

Лекция.

Средства и методы сбора, хранения и передачи информации.

В основе любой информационной деятельности лежат так называемые **информационные процессы** — совокупность последовательных действий (операций), производимых над информацией для получения какого-либо результата (достижения цели). Информационные процессы могут быть различными, но все их можно свести к трем основным: **обработка информации, передача информации и хранение информации.**

Обработка информации

Обработка информации — это целенаправленный процесс изменения формы ее представления или содержания.

Из курса информатики основной школы вам известно, что существует два различных типа обработки информации:

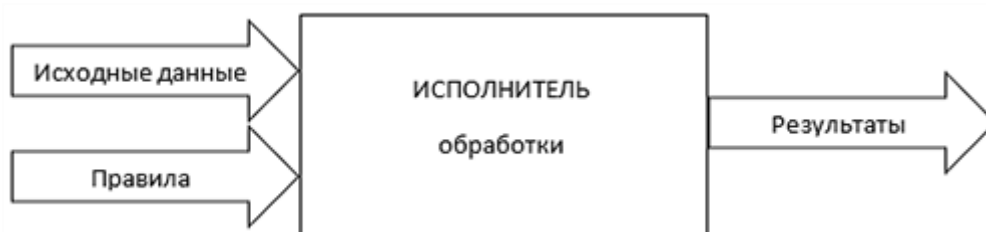
1. обработка, связанная с получением новой информации (например, нахождение ответа при решении математической задачи; логические рассуждения и др.);
2. обработка, связанная с изменением формы представления информации, не изменяющая ее содержания. К этому типу относятся:

— кодирование — переход от одной формы представления информации к другой, более удобной для восприятия, хранения, передачи или последующей обработки; один из вариантов кодирования — шифрование, цель которого — скрыть смысл информации от посторонних;

— структурирование — организация информации по некоторому правилу, связывающему ее в единое целое (например, сортировка);

— поиск и отбор информации, требуемой для решения некоторой задачи, из информационного массива (например, поиск в словаре).

Общая схема обработки информации может быть представлена следующим образом:



Исходные данные — это информация, которая подвергается обработке.

Правила — это информация процедурного типа. Они содержат сведения для исполнителя о том, какие действия требуется выполнить, чтобы решить задачу.

Исполнитель — тот объект, который осуществляет обработку. Это может быть человек или компьютер. При этом человек, как правило, является неформальным, творчески действующим исполнителем. Компьютер же способен работать только в строгом соответствии с правилами, т.е. является формальным исполнителем обработки информации.

Рассмотрим отдельные процессы обработки информации более подробно.

Кодирование информации

Кодирование информации — это обработка информации, заключающаяся в ее преобразовании в некоторую форму, удобную для хранения, передачи, обработки информации в дальнейшем.

Код — это система условных обозначений (кодовых слов), используемых для представления информации.

Кодовая таблица — это совокупность используемых кодовых слов и их значений.

Нам уже знакомы примеры равномерных двоичных кодов — пятиразрядный код Бодо и восьмиразрядный код ASCII.

Самый известный пример неравномерного кода — код Морзе. В этом коде все буквы и цифры кодируются в виде различных последовательностей точек и тире.

A	· —	S	· · ·
B	— · · ·	T	— ·
C	— · — ·	U	· · —
D	— · ·	V	· · · —
E	·	W	· — —
F	· · — ·	X	— · · —
G	— — ·	Y	— · — —
H	· · · ·	Z	— — · ·
I	· ·	1	· — — — —
J	· — — —	2	· · — — —
K	— · —	3	· · · — —
L	· — · ·	4	· · · · —
M	— —	5	· · · · ·
N	— ·	6	— · · · ·
O	— — —	7	— — · · ·
P	· — — ·	8	— — — · ·
Q	— — · —	9	— — — — ·
R	· — ·	0	— — — — —

Чтобы отделить коды букв друг от друга, вводят еще один символ — пробел (пауза). Например, слово «byte», закодированное с помощью кода Морзе, выглядит следующим образом:

— · · · — · — — — ·

При использовании неравномерных кодов важно понимать, сколько различных кодовых слов они позволяют построить.

Пример 1. Имеющаяся информация должна быть закодирована в четырехбуквенном алфавите {A, B, C, D}. Выясним, сколько существует различных последовательностей из 7 символов этого алфавита, которые содержат ровно пять букв A.

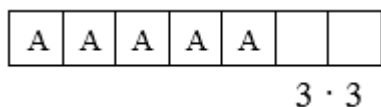
Нас интересует семибуквенная последовательность, т. е.

--	--	--	--	--	--	--

Если бы у нас не было условия, что в ней должны содержаться ровно пять букв A, то для первого символа было бы 4 варианта, для второго — тоже 4, и т. д.

Тогда мы получили бы: $4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 16384$ варианта.

Теперь вернемся к имеющемуся условию и заполним пять первых мест буквой А. Получим:



Так как на 6-м и 7-м местах могут стоять любые из трех оставшихся букв В, С, D, то всего существует 9 ($3 \cdot 3$) вариантов последовательностей.

Но ведь буквы А могут находиться на любых пяти из семи имеющихся позиций. А сколько таких вариантов всего?

Вспоминая комбинаторику, найдем число сочетаний $C_7^5 = 21$, т. е. существует 21 вариант выбора в семибуквенной последовательности ровно пяти мест для размещения букв А. Для каждого из этих 21 вариантов имеется 9 разных вариантов заполнения двух оставшихся мест. В итоге существует 189 ($21 \cdot 9$) различных последовательностей.

Главное условие использование неравномерных кодов — возможность однозначного декодирования записанного с их помощью сообщения. Именно поэтому в технических системах широкое распространение получили особые неравномерные коды — префиксные коды.

Префиксный код — код со словом переменной длины, обладающий тем свойством, что никакое его кодовое слово не может быть началом другого (более длинного) кодового слова.

Например:

1. Код, состоящий из слов 0, 10 и 11, является префиксным.
2. Код, состоящий из слов 0, **10**, 11 и **100**, не является префиксным.

Условие, определяющее префиксный код, называется прямым условием Фано (в честь Роберта Марио Фано), и позволяет однозначно декодировать сообщения, записанные с помощью неравномерных кодов.

Также достаточным условием однозначного декодирования неравномерного код является обратное условие Фано. В нем требуется, чтобы никакой код не был окончанием другого (более длинного) кода.

Пример 2. Двоичные коды для 5 букв латинского алфавита представлены в таблице:

A	B	C	D	E
000	01	100	10	011

Выясним, какое сообщение закодировано с помощью этих кодов двоичной строкой: 0110100011000.

Можно заметить, что для заданных кодов не выполняется прямое условие Фано:

$B=01$, $E=011$, и $D=10$, $C=100$.

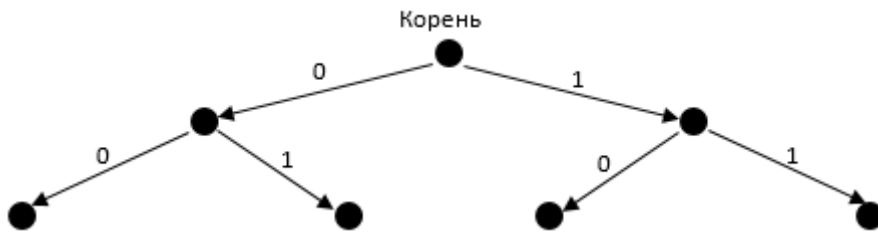
А вот обратное условие Фано выполняется: никакое кодовое слово не является окончанием другого. Следовательно, имеющуюся строку нужно декодировать справа налево (с конца). Получим

01 10 100 011 000 = BDCEA

Для построения префиксных кодов удобно использовать бинарные деревья, в которых от каждого узла отходят только два ребра, помеченные цифрами 0 и 1.

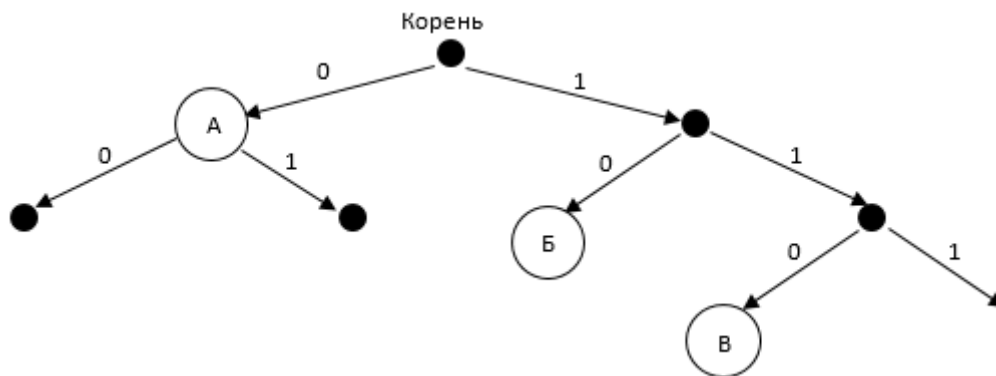
Пример 3. Для кодирования некоторой последовательности, состоящей из букв А, Б, В и Г, решили использовать неравномерный двоичный код, позволяющий однозначно декодировать полученную двоичную последовательность. При этом используются такие кодовые слова: А — 0, Б — 10, В — 110. Каким кодовым словом может быть закодирована буква Г? Если таких слов несколько, укажите кратчайшее из них.

Построим бинарное дерево:

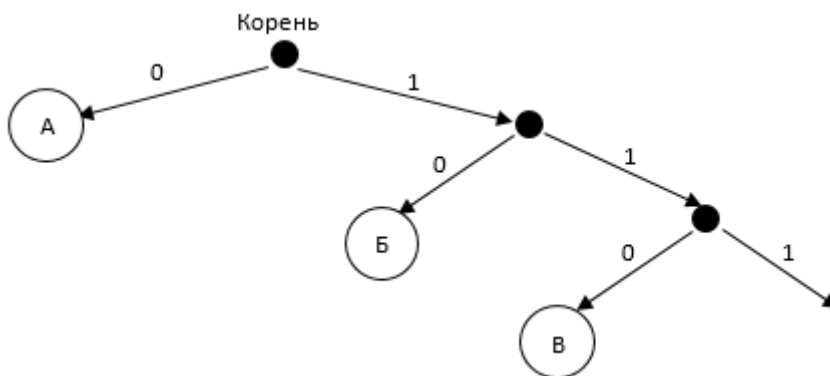


Чтобы найти код символа, нужно пройти по стрелкам от корня дерева к нужному листу, выписывая метки стрелок, по которым мы переходим.

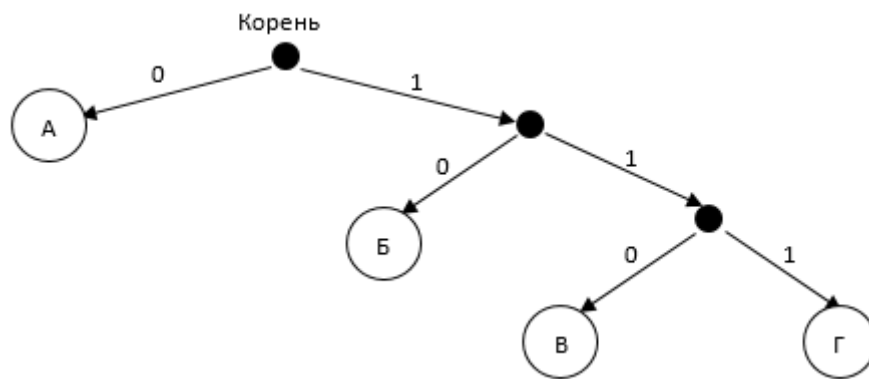
Определим положение букв А, Б и В на этом дереве, зная их коды. Получим:



Чтобы код был префиксным, ни один символ не должен лежать на пути от корня к другому символу. Уберем лишние стрелки:



На получившемся дереве можно определить подходящее расположение буквы Г и его код.



Г — 111.

Поиск информации

Задача поиска обычно формулируется следующим образом. Имеется некоторое хранилище информации — информационный массив (телефонный справочник, словарь, расписание поездов, диск с файлами и др.). Требуется найти в нем информацию, удовлетворяющую определенным условиям поиска (телефон какой-то организации, перевод слова, время отправления поезда, нужную фотографию и т. д.). При этом, как правило, необходимо сократить время поиска, которое зависит от способа организации данных и используемого алгоритма поиска.

Алгоритм поиска, в свою очередь, также зависит от способа организации данных.

Если данные никак не упорядочены, то мы имеем дело с неструктурированным набором данных. Для осуществления поиска в таком наборе применяется метод последовательного перебора.

При последовательном переборе просматриваются все элементы подряд, начиная с первого. Поиск при этом завершается в двух случаях:

— искомый элемент найден;

— просмотрен весь набор данных, но искомого элемента среди них не нашлось.

Зададимся вопросом: какое среднее число просмотров приходится выполнять при использовании метода последовательного перебора? Есть два крайних случая:

— искомый элемент оказался первым среди просматриваемых. Тогда просмотр всего один;

— искомый элемент оказался последним среди просматриваемых. Тогда количество просмотров равно N , где N — размер набора данных. Столько же просмотров нам придется выполнить даже если не сможем найти искомого элемента.

Если же провести поиск последовательным перебором достаточно много раз, то окажется, что в среднем на поиск требуемого элемента уходит $N/2$ просмотров. Эта величина определяет длительность поиска — главную характеристику поиска.

Если же информация упорядочена, то мы имеем дело со структурой данных, в которой поиск осуществляется быстрее, можно построить оптимальный алгоритм.



Одним из оптимальных алгоритмов поиска в структурированном наборе данных может быть метод половинного деления.

Напомним, что при этом методе искомый элемент сначала сравнивается с центральным элементом последовательности. Если искомый элемент меньше центрального, то поиск продолжается аналогичным образом в левой части последовательности. Если больше, то — в правой. Если же значения искомого и центрального элемента совпадают, то поиск завершается.

Пример 4. В последовательности чисел 61 87 180 201 208 230 290 345 367 389 456 478 523 567 590 требуется найти число 180.

Процесс поиска представлен на схеме:

Просмотр 1. Работаем со всей последовательностью. Определяем центральный элемент:

61 87 180 201 208 230 290 345 367 389 456 478 523 567 590

Сравниваем искомый элемент с центральным ($180 < 345$) и отбрасываем правую часть:

61 87 180 201 208 230 290

Просмотр 2. Определяем центральный элемент в левой части последовательности:

61 87 180 201 208 230 290

Сравниваем искомый элемент с центральным ($180 < 201$) и отбрасываем правую часть:

61 87 180

Просмотр 3. Определяем центральный элемент в левой части последовательности:

61 87 180

Сравниваем искомый элемент с центральным ($180 > 87$) и отбрасываем левую часть:

180

Оставшийся элемент совпадает с искомым. Поиск завершен.

Передача информации

Передача информации — это процесс распространения информации от источника к приемнику через определенный канал связи.

На рисунке представлена схема модели процесса передачи информации по техническим каналам связи, предложенная Клодом Шенноном.



Работу такой схемы можно пояснить на примере записи речи человека с помощью микрофона на компьютер.

Источником информации является говорящий человек. *Кодирующим устройством* — микрофон, с помощью которого звуковые волны (речь) преобразуются в электрические сигналы. *Канал связи* — провода, соединяющие микрофон и компьютер. *Декодирующее устройство* — звуковая плата компьютера. *Приемник информации* — жесткий диск компьютера.

При передаче сигнала могут возникать разного рода помехи, которые искажают передаваемый сигнал и приводят к потере информации. Их называют «**шумом**».

В современных технических системах связи борьба с шумом (защита от шума) осуществляется по следующим двум направлениям:

1. Технические способы защиты каналов передачи от воздействия шумов. Например, применение различных фильтров, использование специальных кабелей.
2. Внесение избыточности в передаваемое сообщение, позволяющее компенсировать потерю какой-то части передаваемой по линиям связи информации. Например, если при разговоре по телефону вас плохо слышно, то, повторяя каждое слово дважды, вы увеличиваете шансы на то, что ваш собеседник поймет вас правильно.

Но чрезмерная избыточность приводит к задержкам и удорожанию связи. Поэтому очень важно иметь алгоритмы получения оптимального кода, одновременно обеспечивающего минимальную избыточность передаваемой информации и максимальную достоверность принятой информации.

В современных системах цифровой связи для борьбы с потерей информации часто применяется следующий приём. Всё сообщение разбивается на порции — блоки. Для каждого блока вычисляется контрольная сумма, которая передаётся вместе с данным блоком. В месте приёма заново вычисляется контрольная сумма принятого блока, и если она не совпадает с первоначальной, то передача данного блока повторяется.

Важной характеристикой современных технических каналов передачи информации является их **пропускная способность** — максимально возможная скорость передачи информации, измеряемая в битах в секунду (бит/с). Пропускная способность канала связи зависит от свойств используемых носителей (электрический ток, радиоволны, свет). Так, каналы связи, использующие оптоволоконные кабели и радиосвязь, обладают пропускной способностью, в тысячи раз превышающей пропускную способность телефонных линий.

Скорость передачи информации по тому или иному каналу зависит от пропускной способности канала, а также от длины закодированного сообщения, определяемой выбранным алгоритмом кодирования информации.

Современные технические каналы связи обладают, перед ранее известными, целым рядом достоинств:

— **высокая пропускная способность**, обеспечиваемая свойствами используемых носителей;

— **надёжность**, связанная с использованием параллельных каналов связи;

— **помехозащищённость**, основанная на автоматических системах проверки целостности переданной информации;

— **универсальность** используемого двоичного кода, позволяющего передавать любую информацию — текст, изображение, звук.

Объём переданной информации I вычисляется по формуле:

$$I = v \times t,$$

где v — пропускная способность канала (в битах в секунду), а t — время передачи.

Рассмотрим пример решения задачи, имеющей отношение к процессу передачи информации.

Пример 5. Документ объемом 10 Мбайт можно передать с одного компьютера на другой двумя способами.

А. Передать по каналу связи без использования архиватора.

Б. Сжать архиватором, передать архив по каналу связи, распаковать.

Какой способ быстрее и насколько, если:

— средняя скорость передачи данных по каналу связи составляет 2^{18} бит/с;

— объем сжатого архиватором документа равен 25% от исходного объема;

— время, требуемое на сжатие документа — 5 секунд, на распаковку — 3 секунды?

Для решения данной задачи диаграмма Ганта не нужна; достаточно выполнить расчёты для каждого из имеющихся вариантов передачи информации.

Рассмотрим вариант А. Длительность передачи информации в этом случае составит:

$$10 \times 2^{23}/2^{18} = 10 \times 2^5 = 320 \text{ с.}$$

Рассмотрим вариант Б. Длительность передачи информации в этом случае составит:

5	+	<u>$0,25 \times 10 \times 2^{23}/2^{18}$</u>	+	3	= 5 + 80 + 3 = 88 с.
↑		↑		↑	
Время на сжатие		Время на передачу в упакованном виде		Время на распаковку	

Итак, вариант Б быстрее на 232 с.

Хранение информации

Сохранить информацию — значит тем или иным способом зафиксировать её на некотором носителе.

Носитель информации — это материальная среда, используемая для записи и хранения информации.

Основным носителем информации для человека является его собственная память. По отношению к человеку все прочие виды носителей информации можно назвать внешними.

Основное свойство человеческой памяти — быстрота, оперативность воспроизведения хранящейся в ней информации. Но наша память не надёжна: человеку свойственно забывать информацию. Именно для более надёжного хранения информации человек использует внешние носители, организует внешние хранилища информации.

Виды внешних носителей менялись со временем: в древности это были камень, дерево, папирус, кожа и др. Долгие годы основным носителем информации была бумага. Развитие компьютерной техники привело к созданию магнитных (магнитная лента, гибкий магнитный диск, жёсткий магнитный диск), оптических (CD, DVD, BD) и других современных носителей информации.

В последние годы появились и получили широкое распространение всевозможные мобильные электронные (цифровые) устройства: планшетные компьютеры, смартфоны, устройства для чтения электронных книг, GPS-навигаторы и др. Появление таких устройств стало возможно, в том числе, благодаря разработке принципиально новых носителей информации, которые:

1. Обладают большой информационной ёмкостью при небольших физических размерах.
2. Характеризуются низким энергопотреблением при работе, обеспечивая наряду с этим высокие скорости записи и чтения данных.
3. Энергонезависимы при хранении.
4. Имеют долгий срок службы.

Всеми этими качествами обладает флеш-память (англ. flash-memory). Выпуск построенных на их основе флеш-накопителей, называемых в просторечии «флэшками», был начат в 2000 году.