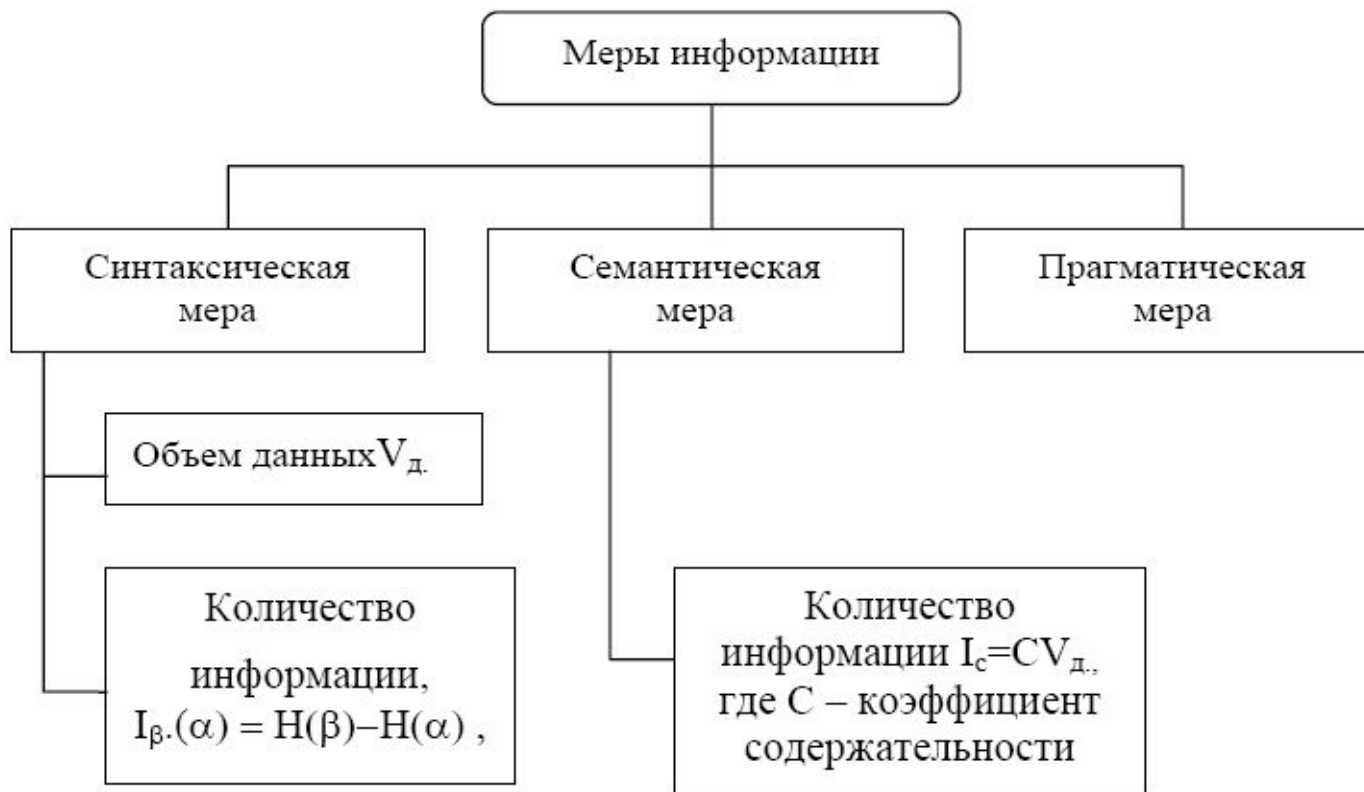




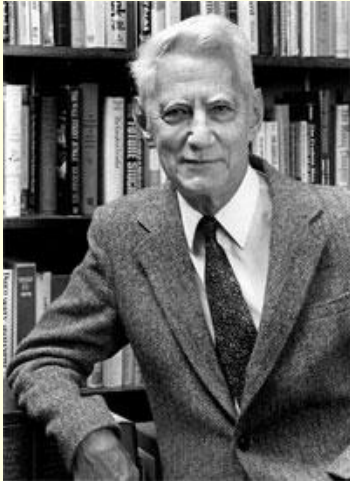
Кафедра «Автоматизированные станочные системы»
Dept. of Automated Manufacturing Systems

Информация, ее свойства и измерение

Информация (лат. information) – по ГОСТ 7.0-99 сведения, воспринимаемые человеком и (или) устройствами как отражение фактов материального или духовного мира в процессе коммуникации. В широком смысле – отражение реального мира; в узком смысле – любые сведения, являющиеся объектом хранения, передачи и преобразования информации.



Теория информации по Шеннону



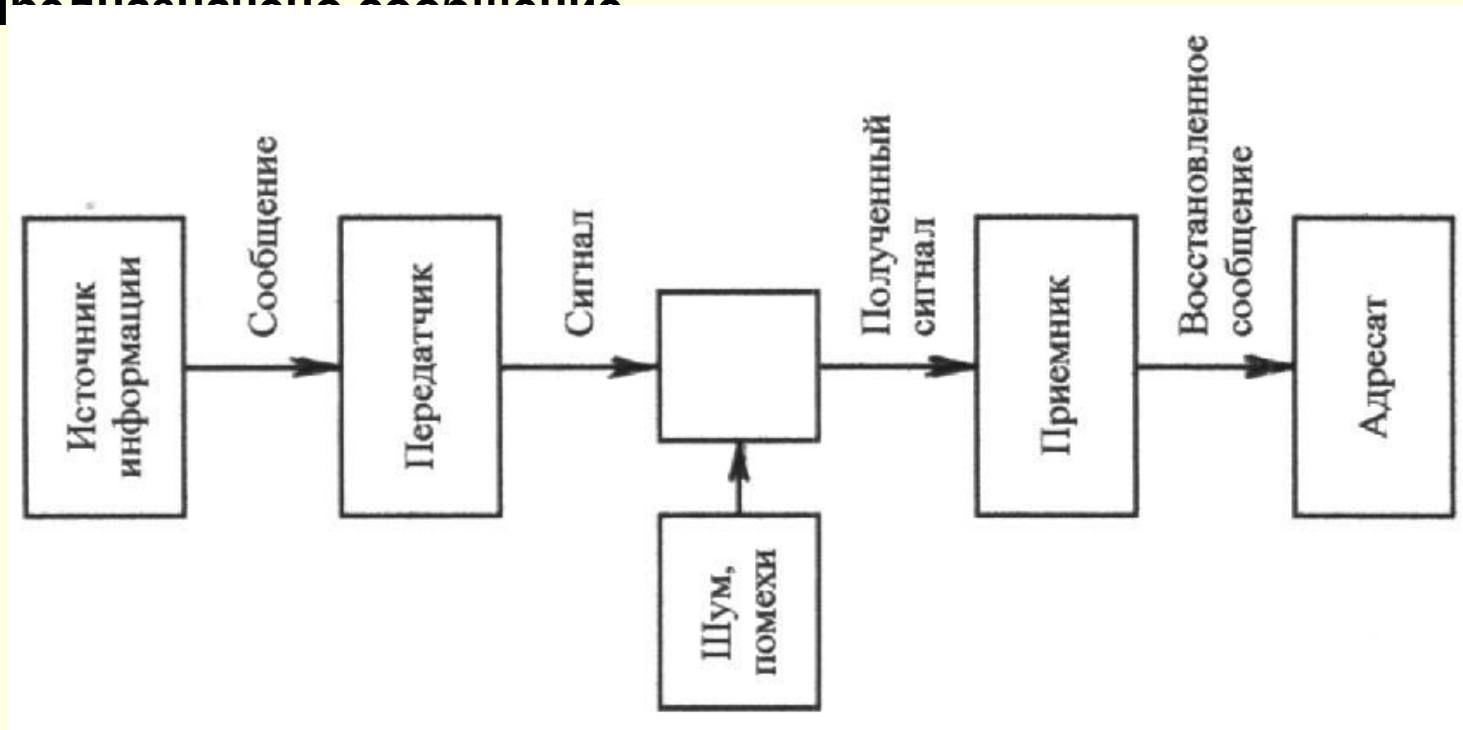
Клод Шеннон (Claude Shannon, 1916-2001),
основоположник теории информации

Дискретное (состоящие из набора символов) сообщение - представляет собой последовательность символов определенного заранее известного алфавита.

Из чего состоит процесс передачи информации:

Дискретное сообщение — это любой набор символов, который формируется **Источником** (им может быть, например, человек). От **Источника** сообщение переходит к **Передатчику**, который преобразует его к виду, который уже можно передавать по **Каналу связи**. Например, **Передатчик** может кодировать сообщение. Преобразованное сообщение называется **Сигналом**. Канал связи — это технический комплекс аппаратуры, который позволяет передать **Сигнал**.

В общем случае в процессе передачи сигнал в канале искажается **шумом**, который исходит от **Источника шума**. **Приемник** обычно выполняет операцию, обратную по отношению к той, что производится передатчиком, — **восстанавливает сообщение по сигналам**. Процесс преобразования сигнала в сообщения, осуществляемый в Приемнике, называют **декодированием**. **Адресат (получатель)** — это человек или аппарат, для которого предназначено сообщение.



Пример процесса передачи информации

Вася передает Юле свое имя. Они сидят по разные стороны стола. Вася пишет на листе бумаги букву и передвигает лист к Юле. Юля читает букву и переносит ее на свой лист бумаги. Если передача завершится успешно, на листе у Юли будет последовательность: «ВАСЯ». По столу бегают котенок, который все время норовит столкнуть передаваемый листок на пол или расцарапать бумагу.

Вася здесь — **Источник сообщений**. Лист бумаги и написанный на нем символ (буква русского алфавита), рука, которая передвигает лист — **Передатчик**. Стол — **Канал связи**. Котенок — **Источник шума**. Юля, получающая лист с написанной на нем буквой — **Приемник**. Она копирует букву на свой лист. Если Юля понимает то, что прочла (то есть она умеет читать и ни один символ не был утрачен во время передачи), значит, **Сообщение** дошло до **Получателя**. Если Юля читать не умеет — передача **Сообщения** не состоится, **Получатель** не может его принять.

Объем информации в сообщении

Попробуем сравнить следующие два вопроса:

На каком из пяти курсов университета учится студент?

Как упадет монета при подбрасывании: вверх «гербом» или «цифрой»?

В первом случае возможны **пять** равновероятных ответов, во втором – **два**. Следовательно, вероятность какого-то ответа во втором случае больше, чем в первом ($1/2 > 1/5$), в то время как неопределенность, снимаемая ответами, больше в первом случае. Любой из возможных ответов на первый вопрос снимает большую неопределенность, чем любой ответ на второй вопрос. Поэтому ответ на первый вопрос несет больше информации.

Следовательно, чем меньше вероятность какого-либо события, тем большую неопределенность снимает сообщение о его появлении и, следовательно, тем большую информацию оно несет.

Энтропия

В физике – мера неупорядоченности термодинамической системы

Пусть сообщение — осмысленное предложение на русском языке. Шеннон заметил, что **при передаче различных букв мы передаем разное количество информации**. Если мы передаем часто встречающиеся буквы, то информации меньше; при передаче редких букв — больше.

Пример – азбука Морзе. Наиболее частые буквы передаются коротко, а для редких используют более длинные цепочки.

Буква «**Е**»: (одна точка)

Буква «**Ш**» – – – – (четыре тире)

Количество информации на одну букву связано с частотой употреблений этой буквы во всех сообщениях, формируемых на языке. Чем более редкую букву мы передаем, тем больше в ней информации.

Буква	Частота встречи, %	
о	10,92	*****
а	8,89	*****
е	8,10	*****
н	6,43	*****
и	6,39	*****
л	5,87	*****
т	5,76	*****
с	5,11	*****
к	4,57	*****
р	4,16	*****
в	3,65	*****
м	3,08	*****
д	3,06	*****
у	3,03	*****
п	2,71	*****
ь	2,32	*****
ы	2,12	*****
э	2,00	*****
я	1,88	*****
г	1,69	*****
ч	1,67	*****
б	1,55	*****
й	1,19	****
ш	1,01	****
ж	0,92	***
х	0,84	***
ю	0,33	*
ц	0,29	*
щ	0,26	*
э	0,14	*
ф	0,06	
ъ	0,03	

Частоты встречаемости букв русского языка

Азбука Морзе

А, А	..	К, К	...	Ф, F
Б, В	Л, L	Х, H
В, W	...	М, M	--	Ц, C
Г, G	---	Н, N	..	Ч
Д, D	---	О, O	---	Ш
Е, E	.	П, P	Щ, Q
Ж, V	Р, R	...	Ъ, X
З, Z	С, S	...	Ы, Y
И, I	..	Т, T	-	Ь
Й, J	У, U	...	Э
Ю	Я	1
2	3	4
5	6	7
8	9	0
.	,	"
-	/	?
:				

Д.И.

8

Энтропия в теории информации — мера непредсказуемости

Пусть вероятность i -го символа алфавита (мера частоты, с которой встречается символ во всех сообщениях языка), равна p_i . Тогда информация в сообщении из n символов:

$$H = - \sum_{i=1}^n p_i \log_2 p_i$$

(здесь \log_2 — логарифм по основанию 2).

Знак "минус" в формуле Шеннона не означает, что количество информации в сообщении – отрицательная величина. Объясняется это тем, что вероятность p , согласно определению, меньше единицы, но больше нуля. Так как логарифм числа, меньшего единицы, $(\log p_i)$ – величина отрицательная, то знак «минус» сделает кол-во информации положительным.

Информация – это устраненная неопределенность.

В простейшем случае неопределенности выбор будет производиться между **двумя** взаимоисключающими друг друга равновероятными сообщениями, например между двумя качественными признаками: положительным и отрицательным импульсами, импульсом и паузой и т.п. Количество информации, переданное в этом простейшем случае, наиболее удобно принять за **единицу количества информации.**

Единица количества информации, представляющая собой выбор из двух **равновероятных** событий, получила название двоичной единицы, или **бита**. Название bit образовано от binary unit

На количество информации, получаемой из сообщения, влияет фактор неожиданности его для получателя, который зависит от вероятности получения того или иного сообщения. Чем меньше эта вероятность, тем сообщение более неожиданно и, следовательно, более информативно. Сообщение, вероятность которого высока и, соответственно, низка степень неожиданности, несет немного информации.

Чем более непредсказуемо сообщение (и чем больше его энтропия, потому что меньше вероятность) — тем больше информации оно несет. **Сенсация** — это редкое событие, предсказуемость которого очень мала, и потому велика его информационная стоимость.

Избыточность сообщений

При передаче сообщения на каком-то языке каждый следующий символ не полностью случаен

То, какой символ последует дальше, зависит от символов, уже переданных. Например, в русском языке после символа «Ъ» не может идти символ согласного звука. После двух подряд гласных «Е» третий гласный «Е» следует крайне редко.

Каждый следующий символ в некоторой степени предопределен.

Избыточность обычного текста составляет примерно 50%. Это значит, что примерно половина букв предопределяется структурой языка и лишь половина выбирается свободно.

Пример:

Прду в срд 12 янв рейс 123 Мск-Спб прлт 13:10 встрч

Дополнительные единицы измерения информации

1 байт = 8 бит;

1 Кбайт = 1024 байт;

1 Мбайт = 1024 Кбайт;

1 Гбайт = 1024 Мбайт.

В вычислительной технике и сетях передачи данных обычно значения бита 0 и 1 передаются различными уровнями напряжения либо тока. Например, в микросхемах 0 представляется напряжением в диапазоне от +0 до + 3 В, а 1 в диапазоне от 4,5 до 5,0 В.

Измерения в байтах					
Десятичная приставка			Двоичная приставка		
Название	Символ	Степень	Название	Символ	Степень
килобайт	кВ	10^3	кибибайт	KiB	2^{10}
мегабайт	МВ	10^6	мебибайт	MiB	2^{20}
гигабайт	ГВ	10^9	гибибайт	GiB	2^{30}
терабайт	ТВ	10^{12}	тебибайт	TiB	2^{40}
петабайт	ПВ	10^{15}	пебибайт	PiB	2^{50}
экзабайт	ЕВ	10^{18}	эксбибайт	EiB	2^{60}
зеттабайт	ЗВ	10^{21}	зебибайт	ZiB	2^{70}
йоттабайт	УВ	10^{24}	йобибайт	YiB	2^{80}

Машинное слово

машинно-зависимая и платформозависимая величина, измеряемая в битах или байтах, равная **разрядности регистров процессора** и/или **разрядности шины данных** (обычно некоторая степень двойки). Размер слова совпадает также с минимальным размером адресуемой информации (разрядностью данных, расположенных по одному адресу).

8 бит	12 бит	16 бит	18 бит	32 бит	36 бит	48 бит	60 бит	64 бит
1 байт	1,5 байт	2 байт	2,25 байт	4 байт	4,5 байт	6 байт	7,5 байт	8 байт
Spectrum ZX Поиск-1	PDP-8	PDP-11 i8086 i8088 i80286	PDP-1	IBM/360 i80386 i80486 Intel Pentium PowerPC (до G5)	было очень популярно до распространения IBM/360	БЭСМ-6	CDC 6600 CDC 7600	Intel Itanium AMD Opteron AMD Athlon 64 DEC Alpha Sun SparcStation PowerPC G5

Качество информации

репрезентативность

содержательность

достаточность

актуальность

своевременность

точность

достоверность

устойчивость

Репрезентативность связана с адекватным отражением свойств объекта. Важнейшее значение здесь имеют:
правильность концепции, на базе которой сформулировано исходное понятие;
обоснованность отбора существенных признаков и связей отображаемого явления.

Содержательность отражает семантическую емкость, равную отношению количества **семантической информации** в сообщении к объему обрабатываемых данных

С увеличением содержательности информации растет семантическая пропускная способность информационной системы (для получения одних и тех же сведений требуется преобразовать меньший объем данных).

Достаточность (полнота) информации означает, что она содержит минимальный, но достаточный для принятия правильного решения состав (набор показателей). Понятие **полноты информации** связано с ее смысловым содержанием (семантикой) и прагматикой. Как **неполная**, т.е. недостаточная для принятия правильного решения, так и **избыточная** информация снижает эффективность принимаемых пользователем решений.

Доступность восприятию обеспечивается выполнением соответствующих процедур ее получения и **преобразования**. Например, в информационной системе информация преобразовывается к доступной и удобной для восприятия пользователя форме.

Информация в
памяти компьютера:

0000005280:	00 00 02 00 16 00 03 00	5D 08 81 00 4C 00 0A 00
0000005290:	01 00 02 00 4C 00 0D 01	00 00 32 41 86 00 00 00
00000052A0:	0B 00 23 04 3A 04 30 04	37 04 30 04 42 04 35 04
00000052B0:	3B 04 4C 04 20 00 31 00	00 00 18 00 17 00 0F 84
00000052C0:	1D 01 11 84 E3 FE 12 64	F0 00 01 00 5E 84 1D 01
00000052D0:	60 84 E3 FE 04 00 43 4A	18 00 8C 00 FE 0F 91 01
00000052E0:	82 01 8C 00 0C 00 00 00	31 4C 0E 00 00 00 22 00
00000052F0:	21 04 42 04 38 04 3B 04	4C 04 20 00 1D 04 30 04
0000005300:	37 04 32 04 30 04 3D 04	38 04 35 04 20 00 3E 04
0000005310:	31 04 4A 04 35 04 3A 04	42 04 30 04 20 00 2B 00
0000005320:	20 00 3F 04 3E 04 20 00	46 04 35 04 3D 04 42 04
0000005330:	40 04 43 04 00 00 1A 00	18 00 03 24 01 11 84 00
0000005340:	00 12 64 F0 00 01 00 14	A4 00 00 60 84 00 00 61
0000005350:	24 01 13 00 35 08 81 43	4A 18 00 61 4A 18 00 6D
0000005360:	48 19 04 73 48 19 04 00	64 00 22 00 01 00 02 00
0000005370:	64 00 0C 01 00 00 C4 44	EF 00 00 00 10 00 1D 04
0000005380:	30 04 37 04 32 04 30 04	3D 04 38 04 35 04 20 00
0000005390:	3E 04 31 04 4A 04 35 04	3A 04 42 04 30 04 00 00
00000053A0:	21 00 19 00 03 24 01 11	84 00 00 2D 44 00 01 4D
00000053B0:	C6 0A 00 00 00 FF FF FF	FF 00 00 00 60 84 00 00
00000053C0:	61 24 01 00 08 00 43 4A	18 00 61 4A 18 00 80 00
00000053D0:	FE 0F 91 01 A2 01 80 00	0C 00 00 00 4B 5B 3A 00
00000053E0:	00 00 23 00 21 04 42 04	38 04 3B 04 4C 04 20 00

Та же информация в
преобразованном виде

План лекций
"Информатика САПР"
1. Введение
2. Понятие информации
и ее измерение
3. Понятие информационной
технологии
4. Модуляция и кодирование
сигналов

Актуальность информации определяется степенью сохранения ценности информации для управления в момент ее использования, зависит от динамики изменения ее характеристик и от интервала времени, прошедшего с момента возникновения данной информации.

Своевременность информации означает ее поступление не позже заранее назначенного момента времени, согласованного со временем решения поставленной задачи.

Точность информации определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т.п.

Семантика (смысл) информации может искажаться при ее передаче от человека к человеку

Пример искажения информации

Полковник - подполковнику:

— Завтра в 9.00 произойдет солнечное затмение, что случается не каждый день. Всех построить во дворе, солдатам иметь с собой закопченные стекла, я сам буду давать объяснения. Если будет дождь, то всех собрать в гимнастическом зале.

Подполковник — капитану:

— По приказанию полковника завтра в 9.00 произойдет затмение солнца, в случае дождя вы его не увидите и затмение состоится в гимнастическом зале, что случается не каждый день, но для чего иметь с собой закопченные стекла.

Капитан — лейтенанту:

— По приказу полковника завтра в 9.00 в гимнастическом зале состоится торжественное затмение солнца, для чего закоптить стекла. Если потребуется дождь, полковник даст особый приказ, что случается не каждый день.

Лейтенант — сержанту:

— Завтра в 9.00 полковник затмит солнце в гимнастическом зале при закопченных стеклах, что случается не каждый день в ясную погоду, когда идет дождь.

Сержант — солдатам:

— Завтра в 9.00 состоится затмение полковника по случаю солнца. Если в гимнастическом зале будет дождь, что случается не каждый день, вас соберут во дворе, но каждому закоптить по окну в спортзале.

Солдат — солдату:

— Послушай, завтра, когда мы будем коптить окна в гимнастическом зале, то случится дождь и солнце закоптит полковника.

Для информации, отображаемой цифровым кодом, известны четыре понятия точности:

формальная точность, измеряемая значением единицы младшего разряда числа;

реальная точность, определяемая значением единицы последнего разряда числа, верность которого гарантируется;

максимальная точность, которую можно получить в конкретных условиях функционирования системы;

необходимая точность, определяемая функциональным назначением показателя.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой **точностью**. Измеряется достоверность информации вероятностью необходимой точности, т.е. вероятностью того, что отображаемое информацией значение параметра отличается от истинного значения этого параметра в пределах необходимой точности.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности.
