

ТЕМА 1. ИНФОРМАТИКА И ИНФОРМАЦИЯ

Содержание

1. Понятие информатики и ее составляющих
2. Понятие данных и информации
3. Общая характеристика процессов сбора, передачи, обработки и накопления информации
4. Виды и типы данных
5. Основные критерии качества информации
6. Основные структуры данных
7. Единицы представления, измерения, хранения и передачи данных
8. Контрольные вопросы

1. Понятие информатики и ее составляющих

Информатика – это техническая прикладная наука, систематизирующая приемы создания, хранения, воспроизведения, обработки и передачи данных (информации) средствами вычислительной техники, а также принципы функционирования этих средств и методы управления ими.

Основной задачей информатики является систематизация приемов и методов работы с аппаратными и программными средствами вычислительной техники.

Основопологающим понятием информатики является информационная модель.

Основной метод, используемый в информатике, это ***моделирование информационных процессов с помощью компьютера.***

Предмет «Информатика» – это учебная дисциплина, в рамках которой изучаются технологии создания, хранения, воспроизведения и обработки данных (информации) средствами вычислительной техники, а также принципы функционирования этих средств и методы управления ими.

2. Понятие данных и информации

Все существующие объекты, процессы или явления обладают определенными свойствами, которые можно регистрировать теми или иными средствами и способами. Зарегистрированные свойства характеризуют объекты, процессы или явления. **Следовательно, зарегистрированные свойства – это данные об объектах, процессах или явлениях.**

Обработка данных адекватными им методами создает новый продукт – информацию. Таким образом, информация возникает и существует в момент взаимодействия объективных данных и субъективных методов.

Итак, *информация* – это продукт взаимодействия данных и адекватных им методов или мера устранения неопределенности в отношении, интересующего нас события.

3. Общая характеристика процессов сбора, передачи, обработки и накопления информации

В структуре возможных операций с данными и информацией можно выделить следующие операции:

- ✓ **сбор данных и информации** – накопление информации с целью обеспечения достаточной полноты для принятия решения;
- ✓ **форматизация данных (информации)** – это приведение данных, поступающих из разных источников, к одинаковой форме, чтобы сделать их сопоставимыми между собой, то есть повысить их уровень доступности;
- ✓ **фильтрация данных и информации** – отсеивание тех данных, в которых нет необходимости для принятия решения; при этом должны возрастать достоверность и адекватность информации;

- ✓ **сортировка данных и информации** – упорядочение данных по определенному признаку с целью удобства их использования; при этом должна повышаться доступность информации;
- ✓ **архивация данных и информации** – организация хранения данных в удобной и легкодоступной форме; служит для снижения экономических затрат по хранению данных и повышает общую надежность информационного процесса в целом;
- ✓ **преобразование данных и информации** – перевод данных из одной формы в другую или из одной структуры в другую.

4. Виды и типы данных

Данные могут быть представлены следующими видами:

- **числами;**
- **текстом;**
- **мультимедиа (графическими объектами, звуковыми сигналами, цветными изображениями).**

Виды данных подразделяются на следующие типы:

а) числовой вид:

- **байтовый тип (byte, диапазон: 0 ... 256);**
- **целые числа простой точности(integer, диапазон: -32768 ... 32767);**
- **длинные целые числа(long);**
- **вещественные числа простой точности(single);**
- **вещественные числа двойной точности(double);**

б) текстовый вид:

- **строковый тип(string);**
- **текстовый тип (в базах данных);**

в) мультимедийный вид самостоятельных типов не имеет, однако для представления графических объектов может использоваться тип объектов, а свойства этого вида задаются целыми числами.

Дата и время могут представляться в виде чисел, однако к числовому виду они не относятся. Для них применяется тип даты и времени.

Наряду с названными типами, относящимися к тем или иным видам, существуют также:

- **логический тип (Boolean);**
- **тип объектов.**

Кроме того, в некоторых языках программирования предусмотрена возможность создания пользовательского типа, а также имеется тип *Variant* (например, в объектно-ориентированном языке Visual Basic for Applications), совместимый со всеми другими типами.

5. Основные критерии качества информации

С технологической точки зрения информация является продукцией информационных систем. Как и для всякого продукта, для информации большое значение имеет ее качество, то есть способность удовлетворять определенные информационные потребности.

Качество информации является сложным понятием, его основу составляет базовая система показателей, включающая показатели трех классов:

а) класс выдачи (своевременность, актуальность, полнота, доступность и другие):

- ***Своевременность*** информации оценивается временем выдачи (получения), в течение которого информация не потеряла свою актуальность.
- ***Актуальность*** информации – это степень ее соответствия текущему моменту времени.

Устаревшая и потерявшая свою актуальность информация может приводить к ошибочным решениям и тем самым теряет свою практическую ценность.

- ***Полнота*** информации определяет достаточность данных для принятия решений или для создания новых данных на основе имеющихся. Чем полнее данные, тем проще подобрать метод, вносящий минимум погрешностей в ход информационного процесса.

в) класс обработки (достоверность, адекватность):

- ***Достоверность* информации – это степень соответствия между получаемой и исходящей информацией.**
- ***Адекватность* информации – это степень соответствия реальному объективному состоянию дела. Неадекватная информация может образовываться при создании новой информации на основе неполных или недостаточных данных. Однако и полные, и достоверные данные могут приводить к созданию неадекватной информации в случае применения к ним неадекватных методов.**

с) класс защищенности (доступность, физическая целостность, логическая целостность, безопасность):

- ***Доступность* информации – мера возможности получить ту или иную информацию. Отсутствие доступа к данным или отсутствие адекватных методов обработки данных приводят к одинаковому результату: информация оказывается недоступной**
- ***Безопасность информации* – один из наиболее существенных показателей качества информации.**

6. Основные структуры данных

Работа с большими наборами данных автоматизируется проще, когда данные упорядочены, то есть образуют заданную структуру. При работе с данными чаще всего используются *линейная, табличная, иерархическая* структуры. При создании любой структуры данных необходимо обеспечить решение двух задач: разделение элементов данных между собой и поиск нужных элементов.

Линейные структуры – это хорошо знакомые списки. Список – это простейшая структура данных, отличающаяся тем, что каждый элемент данных однозначно определяется своим уникальным номером в массиве (списке).

Табличные структуры данных подразделяются на двумерные и многомерные.

Двумерные табличные структуры данных (матрицы) – это упорядоченные структуры, в которых адрес элемента определяется номером столбца и номером строки, на пересечении которых находится ячейка, содержащая искомый элемент.

Многомерные таблицы – это упорядоченные структуры данных, в которых адрес элемента определяется тремя и более измерениями. Для отыскания нужного элемента в таких таблицах необходимо знать параметры всех измерений (размерностей).

Иерархические структуры – это структуры, объединяющие нерегулярные данные, которые трудно представить в виде списка или таблицы. В иерархической структуре адрес каждого элемента определяется маршрутом, ведущим от вершины структуры к данному элементу. Эти структуры по форме сложнее, чем линейные и табличные, но они не создают проблем с обновлением данных. Их легко развивать путем создания новых уровней. Недостатками иерархических структур являются относительная трудоемкость записи адреса элемента данных и сложность упорядочения. Поэтому для упорядочения в таких структурах применяется метод предварительной индексации. При этом каждому элементу данных присваивается свой уникальный индекс, который используется при поиске, сортировке и других действиях по обработке информации. В качестве примера иерархической структуры может служить система почтовых адресов.

7. Единицы представления, измерения, хранения и передачи данных

Одной из систем представления данных, принятых в информатике и вычислительной технике является система двоичного кодирования. Наименьшей единицей такого представления является *бит* (двоичный разряд).

Совокупность двоичных разрядов, выражающих числовые или иные данные, образует некий битовый рисунок. С битовым представлением удобнее работать, если этот рисунок имеет регулярную форму. В качестве таких форм используются группы из **8 битов**, каждая из которых называется *байтом*. Однако во многих случаях целесообразно использовать 16-разрядное, 24-разрядное, 32-разрядное, 64-разрядное кодирование. Байт является наименьшей единицей измерения количества данных (информации).

Более крупные единицы измерения данных образуются добавлением префиксов: кило-, мега-, гига-, тера-, пета-, экса-, зетта-, йотта-.

1 килобайт (Кбайт) = 1024 байт = 2^{10} байт.

1 мегабайт (Мбайт) = 1024 Кбайт = 2^{20} байт.

1 гигабайт (Гбайт) = 1024 Мбайт = 2^{30} байт.

1 терабайт (Тбайт) = 1024 Гбайт = 2^{40} байт.

1 петабайт (Пбайт) = 1024 Тбайт = 2^{50} байт.

1 эксабайт (Эбайт) = 1024 Пбайт = 2^{60} байт.

1 зеттабайт (Збайт) = 1024 Эбайт = 2^{70} байт.

1 йоттабайт (Йбайт) = 1024 Збайт = 2^{80} байт.

В качестве единицы хранения данных (информации) принят объект переменной величины, называемый файлом.

Файл – это последовательность байтов произвольного числа, обладающая уникальным собственным именем.

Поскольку в определении файла нет ограничений на его размер, то можно представить себе файл, имеющий 0 байтов (пустой файл), и файл, имеющий любое число байтов.

В определении файла особое внимание уделяется имени. Имя файла фактически несет в себе адресную информацию, без которой данные, хранящиеся в файле, не станут информацией из-за отсутствия методов доступа к ним. Кроме адресной информации в имени файла могут содержаться сведения о типе данных, заключенных в нем.

Требование уникальности имени файла в вычислительной технике обеспечивается автоматически – создать файл с именем, тождественным уже существующему, не может ни пользователь, ни автоматика.

Хранение файлов организуется в иерархической структуре, которая называется файловой структурой. В качестве вершины структуры служит имя носителя, на котором сохраняются файлы. Далее файлы группируются в каталоги (папки), внутри которых могут быть созданы вложенные каталоги (папки). *Путь доступа* к файлу начинается с имени устройства и включает все имена каталогов (папок), через которые проходит. В качестве разделителя используется символ «\» (обратная косая черта).

Все эти сведения образуют *спецификацию файла*, которая записывается следующим образом:

Имя носителя: \ Имя каталога 1 \ ... \ Имя каталога N \ Составное имя файла.

Пример: *C:\Игры\Стрелялки\Кролики.exe.*

Передача данных в компьютерных системах происходит с разной скоростью. Единицей измерения скорости передачи данных через последовательные порты является бит в секунду (бит/с, Кбит/с, Мбит/с). Единицей измерения скорости передачи данных через параллельные порты является байт в секунду (байт/с, Кбайт/с, Мбайт/с).

8. Кодирование данных двоичным кодом

Для автоматизации работы с данными, относящимися к различным типам, очень важно унифицировать их форму представления. Для этого используется прием кодирования, то есть выражение данных одного типа через данные другого типа. В вычислительной технике применяется система кодирования двоичным кодом. Она основана на представлении данных последовательностью всего двух знаков **0** и **1**. Эти знаки называются двоичными цифрами, каждая из которых представляет 1 бит. Одним битом могут быть выражены два понятия: 0 или 1 (да или нет, черное или белое, истина или ложь и т. п.). Двумя битами можно выразить четыре различных понятия: 00, 01, 10, 11. Тремя битами можно закодировать 8 различных значений: 000, 001, 010, 011, 100, 101, 110, 111. При увеличении количества разрядов в системе двоичного кодирования на единицу количество кодируемых значений увеличивается в два раза.

Кроме системы двоичного кодирования в программировании используются системы восьмеричного и шестнадцатеричного кодирования, а вообще-то в качестве основания системы счисления, используемой для кодирования, может выступать любое число.

Институт стандартизации США (**ANSI** – American National Standard Institute) ввел в действие систему кодирования **ASCII** (American Standard Code for Information Interchange – стандартный код информационного обмена США). В системе ASCII закреплены две таблицы кодирования – базовая и расширенная.

Базовая таблица содержит значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 по 255.

Первые 32 кода (от 0 до 31) базовой таблицы выделены производителям аппаратных средств (в первую очередь компьютеров и печатающих устройств). Это управляющие коды, которым не соответствуют никакие символы, зато с их помощью можно управлять работой технических устройств.

Коды от 32 по 127 предназначены для кодирования символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов. Символы русского алфавита и другие специальные символы кодируются кодами расширенной таблицы от 128 по 255.

Однако рассмотренная выше система кодирования ASCII не обеспечивает кодирование алфавитов многих других языков планеты. С целью устранения этого недостатка в настоящее время используется универсальная система – **UNICODE**, основанная на 16-разрядном (2 байта) кодировании символов. Эта система позволяет обеспечить уникальные коды для 65536 различных символов. Этого количества достаточно для размещения в одной таблице символов практически всех различных алфавитов планеты.

Кодирование целых и вещественных чисел

Для кодирования целых чисел от 0 до 255 достаточно иметь 8 разрядов двоичного кода (8 бит). Для кодирования чисел от 0 до 65535 потребуется 16 разрядов (16 бит). Используя 24 разряда (24 бита), можно закодировать более 16,5 миллиона разных значений.

Для кодирования вещественных чисел используется 80 разрядов (80 бит). При этом действительное число предварительно преобразуется в нормализованную форму:

$$41,2346785 = 0,412346785 \times 10^2.$$

Первая часть нормализованного числа называется мантиссой, а вторая – характеристикой. При этом значительная часть из 80 бит задействуется для хранения мантиссы (вместе со знаком числа) и некоторое фиксированное количество бит отводится для хранения характеристики (тоже со знаком степени).

При восьмеричном кодировании десятичное число последовательно делится на 8, а при шестнадцатеричном кодировании десятичное число последовательно делится на 16. Формирование восьмеричного кода осуществляется аналогично двоичному кодированию.

Кодирование десятичных чисел двоичным, восьмеричным и шестнадцатеричным кодами, а также обратное преобразование достаточно легко можно выполнить с помощью калькулятора, имеющегося в составе пакета программ «Стандартные». При этом необходимо использовать «Инженерный» режим работы калькулятора.

Кодирование графических данных

Если графическое изображение рассматривать как комбинацию мельчайших точек (пикселей), образующих определенный узор, называемый растром, то для его кодирования, используя систему линейных координат и индивидуальные свойства каждой точки, записываемые с помощью целых чисел, можно применить двоичную систему. К индивидуальным свойствам точки относятся яркость и цвет. Черно-белые иллюстрации представляются в виде комбинации точек с 256 градациями серого цвета. Таким образом, для кодирования яркости любой точки достаточно 8 разрядов двоичного числа.

Кодирование цветных графических изображений осуществляется с помощью декомпозиции произвольного цвета на основные составляющие. В качестве таких составляющих используются три цвета: красный (**Red, R**), зеленый (**Green, G**) и синий (**Blue, B**). Такой принцип кодирования называется системой **RGB**. При этом если для кодирования яркости каждой из основных составляющих использовать по 256 значений (8 двоичных разрядов), то на кодирование цвета одной точки требуется 24 разряда. Такая система кодирования обеспечивает 16,5 миллионов цветов. Эта система является полноцветной и называется **True Color**.

Если уменьшить количество двоичных разрядов, используемых для кодирования цвета каждой точки, то можно сократить объем данных, но при этом заметно сократится диапазон кодируемых цветов. Кодирование цветной графики двоичными числами, содержащими 16 разрядов, называется **High Color**.

На практике применяется индексный метод кодирования информации о цвете. При этом код каждой точки растра выражает не цвет сам по себе, а только его номер (индекс) в справочной таблице, называемой *палитрой*, которая прилагается к графическим данным.

Кодирование звука

Для кодирования звуковой информации применяется метод таблично-волнового синтеза (**Wave-Table**). Сущность этого метода состоит в том, что используются заранее подготовленные таблицы образцов звуков. В технике такие образцы называют *сэмплами*. Числовые коды звуковой информации выражают тип инструмента и номер его модели, высоту тона, продолжительность, интенсивность звука и динамику его изменения, а также некоторые параметры среды, в которой происходит звучание, и прочие параметры, характеризующие особенности звука.

9. Контрольные вопросы

- 1) Назовите одно из основных направлений развития информатики.
- 2) Назовите основополагающее понятие информатики.
- 3) Назовите основной метод, используемый в информатике.
- 4) Как называется соответствие информации текущему моменту времени?
- 5) Определите понятие данных.
- 6) Определите понятие информации.
- 7) Как называются сведения об объектах, процессах и явлениях, повышающие степень достоверности?
- 8) Сколько символов позволяет закодировать таблица Unicode?
- 9) Каким понятием определяется информация, пригодная для переработки автоматизированными или автоматическими средствами?
- 10) Перечислите виды данных.
- 11) Назовите числовые типы данных.

- 12) Что относится к мультимедийным видам данных?
- 13) В каком виде данные входят в состав команд компьютера?
- 14) Какого вида известна информация, если задан целочисленный тип данных?
- 15) Перечислите типовые структуры данных.
- 16) Назовите простые структуры данных.
- 17) Какие наименьшие единицы представления, измерения и хранения данных вы знаете?
- 18) Как называются наименьшие единицы передачи данных?
- 19) Перечислите основные цвета модели RGB.
- 20) Какие коды используются для кодирования букв русского алфавита в ASCII?
- 21) Какие коды используются для кодирования букв английского алфавита, арабских цифр и специальных символов в ASCII?
- 22) Какие коды используются для кодирования команд управления