Лекции: «Информационные технологии»

Основные понятия и определения

Информатика – наука, изучающая все аспекты получения, хранения, преобразования, передачи и использования информации.

Информатика состоит из следующих научных направлений:

- Теоретическая информатика;
- Кибернетика;
- Программирование;
- Искусственный интеллект;
- Информационные системы.

Теоретическая информатика – математическая дисциплина. Она использует методы математики для построения и изучения модели обработки, передачи и использования информации, создает теоретический фундамент для всех разделов информатики.

Теоретическая информатика изучает следующие дисциплины:

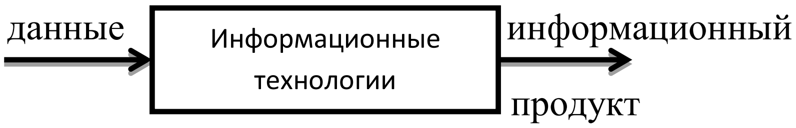
- Математическая логика;
- Вычислительная математика и геометрия;
- Теория информации и кодирования;
- Системный анализ;
- Теория принятия решений.

Из определения информатика можно увидеть, что центральное место информатике занимает понятие информации. Информация является абстрактной категорией и связана с процессом познания человеком окружающего мира. Поэтому понятие «информация» используется в различных смыслах в зависимости от конкретной области приложения.

Информация – это совокупность каких-либо сведений, данных, передаваемых устно (в форме речи), письменно (в виде текста, таблиц) либо другим способом (например, с помощью звуковых или световых сигналов, электрических и нервных импульсов).

В информатике информацию понимают как абстрактное значение выражений, графических изображений, указаний и высказываний.

Информация является одним из важнейших ресурсов общества, наряду с такими материальными видами ресурсов, как нефть, газ, а соответственно процесс ее переработки по аналогии с процессами переработки материальных ресурсов можно воспринимать как технологию.



<u>Информационная технология</u> (ИТ) – процесс, использующий совокупность средств и методов сбора, обработки и передачи данных (первичной информации) для получения информации нового качества о состоянии объекта, процесса или явления (информационного продукта).

<u>Цель</u> информационной технологии – производство информации для ее анализа человеком и принятия на его основе решения по выполнению какого-либо действия.

В связи с применением ИТ, основанной на использовании средств связи, компьютеров, широко используется понятие информационная система.

<u>Информационная система</u> (ИС) – это взаимосвязанная совокупность средств, методов и персонала, используемых для хранения, обработки и выдачи информации в интересах достижения постоянной цели.

Структуру информационной системы составляет совокупность отдельных ее частей, называемых <u>подсистемами</u>.

Виды подсистем:

- Подсистема технического обеспечения;
- Подсистема математического обеспечения;
- Подсистема программного обеспечения;
- Подсистема информационного обеспечения и т.д.

Информационные системы тесно связаны с информационными технологиями, но это разные понятия. Так как ИТ является основной составляющей частью любой ИС, то ИТ является более емким понятием, отражающим современное представление о процессах преобразования информации в информационном обществе.

В первом семестре мы познакомимся с разделом теоретической информатики теорией информации и кодирования.

Во втором семестре мы познакомимся с информационными технологиями и научимся применять их на практике.

1. Единицы информации

Информация, вводимая в компьютер должна быть конкретной и однозначной. Издавна люди пользовались шифрами. Самыми простыми и удобными из них были цифровые шифры. Для обработки компьютером любая информация кодируется с помощью цифр. Цифры представляются электрическими сигналами, с которыми работает компьютер. Для удобства различения в компьютере используют сигналы двух уровней. Один из них соответствует цифре 1, другой - 0**.** Цифры 1 и 0 называются <u>двоичными</u>. Они являются символами, из которых состоит язык, понимаемый и используемый компьютером. Т.о., любая информация в компьютере представляется с помощью двоичных цифр.

Наименьшей единицей информации является <u>бит</u> (bit).

<u>Бит</u> - это количество информации, необходимое для однозначного определения одного из двух равновероятных событий.

Один бит информации получает человек, когда он узнает, опаздывает с прибытием нужный ему поезд или нет, был ночью мороз или нет, присутствует на лекции студент Иванов или нет и т.д.

В информатике принято рассматривать последовательности длиной 8 бит. Такая последовательность называется <u>байтом</u>. С помощью одного байта можно записать двоичные коды 256 (2^8) чисел от 0 до 255.

Единицы измерения информации:

1 байт=8 бит
1 килобайт (Кб) = 1024=2^{10} байт
1 мегабайт (Мб) = 1024 килобайт
1 гигабайт (Гб) = 1024 мегабайт
1 терабайт (Тб) = 1024 гигабайт

2. Системы счисления

Все возможности вычислительной техники (ВТ) реализуются путем создания разнообразных комбинаций сигналов высокого и низкого уровней, которые условились называть «единицами» и «нулями».

Система счисления (СС) - это система записи чисел с помощью определенного набора цифр.

СС называется *позиционной*, если одна и та же цифра имеет различное значение, которое определяется ее местом в числе, иначе непозиционная.

Пример:

- Десятичная СС является позиционной: 9 – стоит на 9-ой позиции.
- Римская СС является непозиционной: значение цифры X в числе XXI остается неизменным при вариации ее положения в числе.

Количество различных цифр, употребляемых в позиционной СС называется основанием СС.

Развернутая форма числа - это запись, которая представляет собой сумму произведений цифр числа на значение позиций.

Например: $8527=8*10^3+5*10^2+2*10^1+7*10^0$

Развернутая форма записи чисел произвольной системы счисления имеет вид:

$$X = \sum_{i=n-1}^{-m} a_i q^i ,$$

где X - число;

a - основа системы исчисления;

i - индекс;

m - количество разрядов числа дробной части;

n - количество разрядов числа целой части.

Десятичная СС – *это* система, в которой для записи чисел используются цифры от 0 до 9. Основанием десятичной системы счисления является число 10.

Двоичная СС - это система, в которой для записи чисел используются две цифры 0 и 1. Основанием двоичной системы счисления является число 2.

Если основание используемой СС больше десяти, то для цифр вводят буквенное обозначение.

Например: если 10=A, а 11=B.

В шестнадцатеричной СС основа - это цифры 0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15 с соответствующими обозначениями 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F. Примеры чисел: 17D, F12AH.

В ВТ применяют позиционные СС с недесятичным основанием: двоичную, восьмеричную, шестнадцатеричную. Для обозначения используемой СС число снабжают верхним или нижним индексом, в котором записывают основание СС. Другой способ – использование латинских букв после записи числа:

D – десятичная СС
B – двоичная СС
O – восьмеричная СС
H – 16-ричная СС.

Несмотря на то, что 10-тичная СС имеет широкое распространение, цифровые ЭВМ строятся на двоичных элементах, т.к. реализовать элементы с 10 четко различимыми состояниями сложно. Историческое развитие ВТ сложилось таким образом, что ЭВМ строятся на базе двоичных цифровых устройств: триггеров, регистров, счетчиков, логических элементов и т.д.

16-ая и 8-ая СС используются при составлении программ на языке машинных кодов для более короткой и удобной записи двоичных кодов – команд, данных, адресов и операндов.

Задача перевода из одной СС в другую часто встречается при программировании, особенно, на языке Ассемблера. Например, отыскать неисправность в ЭВМ невозможно без представлений о двоичной СС.

В таблице приведены некоторые числа, представленные в различных СС.

Двоичные числа	**Восьмеричные числа**	**Десятичные числа**	**Шестнадцатеричные числа**
0	0	0	0
1	1	1	1
10	2	2	2
11	3	3	3
100	4	4	4
101	5	5	5
110	6	6	6
111	7	7	7
1000	10	8	8
1001	11	9	9
1010	12	10	A
1011	13	11	B
1100	14	12	C
1101	15	13	D
1110	16	14	E
1111	17	15	F

2.1 Перевод чисел из произвольной системы в десятичную

Для перевода числа из любой позиционной СС в десятичную необходимо использовать развернутую форму числа, заменяя, если это необходимо, буквенные обозначения соответствующими цифрами.

Пример:

$$1101_2=1*2^3+1*2^2+0*2^1+1*2^0=13_{10}$$

$$17D.ECH=12\cdot16^{-2} + 14\cdot16^{-1} +13\cdot16^0 + 7\cdot16^1 + 1\cdot16^2=381.921875$$

2.2 Перевод чисел из десятичной СС в заданную

1) Для преобразования целых чисел десятичной системы счисления в число любой системы счисления последовательно выполняют деление

нацело на основание СС, пока не получат нуль. Числа, которые возникают как остаток от деления на основание СС, представляют собой последовательную запись разрядов числа в выбранной СС от младшего разряда к старшему. Поэтому для записи самого числа остатки от деления записывают в обратном порядке.

Например:

```
475|2
   1  237 |2
      1  118|2
            0  59|2
                1  29|2
                     1  14|2
                          0  7|2
                             1 3|2
                               1 1|2
                                  1 0
```

Читая остатки от деления снизу вверх, получим 111011011.

Проверка:

$$1*2^8+1*2^7+1*2^6+0*2^5+1*2^4+1*2^3+0*2^2+1*2^1+1*2^0=$$

$$=1+2+8+16+64+128+256=475_{10}.$$

2) Для преобразования <u>десятичных дробей</u> десятичной СС в число любой СС последовательно выполняют умножение на основание системы счисления , пока дробная часть произведения не станет равной нулю. Полученные целые части являются разрядами числа в новой системе, и их необходимо представлять цифрами этой новой системы счисления. Целые части в дальнейшем отбрасываются.

<u>Например</u>: перевести число 0.375_{10} в двоичную СС.

```
  0.375  →0
      2
 0.750  →0
     2
  1.50   →1
  0.50
     2
  1.00   →1
```

Полученный результат - 0.011_2.

Необходимо отметить, что не каждое число может быть точно выражено в новой системе счисления, поэтому иногда вычисляют только требуемое количество разрядов дробной части, округляя последний разряд.

3. Коды для представления чисел

При выполнении арифметических операций в ЭВМ применяют специальные коды для представления чисел (с целью упрощения арифметических операций): прямой, обратный и дополнительный коды чисел. Например, упрощается определение знака результата операции, вычитание есть сложение кодов, облегчено определение переполнения разрядной сетки.

Прямой код (представление в виде абсолютной величины со знаком) двоичного числа – это само двоичное число, в котором все цифры, изображающие его значение, записываются как в математической записи, а знак числа записывается двоичной цифрой.

Прямой код почти не отличается от принятого в математике: для выявления абсолютной величины (модуля) числа, надо отбросить цифру, обозначающую его знак. Старший бит слова является битом хранения знака или знаковым разрядом. Все последующие биты слова представляют значащие разряды числа.

Прямой код используется при хранении чисел в памяти ЭВМ, а также при выполнении операций умножения и деления, но формат представления чисел в прямом коде неудобен для использования в вычислениях, поскольку сложение и вычитание положительных и отрицательных чисел выполняется по–разному, а потому требуется анализировать знаковые разряды операндов. Поэтому прямой код практически не применяется при реализации в АЛУ арифметических операций над целыми числами. Вместо этого формата широкое распространение получили форматы представления чисел в обратном и дополнительном кодах.

Пример: Дано число X=-1011. Перевести число в прямой код.

$X_{ПК}$=1.1011

Обратный код положительного числа совпадает с прямым, а при записи отрицательного числа все его цифры, кроме цифры, изображающей знак числа, заменяются на противоположные (0 заменяется на 1, а 1 – на 0).

Пример: Дано число X=-1011. Перевести число в обратный код.

$X_{обр}$=1.0100

Дополнительный код (представление в виде дополнения до двойки) положительного числа совпадает с прямым, а код отрицательного числа образуется как результат увеличения на 1 его обратного кода. То есть, процесс построения дополнительного кода отрицательного числа можно разбить на два этапа – построить обратный код, а затем из него построить дополнительный.

Пример: Дано число X=-1011. Перевести в дополнительный код.

$X_{доп}$=1.0101

Пример: Дано число X=-0,1011. Перевести число в прямой, обратный и дополнительный код, при условии, что разрядная сетка содержит 8 разрядов.

$X_{пр}$=1,1011000

$X_{обр}$=1,0100111

$X_{доп}$=1,0101000

Основным достоинством дополнительного кода является то, что в нем единообразно реализуются операции сложения чисел разных знаков, а операцию вычитания можно свести к операции сложения заменой знака вычитаемого на обратный, причем для реализации дополнительного кода, не требуется ни каких дополнительных аппаратных устройств.

3.1 Операция сложения в обратном и дополнительном кодах

Сложение и вычитание чисел в обратном и дополнительном кодах выполняется с использованием обычного правила арифметического сложения многоразрядных чисел. Общей для этих кодов особенностью является то, что при поразрядном сложении чисел разряды, изображающие знаки числа рассматриваются как равноправные разряды двоичного числа, которые складываются друг с другом и с единицей переноса из предыдущего разряда числа по обычным правилам арифметики. Различия же обратного и дополнительного кодов связаны с тем, что делается с единицей переноса из старшего разряда.

При сложении чисел в дополнительном коде: единица переноса из старшего разряда игнорируется (теряется).

При сложении чисел в обратном коде эту единицу надо прибавить к младшему разряду результата.

<u>Пример:</u> Сложить числа +12 и -5.

а) В обратном коде

Десятичная форма	Двоичная форма	Прямой код	Обратный код
+12	+1100	00001100	00001100
-5	-101	10000101	11111010

Выполним сложение:

	+	0	0	0	0	1	1	0	0
		1	1	1	1	1	0	1	0
+	1	0	0	0	0	0	1	1	0
									1
	0	0	0	0	0	0	1	1	1

Результат в обратном коде – 00000111. Поскольку знаковый разряд равен 0, результат положительный, и, следовательно, запись кода числа совпадает с записью прямого кода. Теперь можно восстановить алгебраическую запись результата. Он равен +111 (незначащие нули отброшены), или в десятичной форме +7. Проверка (+12-5=+7) показывает, что результат верный.

<u>б) В дополнительном коде</u>

Десятичная форма	Двоичная форма	Прямой код	Обратный код	Дополнительный код
+12	+1100	00001100	00001100	00001100
-5	-101	10000101	11111010	11111011

Выполним сложение в дополнительном коде:

	+	0	0	0	0	1	1	0	0
		1	1	1	1	1	0	1	1
	1	0	0	0	0	0	1	1	1
	0	0	0	0	0	0	1	1	1

Результат в дополнительном коде – 00000111. Поскольку знаковый разряд равен 0, результат положительный. Теперь можно восстановить алгебраическую запись результата. Он равен +111 (незначащие нули отброшены), или в десятичной форме +7. Проверка (+12-5=+7) показывает, что результат верный.

Теория информации и теория кодирования

Теория информации и теория кодирования изучают информацию в виде абстрактного объекта, лишенного конкретного содержания, выявляют общие свойства информации, законы, управляющие ее возникновением, развитием и уничтожением, а атк же изучают вопросы передачи информации по каналам связи, кодирование и декодирование информации при ее посылке.

Теория информации использует понятия и методы теории вероятностей. Причем в ней не придерживаются специальной «информационной» терминологии.

Теория кодирования – это раздел информации, связанный с задачами кодирования и декодирования сообщений, посылаемых из источника к приемнику информации.

Код – это система соответствий между элементами исходного сообщения и сочетаниями символов, при помощи которых эти сообщения могут быть зафиксированы и при необходимости переданы на расстояние или использованы для дальнейшей обработки.

Код – это совокупность условных символов (сигналов), обозначающих определенное сообщение.

Код – это множество слов в некотором алфавите, поставленное во взаимнооднозначное соответствие другому множеству.

Цель кодирования - представить информацию в более компактной форме для дальнейшей передачи и обработки.

Условные сигналы, составляющие код, называют кодовыми комбинациями (словами).

Число элементов или знаков, образующих кодовую комбинацию, называют значимостью кода.

1. Код Шеннона - Фано

В теории кодирования фундаментальное значение имеют две теоремы, доказанные К. Шенноном.

Первая теорема говорит о том, что при канале, не вносящем своих помех, можно закодировать сообщение таким образом, чтобы среднее число элементов кода было бы минимальным.

Вторая теорема Шеннона относится к каналам с помехами. Согласно этой теореме, для таких каналов всегда существует способ кодирования, при котором сообщения будут передаваться с какой угодно достоверностью, если только скорость передачи не превышает пропускной способности канала связи.

Рассмотрим два наиболее распространенных метода кодирования на примере двоичных кодов.

В методе кодирования Шеннона - Фано (1949 г) изначально считается, что буквы статистически не связаны между собой, причем получаемый код является неравномерным и обратимым. Код строится следующим образом: символы алфавита сообщений выписываются в таблицу в порядке убывания вероятностей. Затем (в случае двоичного кода) они разделяются на группы так, чтобы суммы вероятностей в каждой из групп были по возможности одинаковыми. Всем буквам верхней половины первого символа присваивается 0, а второй половины 1. Каждая из полученных групп, в свою очередь, разбивается на две подгруппы с одинаковыми суммарными вероятностями и т.д. Процесс повторяется до тех пор, пока в любой подгруппе не останется по одной букве.

Пример:

Закодировать алфавит $A=\{A_1,\ \ ,A_5\}$ двоичным кодом, если вероятности букв следующие: $p(A_1)=1/4$; $p(A_2)=1/4$; $p(A_3)=1/4$; $p(A_4)=1/8$; $p(A_5)=1/8$ (метод кодирования Шеннона - Фано).

Решение:

<table>
<tr><th>A_i</th><th>$p(A_i)$</th><th colspan="3">Кодирование</th></tr>
<tr><td>A_1</td><td>1/4</td><td rowspan="2">0</td><td>0</td><td rowspan="3"></td></tr>
<tr><td>A_2</td><td>1/4</td><td>1</td></tr>
<tr><td>A_3</td><td>1/4</td><td rowspan="3">1</td><td>0</td></tr>
<tr><td>A_4</td><td>1/8</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>A_5</td><td>1/8</td><td>1</td></tr>
</table>

Ответ:

A_1	00
A_2	01
A_3	10
A_4	110
A_5	111

Для оценки эффективности (экономичности) неравномерного кода применяется средняя длина кодовых слов, которая вычисляется по формуле:

$$\bar{l} = \sum_{i=1}^{n} l_i \cdot p(A_i),$$

где n - общее число сообщений;

l_i - длина кодового обозначения для сообщения;

$p(A_i)$ - вероятность.

2. Код Хаффмена

Близок к коду Шеннона - Фано, но еще выгодней, код Хаффмена. Получаемый по этому методу код является неравномерным и обратимым.

Суть алгоритма кодирования по методу Хаффмена заключается в следующем. Буквы алфавита А выписываются в основной столбец по убыванию их вероятностей. Две последние буквы объединяются в одну вспомогательную букву. В результате получается новый алфавит A_1 из алфавита А путем однократного сжатия последнего. Затем вероятности не участвующие в объединении, и полученная суммарная вероятность снова располагаются в порядке убывания вероятности, и процесс сжатия алфавита A_i повторяется до тех пор пока не получим единственную вероятность буквы, равную единице. При этом условимся приписывать буквам алфавита A_i значения 0 или 1.

Пример:

Закодировать по методу Хаффмена буквы алфавита, имеющие следующие вероятности: А={0.05;0.2;0.4;0.2;0.05;0.1}. Подсчитать среднюю длину кодового слова.

Решение:

A_3	0,4	**0**	0,4	**0**	0,4	**0**	0,4	**0**	0,6	**1**	$A_1 \to$ **11001**
A_2	0,2	**10**	0,2	**10**	0,2	**10**	0,4	**11**	0,4	**0**	$A_2 \to$ **10**
A_4	0,2	**111**	0,2	**111**	0,2	**111**	0,2	**10**			$A_3 \to$ **0**
A_6	0,1	**1101**	0,1	**1101**	0,2	**110**					$A_4 \to$ **111**
A_1	0,05	**11001**	0,1	**1100**							$A_5 \to$ **11000**
A_5	0,05	**11000**									$A_6 \to$ **1101**

Подсчитаем среднюю длину кодового слова по формуле:

$$\bar{l} = \sum_{i=1}^{n} l_i \cdot p(A_i) \quad \Rightarrow \; l = 5*0.05+2*0.2+1*0.4+0.2*3+5*0.05+4*0.1=2.3.$$

В теории кодировании принят следующий *тезис:* ни для какого другого метода кодирования букв некоторого алфавита среднее число элементарных сигналов приходящихся на одну букву, не может быть меньше того, какое получается при кодировании по методу Хаффмена.