когда исход опыта может быть предсказан с полной достоверностью.

3. Энтропия имеет наибольшее значение при условии, когда все вероятности равны между собой:  $p(x_1) = p(x_2) = \dots = p(x_k) = \frac{1}{k}$ .

$$\text{В этом случае } H = -\sum_{i=1}^n \frac{1}{k} \cdot \log \frac{1}{k} = -\log \frac{1}{k} = \log k .$$

Итак, **максимальная энтропия** равна  $H_{\max} = \log k$ .

Количество информации  $I$  в статистической мере о некотором объекте равно разности априорной  $H_1$  и апостериорной  $H_2$  энтропий объекта:

$$I = H_1 - H_2 ,$$

где  $H_1$  – априорная (до проведения опыта) энтропия;

$H_2$  – апостериорная (после проведения опыта) энтропия.

Если неопределённость ситуации снимается полностью, то  $H_2 = 0$

и  $I = H = -\sum_{i=1}^n p(x_i) \cdot \log p(x_i)$ .

Данная формула и является выражением для подсчёта количества информации статистической мерой. Основание логарифма определяет единицу измерения количества информации. При двоичном логарифме единица измерения – бит (binary digital) – по названию совпадает с битом по Хартли. Статистическую меру информации имеет смысл использовать лишь при теоретической оценке пропускной способности линий связи. Для указания объёма памяти используется бит, определённый Хартли.

### 1.3.3. Сравнение статистической меры информации с аддитивной мерой Хартли

Статистическая мера информации совпадает с аддитивной мерой Хартли при условии, что  $l = 1$ , а все элементы алфавита равновероятны, тогда

$$I_x = l \cdot \log_2 h;$$

$$I_{ш} = \log_2 k,$$

где  $I_x$  – количество информации по Хартли;

$I_{ш}$  – количество информации по Шеннону.

Совпадение количества информации по Шеннону и Хартли говорит о полном использовании потенциальной информационно ёмкости.

В случае неравновероятных событий количество информации по Шеннону меньше количества информации по Хартли.

### 1.3.4. Количество алгоритмической информации

Современная теория алгоритмической информации позволила дать точные количественные определения понятиям сложности и простоты.

Объём алгоритмической информации определяется длиной компьютерной программы, необходимой для генерации данных. **Минимальное число битов, необходимых для хранения программы, называется количеством алгоритмической информации.**

Например, бесконечный ряд натуральных чисел 1, 2, 3, ... содержит очень мало алгоритмической информации: все числа ряда можно получить с помощью коротенькой компьютерной программы. Не имеет значения, сколько времени понадобится для выполнения вычислений и какой объём памяти придется использовать, важна лишь длина программы в битах. Разумеется, точное значение количества алгоритмической информации зависит от используемого языка программирования, но об этом позже.

В качестве другого примера возьмем число  $\pi$ , равное 3,14159 ... . Количество алгоритмической информации в нём тоже невелико: для последовательного вычисления всех его знаков можно использовать довольно короткий алгоритм, например, вычисляющий  $\pi$  по формуле Гаусса, формуле Чудновского (Chudnovsky).

В противоположность этому случайное число, содержащее всего миллион знаков, например 1,341285...64, характеризуется гораздо бóльшим количеством алгоритмической информации. Поскольку такое число не подчиняется закону, длина самой короткой программы, необходимой для его написания, будет близка к длине самого числа:

Начать

Напечатать <<1,341285...64>>

Конец

В программу должны быть включены все цифры, замененные многоточием. Никакая более короткая программа не позволит рассчитать подобную последовательность цифр: её невозможно сжать, в ней нет избыточности. Самое лучшее, что можно сделать, – просто передать как она есть. Такие последовательности называются неприводимыми, или алгоритмически случайными.

Так как имеется много разных вычислительных машин и разных языков программирования, то для определенности задаются некоторой конкретной вычислительной машиной, например машиной Тьюринга.

**В этом случае количество алгоритмической информации в сообщении определяется как минимальное число внутренних состояний машины Тьюринга, требующееся для его воспроизведения.**

### **Контрольные вопросы задания к теме**

1. Дайте определение термину «информация».
2. Чем отличаются данные от информации?
3. Какие этапы проходит информация в автоматизированной системе управления? Кратко опишите каждый этап.
4. С помощью чего передаются данные?
5. Перечислите структурные меры количества информации. Приведите формулы подсчета количества информации различными структурными мерами.
6. Приведите формулу подсчета количества информации по Шеннону.
7. Сравните бит по Хартли и бит по Шеннону.
8. Какому количеству бит равен ниббл?
9. Чем отличаются двоичные укрупняющие приставки в единицах измерения данных от принятых в международной системе единиц?
10. Вы купили накопитель на жестких магнитных дисках, в характеристиках которого указана емкость 200 Гбайт. При подключении к компьютеру операционная система указала емкость 186,2645 Гбайт. В чем причина расхождения значений емкости?
11. В коммуникационных системах вычислительной техники используется единица «октет». Дайте пояснение к этому термину.