

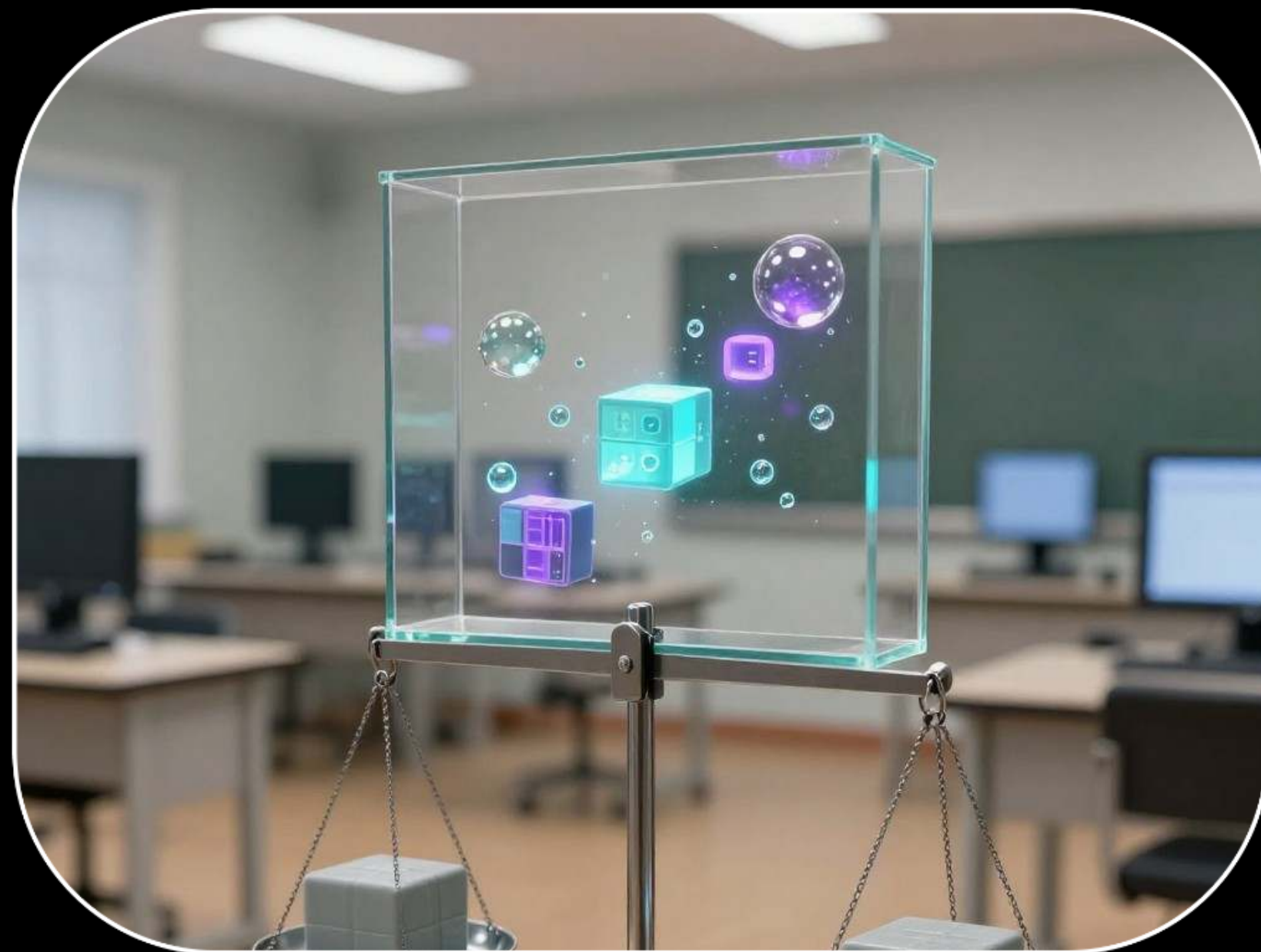
Измерение количества информации: основы для школьной информатики

Узнаем, как считать информацию в сообщениях и данных: от битов до файлов и сетей.



Зачем в информатике измеряют информацию

Информация в информатике — это данные, которые можно хранить, передавать и обрабатывать. Чтобы сравнивать тексты, изображения, аудио и видео, нужна количественная мера. В основе — двоичная запись и подсчёт 0 и 1, а также связь с размером файлов и скоростью сети.



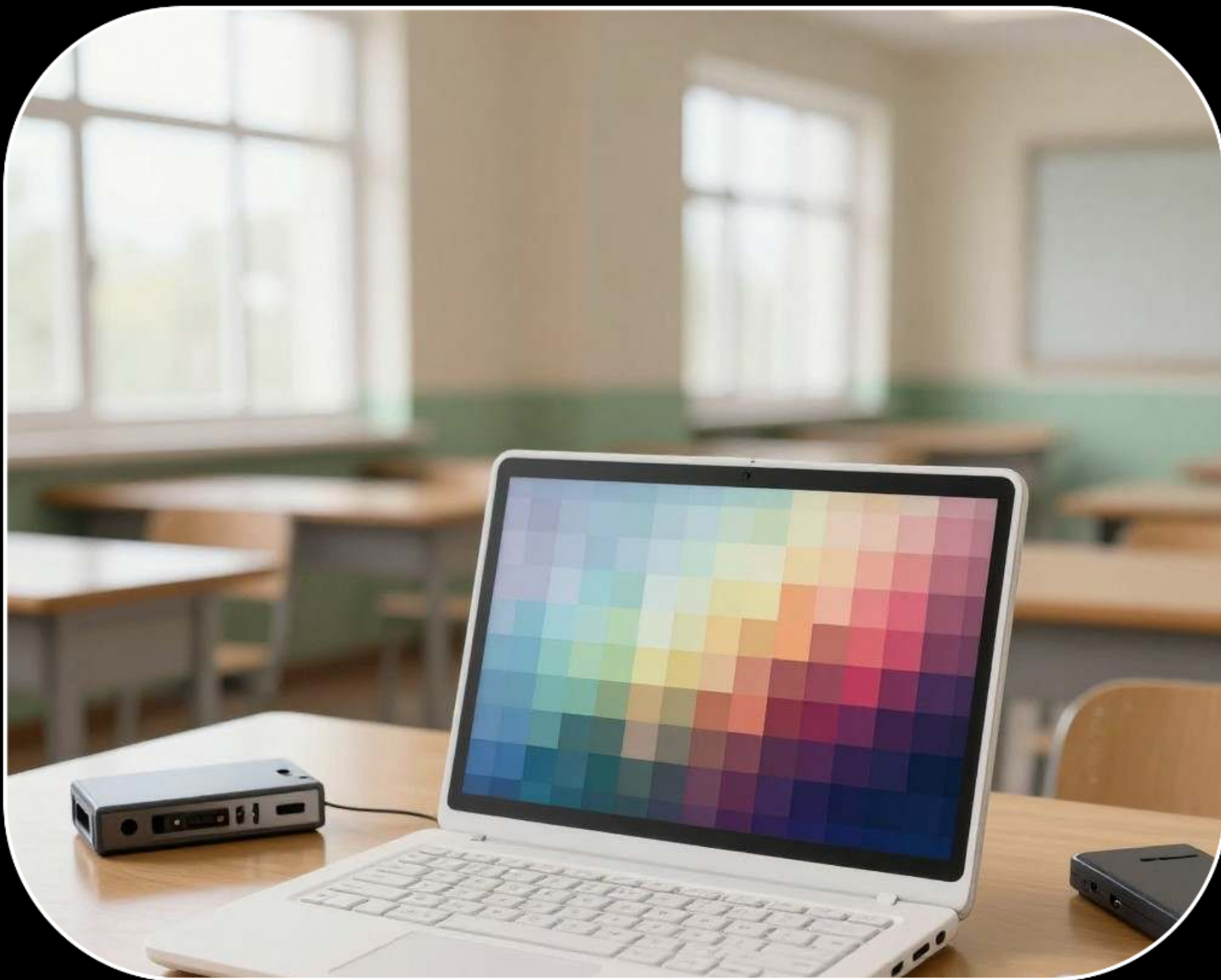
Бит и байт как базовые единицы

Бит — минимальная единица количества информации: она соответствует выбору между двумя равновозможными вариантами. На практике это похоже на вопрос с ответами «да/нет» или «0/1».

Байт равен 8 битам. Он удобен для оценок объёма памяти и размеров файлов, потому что реальные данные почти никогда не измеряются отдельными битами на уровне повседневных задач.

В школьной практике чаще используют байты, килобайты и мегабайты: так оценивают тексты, изображения и приложения. Понимание этих единиц помогает осмысленно сравнивать объёмы и понимать работу устройств.

Двоичное кодирование данных в компьютере



От смысла к двоичному представлению

Все данные в компьютере хранятся и обрабатываются в двоичном виде. Текст представляют как коды символов, изображения — как набор пикселей с числовыми значениями цвета, а звук — как последовательность числовых отсчётов, полученных при дискретизации.

Один и тот же объект — разный объём

Объект может занимать разный объём в зависимости от способа кодирования и формата файла. Поэтому один и тот же смысл (например, картинка) в разных форматах нередко требует разное число байтов для хранения и передачи.

Основные единицы измерения информации

Единица	Соотношение	Бытовой пример
Бит	8? нет: 1 бит	выбор между двумя вариантами
Байт	1 байт = 8 бит	один символ в ASCII
Килобайт	1 КБ = 1024 байта	небольшой текст
Мегабайт	1 МБ = 1024 КБ	фотография
Гигабайт	1 ГБ = 1024 МБ	память ноутбука/накопителя
Терабайт	1 ТБ = 1024 ГБ	архив данных (в школьном масштабе)

Единицы измеряют объём данных в байтах и их кратных величинах; в вычислениях применяются соотношения с основанием 1024.

Чем крупнее единица, тем удобнее оценивать реальные файлы: текст и картинка обычно «укладываются» в КБ и МБ, а накопители — в ГБ и ТБ.

Учебные определения в двоичных кратностях: 1 КБ = 1024 байта и т. д.

Формула Хартли для равновероятных вариантов

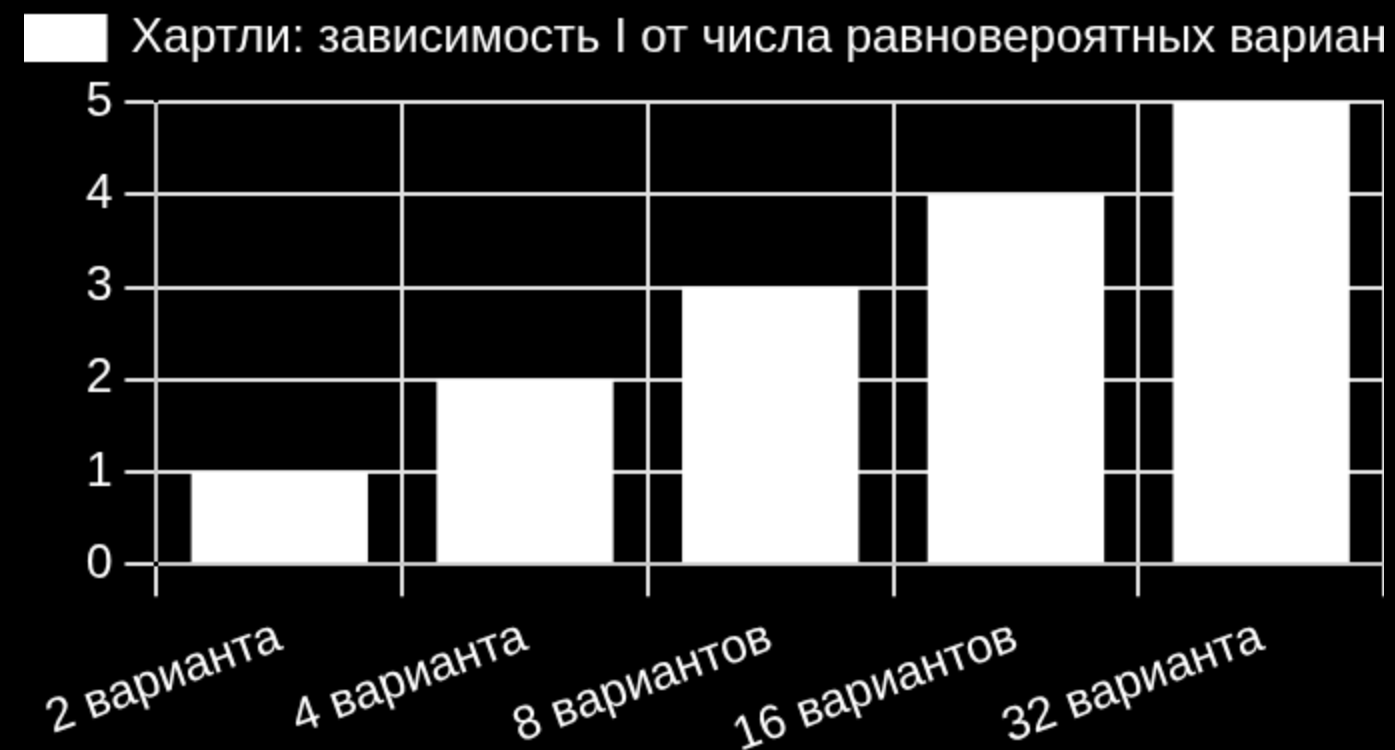
Если все варианты сообщения заранее считаются равновероятными, количество информации вычисляют по формуле $I = \log_2 N$. Здесь N — число возможных вариантов, между которыми делается выбор.

Для понимания формулы: при $N = 2$ получается $I = 1$ бит, при $N = 4$ — 2 бита, при $N = 8$ — 3 бита, при $N = 16$ — 4 бита. Это удобно связывать с «разветвлениями» решения.

Рост N увеличивает I не линейно, а логарифмически: чтобы добавить ещё 1 бит, нужно вдвое больше равновероятных вариантов. Так видно, что «сложность выбора» растёт постепенно.

Формула Хартли полезна, когда неопределённость именно равномерная: например, выбор одного ответа из нескольких одинаково вероятных вариантов теста.

Зависимость количества информации от числа вариантов



На графике видно закономерное увеличение: при удвоении числа равновероятных вариантов информация растёт на 1 бит.

Логарифмический характер формулы хорошо считывается визуально: «добавочный бит» требует нового удвоения числа вариантов.

Расчёт по формуле Хартли $I = \log_2 N$ (равновероятные варианты).

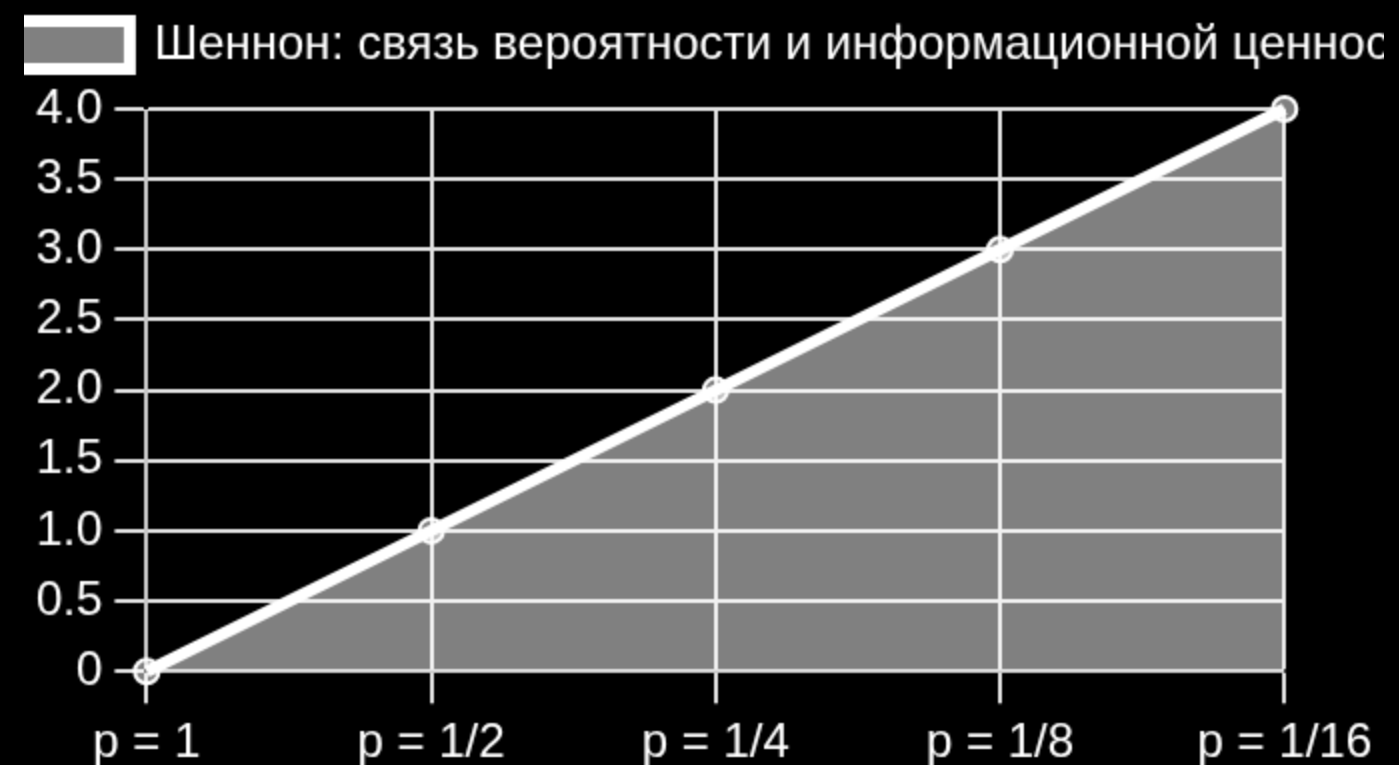
Вероятностный подход Шеннона

Когда варианты сообщения имеют разные вероятности, редкое событие обычно несёт больше информации, чем частое. Интуитивно это понятно: неожиданность повышает информационную ценность.

Для оценки применяют формулу $I = -\log_2 p$, где p — вероятность события. При $p = 1/2$ получаем 1 бит, при $p = 1/4$ — 2 бита, при $p = 1/8$ — 3 бита.

Подход Шеннона описывает ситуации, в которых неопределённость неодинакова заранее. Он помогает анализировать реальные сообщения, где вероятность отдельных исходов не равномерна.

Информация и вероятность события



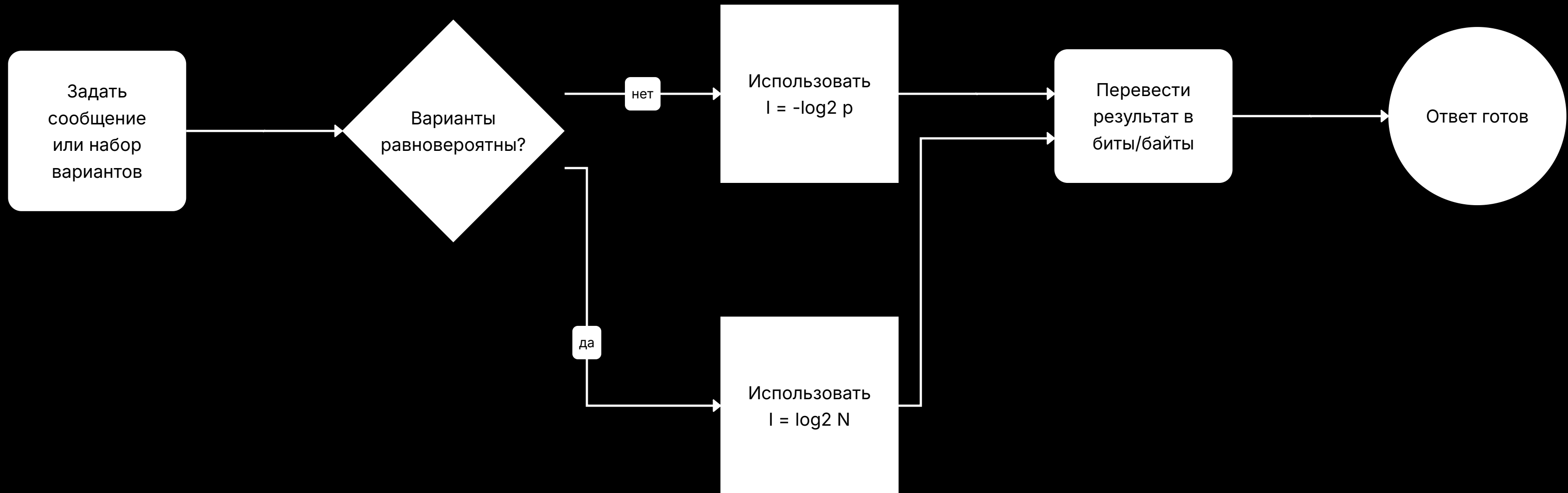
При уменьшении вероятности значение I растёт: редкий исход даёт больше информации, чем ожидаемый.

График закрепляет ключевой принцип: информационная ценность обратно связана с вероятностью, даже если без углубления в вычисления.

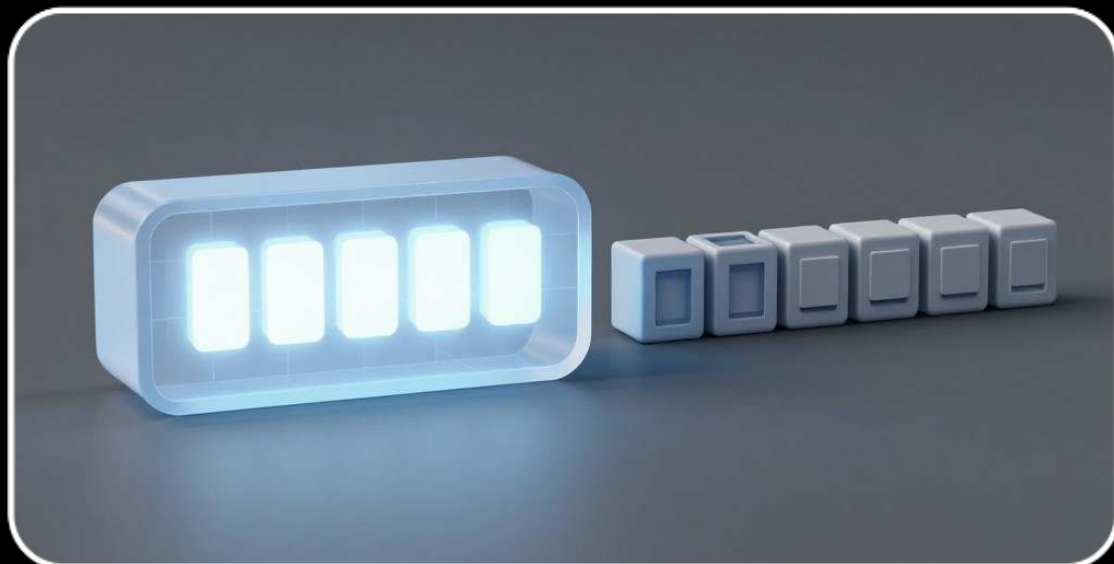
Расчёт по формуле Шеннона $I = -\log_2 p$ для значений $p = 1, 1/2, 1/4, 1/8, 1/16$.

Алгоритм вычисления количества информации

Подсчёт ведётся по числу равновероятных вариантов (Хартли) или по вероятности события (Шеннон).



Кодирование текста и объём символов



ASCII: компактно и предсказуемо

ASCII удобен для самых базовых задач: набор символов ограничен, поэтому один символ кодируется одним байтом. Представьте короткую английскую фразу: её длина в символах напрямую превращается в объём. Это делает оценки размера текста особенно простыми для школьных расчётов.



Unicode и UTF-8: больше символов — больше вариантов

В Unicode одна буква, знак или цифра могут занимать разное число байт. Для UTF-8 длина кодирования одного символа варьируется: от 1 до 4 байт. Поэтому русские тексты часто оказываются «тяжелее» по объёму, чем те же фразы, набранные латиницей. При этом смысл сохраняется, меняется лишь представление данных.



Один и тот же текст — разные размеры в разных форматах

Если посмотреть на один и тот же фрагмент текста, его размер зависит не только от кодировки, но и от файла, в котором он хранится. Например, дополнительные служебные символы, переносы строк и особенности формата могут добавлять байты сверх «чистых» символов. Для иллюстрации удобно показать таблицу символов и фрагмент строки с кодами.

1920×1080

Такое разрешение означает более двух миллионов пикселей, для которых нужно передать цветовую информацию.

Объём растрового изображения оценивают как ширина × высота × бит на пиксель. Для 1920×1080 при 24 битах на пиксель получаются значения порядка нескольких мегабайт без учёта сжатия.

Обобщённая формула оценки объёма растровых данных (разрешение и глубина цвета), учебная практика

Количество информации в звуке и видео

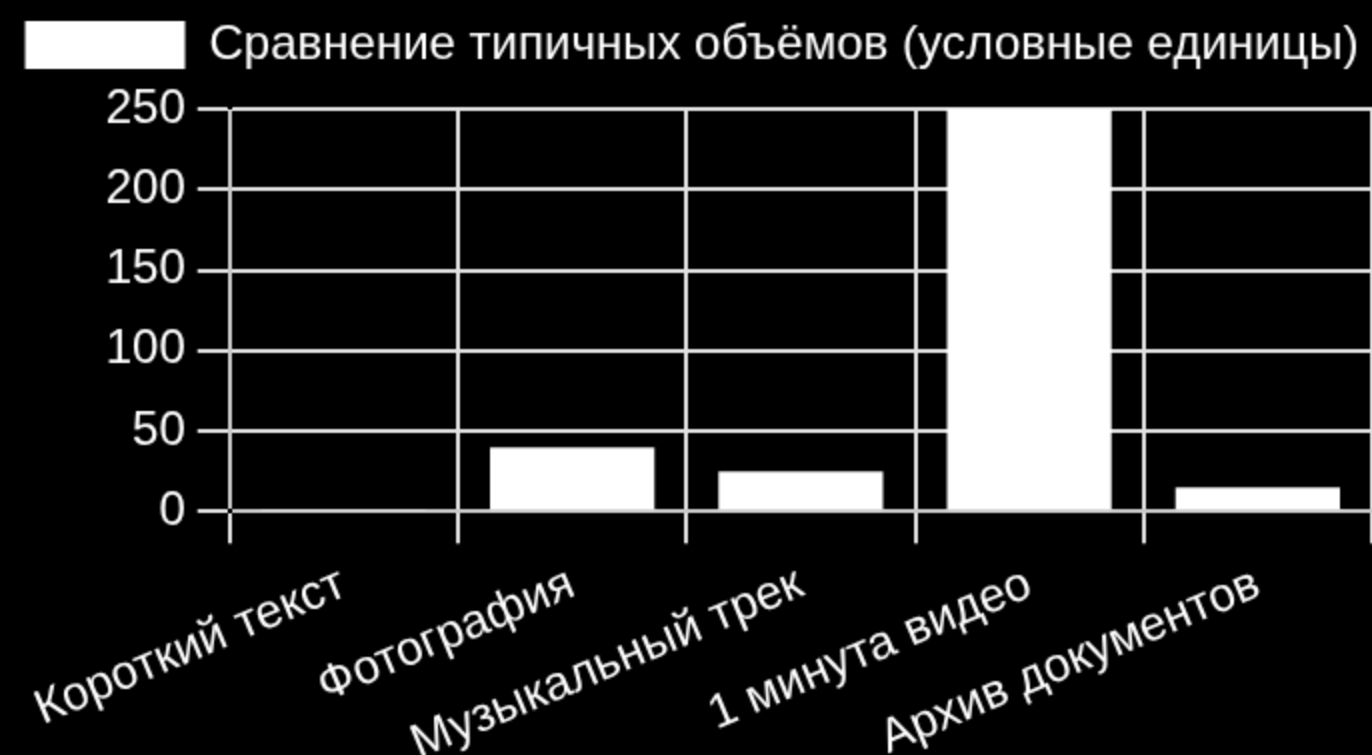
Звук в компьютере представляют как последовательность отсчётов: их частота дискретизации, разрядность и число каналов напрямую определяют объём файла. Чем чаще и точнее измерения, тем больше данных.

Для видео основными «двигателями» объёма становятся разрешение, частота кадров и длительность. Каждая секунда содержит множество картинок, и чем они детальнее, тем больше нужно байтов для хранения и передачи.

Битрейт связывает качество и объём: он показывает, сколько данных приходится на единицу времени. Поэтому длинная запись с высоким битрейтом быстро вырастает по размеру, особенно без эффективного сжатия.

На школьных примерах это видно в отличиях между аудиофайлом, записью урока, видеороликом и стримом: разные настройки создают разные значения параметров и, значит, разный объём данных для сети и памяти.

Сравнение размеров разных типов файлов



На диаграмме показано, что текст и архив обычно намного меньше мультимедийных файлов. Видео при сопоставимом качестве обычно становится самым «тяжёлым» из перечисленного набора данных.

Разрыв объясняется природой данных: изображения и звук состоят из множества числовых отсчётов, а видео добавляет «временную» размерность. Поэтому рост качества почти всегда увеличивает объём.

Типичные размеры файлов в школьной практике

Ориентировочные диапазоны помогают быстро прикидывать объём памяти и трафика. Реальный размер зависит от формата, разрешения, настроек качества и степени сжатия.

Самые заметные скачки происходят между текстом/документами и мультимедиа: видео чаще требует на порядки больше места. Это помогает разумно выбирать формат хранения и способ отправки.

Тип данных	Ориентировочный размер
Короткий текст	2–10 КБ
Фотография со смартфона	2–8 МБ
MP3-трек	3–10 МБ
1 минута HD-видео	60–150 МБ
Архив документов	1–50 МБ

Учебные ориентиры для компьютерной практики (диапазоны зависят от настроек и форматов)

Энтропия как мера неопределённости

Энтропия в теории информации описывает среднюю неопределённость источника сообщений: насколько трудно заранее угадать, какое сообщение будет следующим. Чем меньше предсказуемость, тем выше энтропия.

Если варианты распределены равномерно, каждый исход похож по «шансу», и неопределённость возрастает. Когда же некоторые сообщения встречаются заметно чаще, их можно предугадать — средняя неопределённость снижается.

Именно поэтому энтропия связана со сжатием и эффективностью кодов: чем больше энтропия, тем меньше возможностей для сокращения данных без потери информации. В задачах на школьном уровне важно понимать её как формальную «меру непредсказуемости».

Типичные ошибки при измерении информации

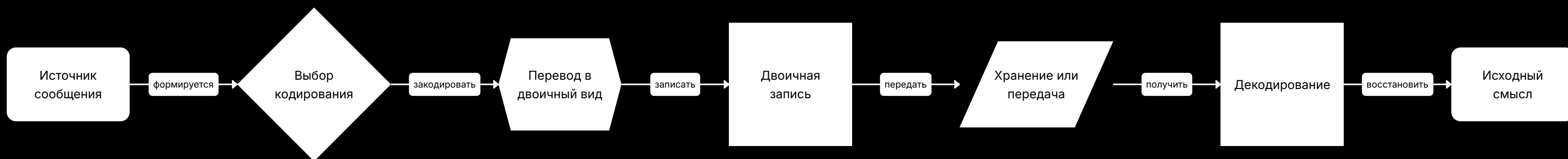


- Одна из самых частых ошибок — путать биты и байты: размер в байтах нужно переводить в биты только при необходимости. Другая ловушка — разница между 1000 и 1024 при пересчёте единиц объёма.

- Ещё одна проблема возникает из-за представления на устройстве: производители могут указывать десятичные значения, а операционная система пересчитывает иначе. Наконец, важно не смешивать объём файла с объёмом несжатых данных при хранении.

Путь информации от сообщения к носителю

Пример: текстовое сообщение, фотография или загрузка файла через сеть



Скорость передачи информации



Чем выше скорость в битах в секунду, тем быстрее канал доставляет данные. Это напрямую влияет на время загрузки файла, скорость просмотра видео и стабильность передачи в сети.

Логарифмический разброс скоростей объясняет разницу ощущений у пользователей: переход от десятков до сотен мегабит обычно радикально ускоряет загрузку и снижает задержки.

Обобщённая шкала скоростей для учебных примеров (единицы Мбит/с и Гбит/с), 2026-05

Информация измерима — значит, ею можно управлять

Количество информации определяется числом возможных вариантов и вероятностью сообщения. Бит и байт делают расчёты понятными, формула Хартли описывает равновероятные случаи, а подход Шеннона — разные вероятности. Тема важна для кодирования, хранения, сжатия и передачи данных.