

Информатика

Лекция 1

ВВЕДЕНИЕ

Учебная дисциплина информатика, появившись в конце второй половины XX в., прочно закрепилась в учебных планах университетов. Сам термин «информатика» возник в 60-х гг. XX в. во Франции для выделения области знаний, связанной с автоматизированной обработкой информации с помощью электронно-вычислительных машин. Informatique (информатика) происходит от слияния французских слов information (информация) и automatique (автоматика) и означает «информационная автоматика или автоматизированная переработка информации». Слово «информатика» вошло в употребление в разных языках: немецком – «informatik», итальянском –

«informatica», польском – «informatyka», испанском – «informatica» и др. В английском языке информатике соответствует термин computer science (дословно «компьютерная наука», «вычислительная наука»).

Несмотря на то, что появление дисциплины произошло благодаря развитию компьютерной техники и немислимо без нее, информатика является широкой областью научных знаний и связана со многими фундаментальными и прикладными дисциплинами, такими как математика, кибернетика, физика, электроника, радиотехника, философия, лингвистика.

Важное место в курсе информатики занимает теоретическая информатика, отвечающая за изучение структуры и общих свойств информации и информационных процессов, разработку общих принципов построения информационной техники и технологии. Данное пособие посвящено основным темам теоретической информатики: понятию информации и измерению ее количества, системам счисления, логическим основам построения цифровых устройств, представлению информации в цифровых устройствах.

ИНФОРМАЦИЯ и Данные

Термин информация происходит от латинского information, что означает разъяснение, осведомление, изложение.

Первоначальное значение этого термина – «сведения, передаваемые людьми устным, письменным или иным способом».

В философском смысле информация есть отражение реального мира с помощью сведений (сообщений). В широком смысле информация – это общенаучное понятие, включающее в себя обмен сведениями между людьми, обмен сигналами между живой и неживой природой, людьми и устройствами.

Мы живем в материальном мире. Все, что нас окружает и с чем мы сталкиваемся ежедневно, относится либо к физическим телам, либо к физическим полям. Из курса физики мы знаем, что состояния абсолютного покоя не существует и физические объекты находятся в состоянии непрерывного движения и изменения, которое сопровождается обменом энергией и ее переходом из одной формы в другую.

Все виды энергообмена сопровождаются появлением сигналов, то есть все сигналы имеют в своей основе материальную энергетическую природу. При взаимодействии сигналов с физическими телами в последних возникают определенные изменения свойств — это явление называется регистрацией сигналов. Такие изменения можно наблюдать, измерять или фиксировать иными способами — при этом возникают и регистрируются новые сигналы, то есть образуются данные.

*Данные – это
зарегистрированные сигналы*

– ОПРЕДЕЛЕНИЕ

Данные — это зарегистрированные сигналы.

Обратим внимание на то, что данные несут в себе информацию о событиях, произошедших в материальном мире, поскольку они являются регистрацией сигналов, возникших в результате этих событий. Однако данные не тождественны информации. Наблюдая излучения далеких звезд, человек получает определенный поток данных, но станут ли эти данные информацией, зависит еще от очень многих обстоятельств. Рассмотрим ряд примеров.

Наблюдая за состязаниями бегунов, мы с помощью механического секундомера регистрируем начальное и конечное положение стрелки прибора. В итоге мы измеряем величину ее перемещения за время забега — это регистрация данных. Однако информацию о времени преодоления дистанции мы пока не получаем. Для того чтобы данные о перемещении стрелки дали информацию о времени забега, необходимо наличие метода пересчета одной физической величины в другую. Надо знать цену деления шкалы секундомера (или знать метод ее определения) и надо также знать, как умножается цена деления прибора на величину перемещения, то есть надо еще обладать математическим методом умножения.

Если вместо механического секундомера используется электронный, суть дела не меняется. Вместо регистрации перемещения стрелки происходит регистрация количества тактов колебаний, произошедших в электронной системе за время измерения. Даже если секундомер непосредственно отображает время в секундах и нам не нужен метод пересчета, то метод преобразования данных все равно присутствует — он реализован специальными электронными компонентами и работает автоматически, без нашего участия.

Прослушивая передачу радиостанции на незнакомом языке, мы получаем данные, но не получаем информацию в связи с тем, что не владем методом преобразования данных в известные нам понятия. Если эти данные записать на лист бумаги или на магнитную ленту, изменится форма их представления, произойдет новая регистрация и, соответственно, образуются новые данные. Такое преобразование можно использовать, чтобы все-таки извлечь информацию из данных путем подбора метода, адекватного их новой форме. Для обработки данных, записанных на листе бумаги, адекватным может быть метод перевода со словарем, а для обработки данных, записанных на магнитной ленте, можно пригласить переводчика, обладающего своими методами перевода, основанными на знаниях, полученных в результате обучения или предшествующего опыта.

Если в нашем примере заменить радиопередачу телевизионной трансляцией, ведущейся на незнакомом языке, то мы увидим, что наряду с данными мы всё-таки получаем определенную (хотя и не полную) информацию. Это связано с тем, что люди, не имеющие дефектов зрения, априорно владеют адекватным методом восприятия данных, передаваемых электромагнитным сигналом в полосе частот видимого спектра с интенсивностью, превышающей порог чувствительности глаза. В таких случаях говорят, что метод известен по контексту, то есть данные, составляющие информацию, имеют свойства, однозначно определяющие адекватный метод получения этой информации. (Для сравнения скажем, что слепому «телезрителю» контекстный метод неизвестен и он оказывается в положении радиослушателя, пример с которым был рассмотрен выше.)

Понятие об информации

Несмотря на то, что с понятием информации мы сталкиваемся ежедневно, строгого и общепризнанного ее определения до сих пор не существует, поэтому вместо определения обычно используют понятие об информации. Понятия, в отличие от определений, не даются однозначно, а вводятся на примерах, причем каждая научная дисциплина делает это по-своему, выделяя в качестве основных компонентов те, которые наилучшим образом соответствуют ее предмету и задачам. При этом типична ситуация, когда понятие об информации, введенное в рамках одной научной дисциплины, может опровергаться конкретными примерами и фактами, полученными в рамках другой науки. Например, представление об информации как о совокупности данных, повышающих уровень знаний об объективной реальности окружающего мира, характерное для естественных наук, может быть опровергнуто в рамках социальных наук. Нередки также случаи, когда исходные компоненты, составляющие понятие информации, подменяют свойствами информационных объектов, например, когда понятие информации вводят как совокупность данных, которые «могут быть усвоены и преобразованы в знания».

Для информатики как технической науки понятие информации не может основываться на таких антропоцентрических понятиях, как знание, и не может опираться только на объективность фактов и свидетельств. Средства вычислительной техники обладают способностью обрабатывать информацию автоматически, без участия человека, и ни о каком знании или незнании здесь речь идти не может. Эти средства могут работать с искусственной, абстрактной и даже с ложной информацией, не имеющей объективного отражения ни в природе, ни в обществе.

Здесь мы даем новое определение информации, основанное на ранее продемонстрированном факте взаимодействия данных и методов в момент ее образования.

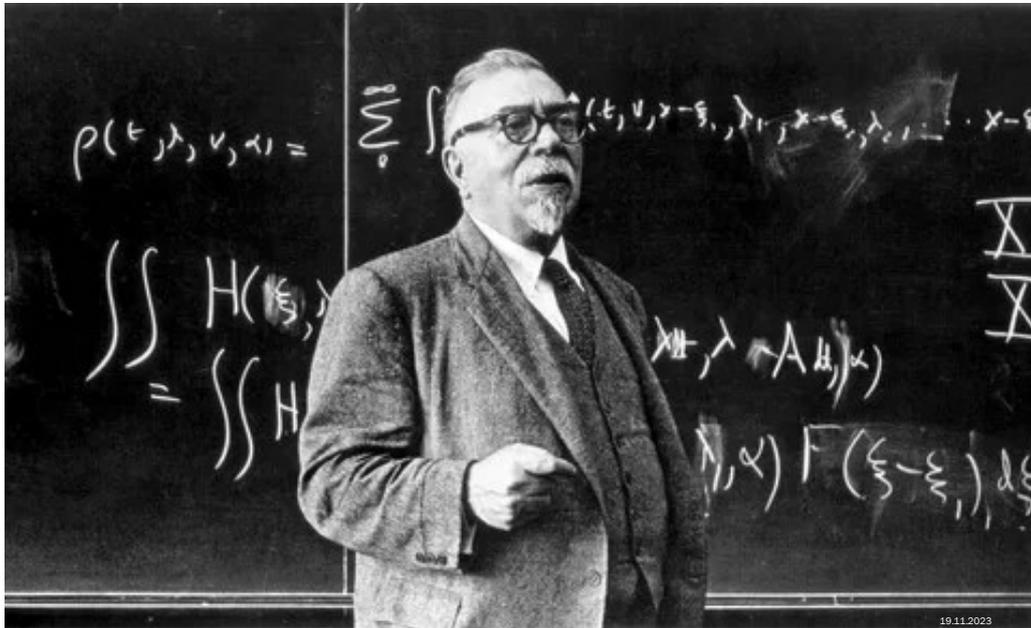
Информация — это продукт взаимодействия данных и адекватных им методов.

— ПОНЯТИЕ

Информация — это продукт взаимодействия данных и адекватных им методов.

Рассмотрим пример, в свое время использованный Норбертом Винером для того, чтобы показать, как информация отдельных членов популяции становится информацией общества.

Норберт Виннер



В своей работе «Кибернетика» Виннер пишет:

Допустим, я нахожусь в лесах вдвоем со смышленным дикарем, который не может говорить на моем языке и на языке, которого я тоже не могу говорить. Даже без какого-либо условного языка знаков, известного нам обоим, я могу многое узнать от него. Мне нужно лишь быть особо внимательным в те моменты, когда он обнаруживает признаки волнения или интереса. Тогда я должен посмотреть вокруг, особенно в направлении его взгляда, и запомнить все, что я увижу и услышу. Не пройдет много времени, как я открою, какие предметы представляются важными для него, — не потому, что он сообщил мне о них словами, но потому, что я сам их заметил. Иначе говоря, сигнал, лишенный внутреннего содержания, может приобрести для моего спутника смысл по тому, что наблюдает он в данный момент, и может приобрести для меня смысл по тому, что наблюдаю я в данный момент. Способность дикаря замечать моменты моего особенно активного внимания сама по себе образует язык, возможности которого столь же разнообразны, как и диапазон впечатлений, доступных нам обоим.

Анализируя этот пример, мы видим, что здесь речь идет о данных и методах. Прежде всего, здесь автор прямо говорит о целой группе методов, связанных с наблюдением и анализом, и даже приводит вариант конкретного алгоритма, адекватного рамкам его гипотетического эксперимента (посмотреть, запомнить, открыть...). Автор неоднократно подчеркивает

требование адекватности метода (дикарь должен быть смысленным, а наблюдатель должен быть особо внимательным), без которого информация может и не образоваться.

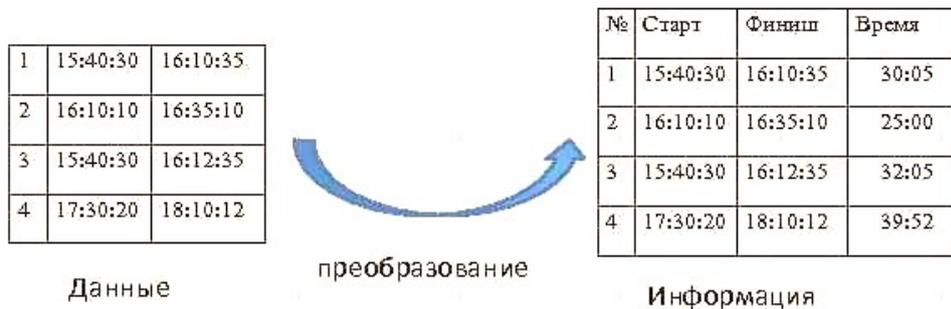


Диалектическое единство данных и методов в информационном процессе

Рассмотрим данное выше определение информации и обратим внимание на следующие обстоятельства.

Данные и информация

Данные – это зарегистрированные сигналы.



- **Динамический характер информации:** информация существует только в момент протекания информационного процесса
- **Адекватность используемых методов:** метод преобразования данных должен соответствовать обрабатываемым данным
- **Диалектический характер взаимодействия данных и методов:** данные - объективны, информация - субъективна

1. **Динамический характер информации.** Информация не является статичным объектом — она динамически меняется и существует только в момент взаимодействия данных и методов. Все прочее время она пребывает в состоянии данных. Таким образом, информация существует только в момент протекания информационного процесса. Все остальное время она содержится в виде данных.

2. **Требование адекватности методов.** Одни и те же данные могут в момент потребления поставлять разную информацию в зависимости от степени адекватности взаимодействующих с ними методов. Например, для человека, не владеющего китайским языком, письмо, полученное из Пекина, дает только ту информацию, которую можно получить методом наблюдения (количество страниц, цвет и сорт бумаги, наличие незнакомых символов и т. п.). Все это информация, но это не вся информация, заключенная в письме. Использование более адекватных методов даст иную информацию.

3. **Диалектический характер взаимодействия данных и методов.** Обратим внимание на то, что данные являются объективными, поскольку это результат регистрации объективно существовавших сигналов, вызванных изменениями в материальных телах или полях. В то же время, методы являются субъективными. В основе искусственных методов лежат алгоритмы (упорядоченные последовательности команд), составленные и подготовленные людьми (субъектами). В основе естественных методов лежат

биологические свойства субъектов информационного процесса. Таким образом, информация возникает и существует в момент диалектического взаимодействия объективных данных и субъективных методов.

Такой дуализм известен своими проявлениями во многих науках. Так, например, в основе важнейшего вопроса философии о первичности материалистического и идеалистического подходов к теории познания лежит не что иное, как двойственный характер информационного процесса. В обоснованиях обоих подходов нетрудно обнаружить упор либо на объективность данных, либо на субъективность методов. Подход к информации как к объекту особой природы, возникающему в результате диалектического взаимодействия объективных данных с субъективными методами, позволяет во многих случаях снять противоречия, возникающие в философских обоснованиях ряда научных теорий и гипотез.

Свойства информации



Итак, информация является динамическим объектом, образующимся в момент взаимодействия объективных данных и субъективных методов. Как и всякий объект, она обладает свойствами (объекты различимы по своим свойствам). Характерной особенностью информации, отличающей ее от других объектов природы и общества, является отмеченный выше дуализм: на свойства информации влияют как свойства данных, составляющих ее содержательную часть, так и свойства методов, взаимодействующих с

данными в ходе информационного процесса. По окончании процесса свойства информации переносятся на свойства новых данных, то есть свойства методов могут переходить на свойства данных.

Можно привести немало разнообразных свойств информации. Каждая научная дисциплина рассматривает те свойства, которые ей наиболее важны. С точки зрения информатики наиболее важными представляются следующие свойства: объективность, полнота, достоверность, адекватность, доступность и актуальность информации.

Объективность и субъективность информации. Понятие объективности информации является относительным. Это понятно, если учесть, что методы являются субъективными. Более объективной принято считать ту информацию, в которую методы вносят меньший субъективный элемент. Так, например, принято считать, что в результате наблюдения фотоснимка природного объекта или явления образуется более объективная информация, чем в результате наблюдения рисунка того же объекта, выполненного человеком. В ходе информационного процесса степень объективности информации всегда понижается. Это свойство учитывают, например, в правовых дисциплинах, где по-разному обрабатываются показания лиц, не посредственно наблюдавших события или получивших информацию косвенным путем (посредством умозаключений или со слов третьих лиц). В не меньшей степени объективность информации учитывают в исторических дисциплинах. Одни и те же события, зафиксированные в исторических документах разных стран и народов, выглядят совершенно по-разному. У историков имеются свои методы для тестирования объективности исторических данных и создания новых, более достоверных данных путем сопоставления, фильтрации и селекции исходных данных. Обратим внимание на то, что здесь речь идет не о повышении объективности данных, а о повышении их достоверности (это совсем другое свойство).

Полнота информации. Полнота информации во многом характеризует качество информации и определяет достаточность данных для принятия решений или для создания новых данных на основе имеющихся. Чем полнее данные, тем шире диапазон методов, которые можно использовать, тем проще подобрать метод, вносящий минимум погрешностей в ход информационного процесса.

Достоверность информации. Данные возникают в момент регистрации сигналов, но не все сигналы являются «полезными» — всегда присутствует какой-то уровень посторонних сигналов, в результате чего полезные данные сопровождаются определенным уровнем «информационного шума». Если полезный сигнал зарегистрирован более четко, чем посторонние сигналы, достоверность информации может быть более высокой. При увеличении уровня шумов достоверность информации снижается. В этом случае для передачи того же количества информации требуется использовать либо больше данных, либо более сложные методы.

Адекватность информации — это степень соответствия реальному объективному состоянию дела. Неадекватная информация может образовываться при создании новой информации на основе неполных или недостоверных данных. Однако и полные, и достоверные данные могут приводить к созданию неадекватной информации в случае применения к ним неадекватных методов.

Доступность информации — мера возможности получить ту или иную информацию. На степень доступности информации влияют одновременно как доступность данных, так и доступность адекватных методов для их интерпретации. Отсутствие доступа к данным или отсутствие адекватных методов обработки данных приводит к одинаковому результату: информация оказывается недоступной. Отсутствие адекватных методов для работы с данными во многих случаях приводит к применению неадекватных методов, в результате чего образуется неполная, неадекватная или недостоверная информация.

Актуальность информации — это степень соответствия информации текущему моменту времени. Нередко с актуальностью, как и с полнотой, связывают коммерческую ценность информации. Поскольку информационные процессы растянуты во времени, то достоверная и адекватная, но устаревшая информация может приводить к ошибочным решениям. Необходимость поиска (или разработки) адекватного метода для работы с данными может приводить к такой задержке в получении информации, что она становится неактуальной и ненужной. На этом, в частности, основаны многие современные системы шифрования данных с открытым ключом. Лица, не владеющие ключом (методом) для чтения данных, могут заняться поиском ключа, поскольку алгоритм его работы доступен, но продолжительность этого поиска столь велика, что за время работы информация теряет актуальность и, соответственно, связанную с ней практическую ценность.

Данные

Носители данных

Данные — диалектическая составная часть информации. Они представляют собой зарегистрированные сигналы. При этом физический метод регистрации может быть любым: механическое перемещение физических тел, изменение их формы или параметров качества поверхности, изменение электрических, магнитных, оптических характеристик, химического состава и (или) характера химических связей, изменение состояния электронной системы и многое другое. В соответствии с методом регистрации данные могут храниться и транспортироваться на носителях различных видов.

Носители информации



Самым распространенным носителем данных, хотя и не самым экономичным, по-видимому, является бумага. На бумаге данные регистрируются путем изменения оптических характеристик ее поверхности. Изменение оптических свойств (изменение коэффициента отражения поверхности в определенном диапазоне длин волн) используется также в устройствах, осуществляющих запись лазерным лучом на пластмассовых носителях с отражающим покрытием (CD-ROM, DVD-ROM). В качестве носителей, использующих изменение магнитных свойств, можно назвать магнитные ленты и диски. Регистрация данных путем изменения химического состава поверхностных веществ носителя широко используется в фотографии. На биохимическом уровне происходит накопление и передача данных в живой природе.

Носители данных интересуют нас не сами по себе, а постольку, поскольку свойства информации весьма тесно связаны со свойствами ее носителей. Любой носитель можно характеризовать параметром разрешающей способности (количеством данных, записанных в принятой для носителя единице измерения) и динамическим диапазоном (логарифмическим отношением интенсивности амплитуд максимального и минимального регистрируемого сигналов). От этих свойств носителя нередко зависят такие свойства информации, как полнота, доступность и достоверность. Например, мы можем рассчитывать на то, что в базе данных, размещаемой на компакт-диске, проще обеспечить полноту информации, чем в аналогичной по

назначению базе данных, размещенной на гибком магнитном диске, поскольку в первом случае плотность записи данных на единице длины дорожки намного выше. Для обычного потребителя доступность информации в книге заметно выше, чем той же информации на компакт-диске, поскольку не все потребители обладают необходимым оборудованием. И наконец, известно, что визуальный эффект от просмотра слайда в проекторе намного больше, чем от просмотра аналогичной иллюстрации, напечатанной на бумаге, поскольку диапазон яркостных сигналов в проходящем свете на два-три порядка больше, чем в отраженном.

Задача преобразования данных с целью смены носителя относится к одной из важнейших задач информатики. В структуре стоимости вычислительных систем устройства для ввода и вывода данных, работающие с носителями информации, составляют до половины стоимости аппаратных средств.

Операции с данными



В ходе информационного процесса данные преобразуются из одного вида в другой с помощью методов. Обработка данных включает в себя множество различных операций. По мере развития научно-технического прогресса и общего усложнения связей в человеческом обществе трудозатраты на обработку данных неуклонно возрастают. Прежде всего, это связано с постоянным усложнением условий управления производством и

обществом. Вторым фактором, также вызывающим общее увеличение объемов обрабатываемых данных, тоже связан с научно-техническим прогрессом, а именно с быстрыми темпами появления и внедрения новых носителей данных, средств их хранения и доставки.

В структуре возможных операций с данными можно выделить следующие основные:

- *сбор данных* — накопление информации с целью обеспечения достаточной полноты для принятия решений;
- *формализация данных* — приведение данных, поступающих из разных источников, к одинаковой форме, чтобы сделать их сопоставимыми между собой, то есть повысить их уровень доступности;
- *фильтрация данных* — отсеивание «лишних» данных, в которых нет необходимости для принятия решений; при этом должен уменьшаться уровень «шума», а достоверность и адекватность данных должны возрасти;
- *сортировка данных* — упорядочение данных по заданному признаку с целью удобства использования; повышает доступность информации;
- *архивация данных* — организация хранения данных в удобной и легкодоступной форме; служит для снижения экономических затрат по хранению данных и повышает общую надежность информационного процесса в целом;
- *защита данных* — комплекс мер, направленных на предотвращение утраты, воспроизведения и модификации данных;
- *транспортировка данных* — прием и передача (доставка и поставка) данных между удаленными участниками информационного процесса; при этом источник данных в информатике принято называть сервером, а потребителя — клиентом;
- *преобразование данных* — перевод данных из одной формы в другую или из одной структуры в другую.

Преобразование данных часто связано с изменением типа носителя: например, книги можно хранить в обычной бумажной форме, но можно использовать для этого и электронную форму, и микрофото пленку. Необходимость в многократном преобразовании данных возникает также при их транспортировке, особенно если она осуществляется средствами, не предназначенными для транспортировки данного вида данных. В качестве примера можно упомянуть, что для транспортировки цифровых потоков данных по каналам телефонных сетей (которые изначально были ориентированы только на передачу аналоговых сигналов в узком диапазоне частот) необходимо преобразование цифровых данных в некое подобие звуковых сигналов, чем и занимаются специальные устройства — телефонные модемы.

Приведенный здесь список типовых операций с данными далеко не полон. Миллионы людей во всем мире занимаются созданием, обработкой, преобразованием и транспортировкой данных, и на каждом рабочем месте выполняются свои специфические операции, необходимые для управления социальными, экономическими, промышленными, научными и культурными процессами. Полный список возможных операций составить невозможно, да и не нужно. Сейчас нам важен другой вывод: работа с информацией может иметь огромную трудоемкость и ее надо автоматизировать.

Кодирование данных двоичным кодом

Примеры различных систем кодирования

С	О	М	Р	U	T	E	R	
43	4F	4D	50	55	54	45	52	Код ASCII
-. . .	- - -	- -	. - -	-	.	. - .	Код Морзе
••	•• •	•• •	•• ••	•• ••	•• ••	•• •	•• ••	Код Брайля
								Код морской сигнальный

20.11.2023 10

Для автоматизации работы с данными, относящимися к различным типам, очень важно унифицировать их форму представления — для этого обычно используется прием кодирования, то есть выражение данных одного типа через данные другого типа. Естественные человеческие языки — это не что иное, как системы кодирования понятий для выражения мыслей посредством речи. К языкам близко примыкают азбуки (системы кодирования компонентов языка с помощью графических символов). История знает интересные, хотя и безуспешные попытки создания «универсальных» языков и азбук. По-видимому, безуспешность попыток их внедрения связана с тем, что национальные и социальные образования естественным образом понимают, что изменение системы кодирования общественных данных

непрерывно приводит к изменению общественных методов (то есть норм права и морали), а это может быть связано с социальными потрясениями.

Та же проблема универсального средства кодирования достаточно успешно реализуется в отдельных отраслях техники, науки и культуры. В качестве примеров можно привести систему записи математических выражений, телеграфную азбуку, морскую флажковую азбуку, систему Брайля для слепых и многое другое.

Своя система существует и в вычислительной технике — она называется двоичным кодированием и основана на представлении данных последовательностью всего двух знаков: 0 и 1. Эти знаки называются двоичными цифрами, по-английски — binary digit или, сокращенно, bit (бит).

Двоичное кодирование

Один бит

01

Два бита

00 01 10 11

Три бита

000 001 010 011 100 101 110 111

$$N = 2^m$$

Одним битом могут быть выражены два понятия: 0 или 1 (да или нет, черное или белое, истина или ложь и т. п.). Если количество битов увеличить до двух, то уже можно выразить четыре различных понятия:

00 01 10 11

Тремя битами можно закодировать восемь различных значений:

000 001 010 011 100 101 110 111

Увеличивая на единицу количество разрядов в системе двоичного кодирования, мы увеличиваем в два раза количество значений, которое может быть выражено в данной системе, то есть общая формула имеет вид:

$$N = 2m,$$

где N — количество независимых кодируемых значений; m — разрядность двоичного кодирования, принятая в данной системе.

Кодирование целых и действительных чисел

Целые числа кодируются двоичным кодом достаточно просто — достаточно взять целое число и делить его пополам до тех пор, пока в остатке не образуется ноль или единица. Совокупность остатков от каждого деления, записанная справа налево вместе с последним остатком, и образует двоичный аналог десятичного числа.

Кодирование целых и действительных чисел

$$19 : 2 = 9 \text{ остаток } 1$$

$$9 : 2 = 4 \text{ остаток } 1$$

$$4 : 2 = 2 \text{ остаток } 0$$

$$2 : 2 = 1 \text{ остаток } 0$$

$$1 : 2 = 0 \text{ остаток } 1$$

$$19_{10} = 10011_2.$$

$$19 : 2 = 9 \text{ остаток } 1$$

$$9 : 2 = 4 \text{ остаток } 1$$

$$4 : 2 = 2 \text{ остаток } 0$$

$$2 : 2 = 1 \text{ остаток } 0$$

$$1 : 2 = 0 \text{ остаток } 1$$

Таким образом, $19_{10} = 10011_2$.

Обратный перевод

Кодирование целых и действительных чисел

$$19_{10} = 10011_2$$

$$19 = 10 * 10^1 + 9 * 10^0$$

$$\begin{aligned} 1 * 2^4 + 0 * 2^3 + 0 * 2^2 + 1 * 2^1 + 1 * 2^0 &= \\ = 16 + 0 + 0 + 2 + 1 &= 19 \end{aligned}$$

20.11.2023 12

$$19 = 10 * 10^1 + 9 * 10^0$$

$$1 * 2^4 + 0 * 2^3 + 0 * 2^2 + 1 * 2^1 + 1 * 2^0 = 16 + 0 + 0 + 2 + 1 = 19$$

Для кодирования целых чисел от 0 до 255 достаточно иметь 8 разрядов двоичного кода (8 бит). Шестнадцать бит позволяют закодировать целые числа от 0 до 65 535, а 24 бита — уже более 16,5 млн разных значений.

Для кодирования действительных чисел используют 80-разрядное кодирование.

При этом число предварительно преобразуется в нормализованную форму:

Нормализация действительных чисел

$$3,1415926 = 0,31415926 \cdot 10^1$$

$$300\ 000 = 0,3 \cdot 10^6$$

$$123\ 456\ 789 = 0,123456789 \cdot 10^{10}$$

20.11.2023 14

$$3,1415926 = 0,31415926 \cdot 10^1$$

$$300\ 000 = 0,3 \cdot 10^6$$

$$123\ 456\ 789 = 0,123456789 \cdot 10^{10}$$

Первая часть числа называется мантиссой, а вторая — характеристикой. Большую часть из 80 бит отводят для хранения мантиссы (вместе со знаком) и некоторое фиксированное количество разрядов отводят для хранения характеристики (тоже со знаком).

Кодирование текстовых данных

Если каждому символу алфавита сопоставить определенное целое число (например, порядковый номер), то с помощью двоичного кода можно кодировать и текстовую информацию. Восемью двоичных разрядов достаточно для кодирования 256 различных символов. Этого хватит, чтобы выразить различными комбинациями восьми битов все символы английского и русского языков, как строчные, так и прописные, а также знаки препинания, символы основных арифметических действий и некоторые общепринятые специальные символы, например символ «§».

Технически это выглядит очень просто, однако всегда существовали достаточно веские организационные сложности. В первые годы развития вычислительной техники они были связаны с отсутствием необходимых стандартов, а в настоящее время вызваны, наоборот, избытком одновременно

действующих и противоречивых стандартов. Для того чтобы весь мир одинаково кодировал текстовые данные, нужны единые таблицы кодирования, а это пока невозможно из-за противоречий между символами национальных алфавитов, а также противоречий корпоративного характера.

Для английского языка, захватившего де-факто нишу международного средства общения, противоречия уже сняты. 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 закрепляет значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 до 255.

Первые 32 кода базовой таблицы, начиная с нулевого, отданы производителям аппаратных средств (в первую очередь производителям компьютеров и печатающих устройств). В этой области размещаются так называемые управляющие коды, которым не соответствуют никакие символы языков, и, соответственно, эти коды не выводятся ни на экран, ни на устройства печати, но ими можно управлять тем, как производится вывод прочих данных.

Начиная с кода 32 по код 127 размещены коды символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов.

Базовая таблица кодировки ASCII

32	пробел	48	0	64	@	80	P	96	`	112	p
33	!	49	1	65	A	81	Q	97	a	113	q
34	"	50	2	66	B	82	R	98	b	114	r
35	#	51	3	67	C	83	S	99	c	115	s
36	\$	52	4	68	D	84	T	100	d	116	t
37	%	53	5	69	E	85	U	101	e	117	u
38	&	54	6	70	F	86	V	102	f	118	v
39	'	55	7	71	G	87	W	103	g	119	w
40	(56	8	72	H	88	X	104	h	120	x
41)	57	9	73	I	89	Y	105	i	121	y
42	*	58	:	74	J	90	Z	106	j	122	z
43	+	59	;	75	K	91	[107	k	123	{
44	,	60	<	76	L	92	\	108	l	124	
45	-	61	=	77	M	93]	109	m	125	}
46	.	62	>	78	N	94	^	110	n	126	~
47	/	63	?	79	O	95	_	111	o	127	

Аналогичные системы кодирования текстовых данных были разработаны и в других странах. Так, например, в СССР в этой области действовала система кодирования КОИ-7 (код обмена информацией, семизначный). Однако поддержка производителей оборудования и программ вывела американский код ASCII на уровень международного стандарта, и национальным системам кодирования пришлось «отступить» во вторую, расширенную часть системы кодирования, определяющую значения кодов со 128 по 255. Отсутствие единого стандарта в этой области привело к множественности одновременно действующих кодировок. Только в России можно указать три действующих стандарта кодировки и еще два устаревших.

128 Ъ	144 ђ	160	176 •	192 А	208 Р	224 а	240 р
129 Ѓ	145 ‘	161 Ў	177 ±	193 Б	209 С	225 б	241 с
130 ‚	146 ‚	162 ў	178	194 В	210 Т	226 в	242 т
131 ƒ	147 “	163 Ј	179 i	195 Г	211 У	227 г	243 у
132 ”	148 ”	164 ђ	180 г	196 Д	212 Ф	228 д	244 ф
133 ...	149 •	165 Г	181 μ	197 Е	213 Х	229 е	245 х
134 †	150 —	166 Г	182 ¶	198 Ж	214 Ц	230 ж	246 ц
135 ‡	151 —	167 §	183 ·	199 З	215 Ч	231 з	247 ч
136 ´	152 ´	168 Ё	184 ё	200 И	216 Ш	232 и	248 ш
137 ‰	153 ™	169 ©	185 №	201 Й	217 Щ	233 й	249 щ
138 Љ	154 Љ	170 €	186 €	202 К	218 Ъ	234 к	250 ъ
139 ‹	155 ›	171 «	187 »	203 Л	219 Ы	235 л	251 ы
140 Њ	156 Њ	172 ¬	188 j	204 М	220 Ь	236 м	252 ь
141 К	157 к	173 -	189 S	205 Н	221 Э	237 н	253 э
142 Ћ	158 ћ	174 @	190 s	206 О	222 Ю	238 о	254 ю
143 Ў	159 џ	175 ĩ	191 ĩ	207 П	223 Я	239 п	255 я

Так, например, кодировка символов русского языка, известная как кодировка Windows-1251, была введена «извне» — компанией Microsoft, но, учитывая широкое распространение операционных систем и других продуктов этой компании в России, она глубоко закрепились и нашла широкое распространение. Эта кодировка используется на большинстве локальных компьютеров, работающих на платформе Windows.

Кодировка КОИ8

128	144 ▯	160 —	176 †	192 ю	208 п	224 Ю	240 П
129	145 ▯	161 Ё	177 †	193 а	209 я	225 А	241 Я
130	146 ▯	162 г	178 †	194 б	210 р	226 Б	242 Р
131	147	163 ё	179 Ё	195 ц	211 с	227 Ц	243 С
132	148	164 г	180 †	196 д	212 т	228 Д	244 Т
133	149	165 г	181 †	197 е	213 у	229 Е	245 У
134	150 √	166	182 †	198 ф	214 ж	230 Ф	246 Ж
135	151 »	167	183 †	199 г	215 в	231 Г	247 В
136	152 <	168	184 †	200 х	216 ь	232 Х	248 Ь
137	153 >	169	185 †	201 и	217 ы	233 И	249 Ы
138	154	170	186 †	202 й	218 з	234 Й	250 З
139	155	171	187 †	203 к	219 ш	235 К	251 Ш
140	156 •	172	188 †	204 л	220 э	236 Л	252 Э
141	157 ²	173	189 †	205 м	221 щ	237 М	253 Щ
142	158 ·	174	190 †	206 н	222 ч	238 Н	254 Ч
143	159 ÷	175 †	191 ё	207 о	223 ь	239 О	255 Ъ

Другая распространенная кодировка носит название КОИ-8 (код обмена информацией, восьмизначный) — ее происхождение относится ко временам действия Совета Экономической Взаимопомощи государств Восточной

Европы. Сегодня кодировка КОИ-8 имеет широкое распространение в компьютерных сетях на территории России и в российском секторе Интернета.

Кодировка ISO

В ISO не определены	160	176 А	192 Р	208 а	224 р	240 №
	161 Ё	177 Б	193 С	209 б	225 с	241 ё
	162 Ъ	178 В	194 Т	210 в	226 т	242 ъ
	163 Ѓ	179 Г	195 У	211 г	227 у	243 ѓ
	164 Є	180 Д	196 Ф	212 д	228 ф	244 є
	165 S	181 Е	197 Х	213 е	229 х	245 s
	166 I	182 Ж	198 Ц	214 ж	230 ц	246 I
	167 I	183 З	199 Ч	215 з	231 ч	247 I
	168 J	184 И	200 Ш	216 и	232 ш	248 j
	169 Љ	185 Й	201 Щ	217 й	233 щ	249 ъ
	170 Њ	186 К	202 Ъ	218 к	234 ъ	250 ъ
	171 Ћ	187 Л	203 Ы	219 л	235 ы	251 ћ
	172 K	188 М	204 Ь	220 м	236 ь	252 k
	173 -	189 Н	205 Э	221 н	237 э	253 §
	174 Ÿ	190 О	206 Ю	222 о	238 ю	254 Ÿ
	175 Ÿ	191 П	207 Я	223 п	239 я	255 Ÿ

20.11.2003 18

Международный стандарт, в котором предусмотрена кодировка символов русского алфавита, носит название кодировки ISO (International Standard Organization — Международный институт стандартизации). На практике данная кодировка используется редко (таблица 1.4).

ГОСТ-альтернативная кодировка

128	А	144	Р	160	а	176	⌘	192	⌘	208	⌘	224	р	240	Ё
129	Б	145	С	161	б	177	⌘	193	Г	209	⌘	225	с	241	ё
130	В	146	Т	162	в	178	⌘	194	⌘	210	⌘	226	т	242	€
131	Г	147	У	163	г	179	⌘	195	⌘	211	⌘	227	у	243	е
132	Д	148	Ф	164	д	180	⌘	196	—	212	⌘	228	ф	244	і
133	Е	149	Х	165	е	181	⌘	197	+	213	⌘	229	х	245	і
134	Ж	150	Ц	166	ж	182	⌘	198	⌘	214	⌘	230	ц	246	Ў
135	З	151	Ч	167	з	183	⌘	199	⌘	215	⌘	231	ч	247	ў
136	И	152	Ш	168	и	184	⌘	200	⌘	216	⌘	232	ш	248	·
137	Й	153	Щ	169	й	185	⌘	201	⌘	217	⌘	233	щ	249	·
138	К	154	Ъ	170	к	186	⌘	202	⌘	218	⌘	234	ъ	250	·
139	Л	155	Ы	171	л	187	⌘	203	⌘	219	■	235	ы	251	√
140	М	156	Ь	172	м	188	⌘	204	⌘	220	■	236	ь	252	№
141	Н	157	Э	173	н	189	⌘	205	—	221	■	237	э	253	⌘
142	О	158	Ю	174	о	190	⌘	206	⌘	222	■	238	ю	254	■
143	П	159	Я	175	п	191	⌘	207	⌘	223	■	239	я	255	

20.11.2002 49

На компьютерах, работающих в операционных системах MS-DOS, могут действовать еще две кодировки (кодировка ГОСТ и кодировка ГОСТ-альтернативная). Первая из них считалась устаревшей даже в первые годы появления персональной вычислительной техники, но вторая используется и по сей день.

В связи с изобилием систем кодирования текстовых данных, действующих в России, возникает задача межсистемного преобразования данных — это одна из распространенных задач информатики.

Универсальная система кодирования текстовых данных

Если проанализировать организационные трудности, связанные с созданием единой системы кодирования текстовых данных, то можно прийти к выводу, что они вызваны ограниченным набором кодов (256). В то же время очевидно, что если, например, кодировать символы не восьмиразрядными двоичными числами, а числами с большим количеством разрядов, то и диапазон возможных значений кодов станет намного больше. Такая система, основанная на 16-разрядном кодировании символов, получила название универсальной — UNICODE. Шестнадцать разрядов позволяют обеспечить

уникальные коды для 65 536 различных символов — этого поля достаточно для размещения в одной таблице символов большинства языков планеты.

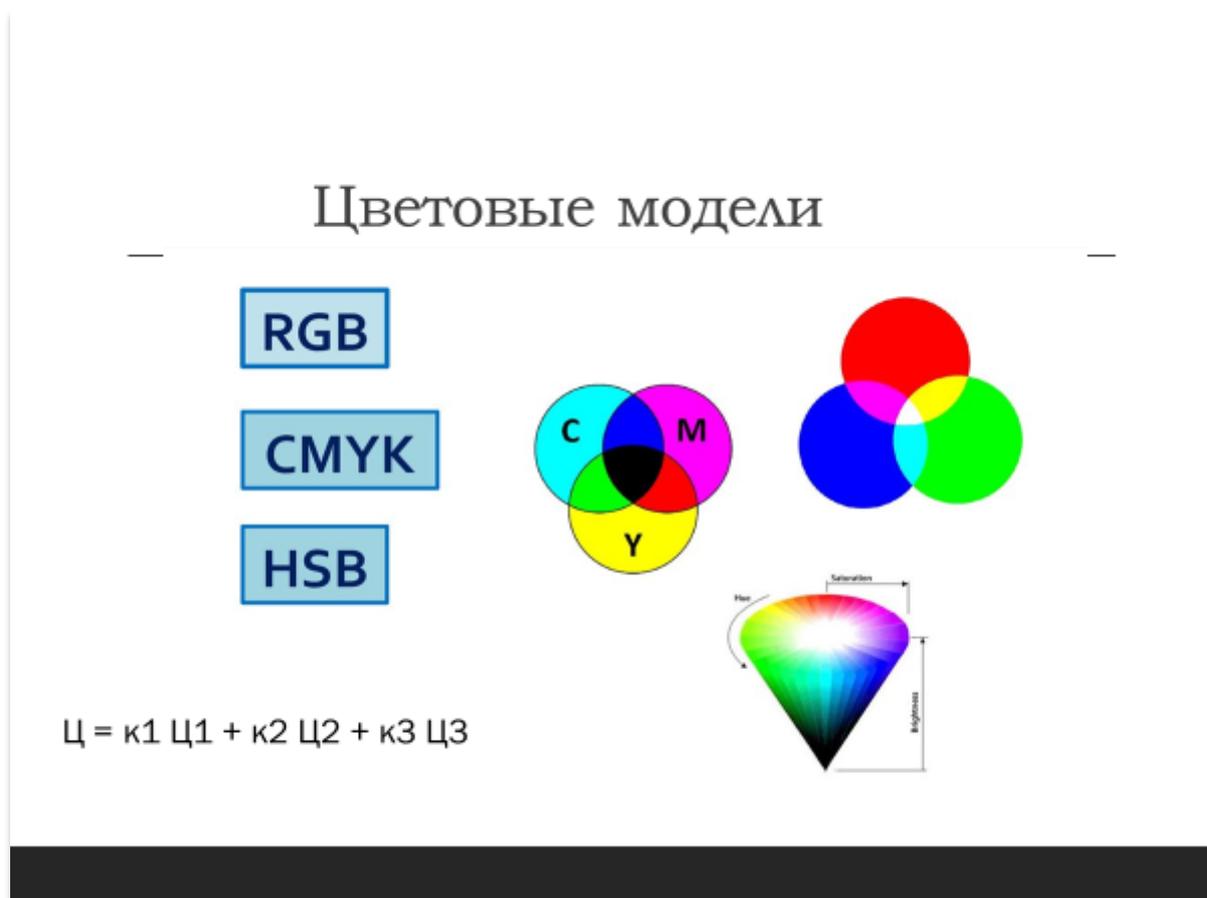
Несмотря на тривиальную очевидность такого подхода, простой механический переход на данную систему долгое время сдерживался из-за недостаточных ресурсов средств вычислительной техники (в системе кодирования UNICODE все текстовые документы автоматически становятся вдвое длиннее). Во второй половине 90-х годов технические средства достигли необходимого уровня обеспеченности ресурсами, и сегодня мы наблюдаем постепенный перевод документов и программных средств на универсальную систему кодирования. Для индивидуальных пользователей это еще больше добавило забот по согласованию документов, выполненных в разных системах кодирования, с программными средствами, но это надо понимать как трудности переходного периода.

Кодирование графических данных

Если рассмотреть с помощью увеличительного стекла черно-белое графическое изображение, напечатанное в газете или книге, то можно увидеть, что оно состоит из мельчайших точек, образующих характерный узор, называемый растром.



Поскольку линейные координаты и индивидуальные свойства каждой точки (яркость) можно выразить с помощью целых чисел, то можно сказать, что растровое кодирование позволяет использовать двоичный код для представления графических данных. Общепринятым на сегодняшний день считается представление черно-белых иллюстраций в виде комбинации точек с 256 градациями серого цвета, и, таким образом, для кодирования яркости любой точки обычно достаточно восьмиразрядного двоичного числа.



Для кодирования цветных графических изображений применяется принцип декомпозиции произвольного цвета на основные составляющие. В качестве таких составляющих используют три основных цвета: красный (Red, R), зеленый (Green, G) и синий (Blue, B). На практике считается (хотя теоретически это не совсем так), что любой цвет, видимый человеческим глазом, можно получить путем механического смешения этих трех основных цветов. Такая система кодирования называется системой RGB по первым буквам названий основных цветов.

Если для кодирования яркости каждой из основных составляющих использовать по 256 значений (восемь двоичных разрядов), как это принято для полутоновых черно-белых изображений, то на кодирование цвета одной точки надо затратить 24 разряда. При этом система кодирования обеспечивает однозначное определение 16,5 млн различных цветов, что на самом деле близко к чувствительности человеческого глаза. Режим представления

цветной графики с использованием 24 двоичных разрядов называется полноцветным (True Color).

Каждому из основных цветов можно поставить в соответствие дополнительный цвет, то есть цвет, дополняющий основной цвет до белого. Нетрудно заметить, что для любого из основных цветов дополнительным будет цвет, образованный суммой пары остальных основных цветов. Соответственно, дополнительными цветами являются: голубой (Cyan, С), пурпурный (Magenta, М) и желтый (Yellow, У). Принцип декомпозиции произвольного цвета на составляющие компоненты можно применять не только для основных цветов, но и для дополнительных, то есть любой цвет можно представить в виде суммы голубой, пурпурной и желтой составляющей.

Такой метод кодирования цвета принят в полиграфии, но в полиграфии используется еще и четвертая краска — черная (Black, К). Поэтому данная система кодирования обозначается четырьмя буквами CMYK (черный цвет обозначается буквой К, потому что буква В уже занята синим цветом), и для представления цветной графики в этой системе надо иметь 32 двоичных разряда. Такой режим тоже называется полноцветным (True Color).

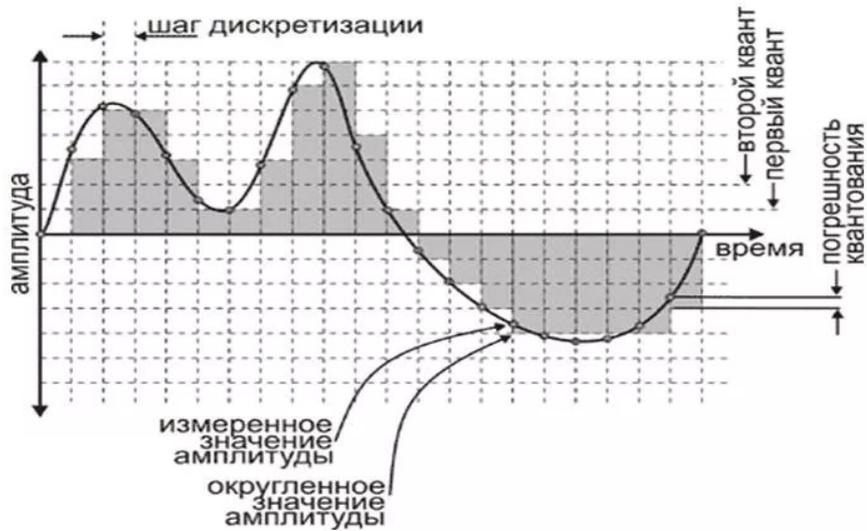
Если уменьшить количество двоичных разрядов, используемых для кодирования цвета каждой точки, то можно сократить объем данных, но при этом диапазон кодируемых цветов заметно сокращается. Кодирование цветной графики 16-разрядными двоичными числами называется режимом High Color.

При кодировании информации о цвете с помощью восьми бит данных можно передать только 256 цветовых оттенков. Такой метод кодирования цвета называется индексным. Смысл названия в том, что, поскольку 256 значений совершенно недостаточно, чтобы передать весь диапазон цветов, доступный человеческому глазу, код каждой точки раstra выражает не цвет сам по себе, а только его номер (индекс) в некоей справочной таблице, называемой палитрой. Разумеется, эта палитра должна прикладываться к графическим данным — без нее нельзя воспользоваться методами воспроизведения информации на экране или бумаге (то есть, воспользоваться, конечно, можно, но из-за неполноты данных полученная информация не будет адекватной:

листва на деревьях может оказаться красной, а небо — зеленым).

Кодирование звуковой информации

Кодирование звуковой информации

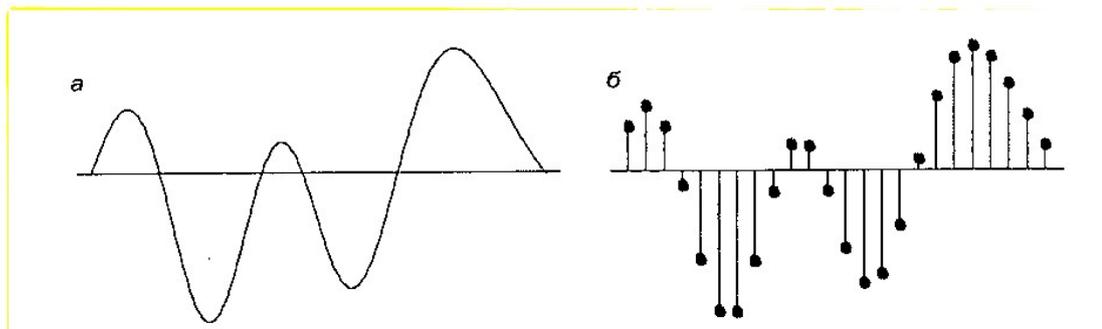


Дискретизация звука

Приемы и методы работы со звуковой информацией пришли в вычислительную технику наиболее поздно. К тому же, в отличие от числовых, текстовых и графических данных, у звукозаписей не было столь же длительной и проверенной истории кодирования. В итоге методы кодирования звуковой информации двоичным кодом далёки от стандартизации. Множество отдельных компаний разработали свои корпоративные стандарты, но если говорить обобщенно, то можно выделить два основных направления.

Кодирование звуковой информации

- **Метод FM (Frequency Modulation).**



Преобразование звукового сигнала в дискретный сигнал:
а – звуковой сигнал на входе АЦП (аналогово-цифровые преобразователи); б – дискретный сигнал на выходе АЦП

20.11.2023 23

Метод FM (Frequency Modulation) основан на том, что теоретически любой сложный звук можно разложить на последовательность простейших гармонических сигналов разных частот, каждый из которых представляет собой правильную синусоиду, а следовательно, может быть описан числовыми параметрами, то есть кодом. В природе звуковые сигналы имеют непрерывный спектр, то есть являются аналоговыми. Их разложение в гармонические ряды и представление в виде дискретных цифровых сигналов выполняют специальные устройства — аналогово-цифровые преобразователи (АЦП). Обратное преобразование для воспроизведения звука, закодированного числовым кодом, выполняют цифро-аналоговые преобразователи (ЦАП). При таких преобразованиях неизбежны потери информации, связанные с методом кодирования, поэтому качество звукозаписи обычно получается не вполне удовлетворительным и соответствует качеству звучания простейших электромузыкальных инструментов с окрасом, характерным для электронной музыки. В то же время, данный метод кодирования обеспечивает весьма компактный код, и потому он нашел применение еще в те годы, когда ресурсы средств вычислительной техники были явно недостаточны.

Кодирование звуковой информации wave-table



Метод таблично-волнового (Wave-Table) синтеза лучше соответствует современному уровню развития техники. Если говорить упрощенно, то можно сказать, что где-то в заранее подготовленных таблицах хранятся образцы звуков для множества различных музыкальных инструментов (хотя не только для них). В технике такие образцы называют сэмплами. Числовые коды выражают тип инструмента, номер его модели, высоту тона, продолжительность и интенсивность звука, динамику его изменения, некоторые параметры среды, в которой происходит звучание, а также прочие параметры, характеризующие особенности звука. Поскольку в качестве образцов используются «реальные» звуки, то качество звука, полученного в результате синтеза, получается очень высоким и приближается к качеству звучания реальных музыкальных инструментов.