

Лекция 3.

Количество информации

- Теория информации: понятие информации и ее измерение; количество и качество информации; единицы измерения информации; информация и энтропия; сообщения и сигналы; кодирование и квантование сигналов; основные виды обработки данных; обработка аналоговой и цифровой информации.
- **Понятие информации**
- Термин информация происходит от латинского слова *informatio*, что означает сведения, разъяснения, изложение.
- Информация — это настолько общее и глубокое понятие, что его нельзя объяснить одной фразой. В это слово вкладывается различный смысл в технике, науке и в житейских ситуациях.
- В обиходе информацией называют любые данные или сведения, которые кого-либо интересуют. Например, сообщение о каких-либо событиях, о чьей-либо деятельности и т.п. Информировать в этом смысле означает сообщить нечто, неизвестное раньше.
- Информация — сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые воспринимают информационные системы (живые организмы, управляющие машины и др.) в процессе жизнедеятельности и работы. Одно и то же информационное сообщение (статья в газете, объявление, письмо, телеграмма, справка, рассказ, чертёж, радиопередача и т.п.) может содержать разное количество информации для разных людей — в зависимости от их предшествующих знаний, от уровня понимания этого сообщения и интереса к нему.
- Информация есть характеристика не сообщения, а соотношения между сообщением и его потребителем. Без наличия потребителя, хотя бы потенциального, говорить об информации бессмысленно. В случаях, когда говорят об автоматизированной работе с информацией посредством каких-либо технических устройств, обычно в первую очередь интересуются не содержанием сообщения, а тем, сколько символов это сообщение содержит.

Информация может существовать в самых разнообразных формах:

- 1) в виде текстов, рисунков, чертежей, фотографий;
- 2) в виде световых или звуковых сигналов;
- 3) в виде радиоволн;
- 4) в виде электрических и нервных импульсов;
- 5) в виде магнитных записей;
- 6) в виде жестов и мимики;
- 7) в виде запахов и вкусовых ощущений;
- 8) в виде хромосом, посредством которых передаются по наследству признаки и свойства организмов и т.д.

- Предметы, процессы, явления материального или нематериального свойства, рассматриваемые с точки зрения их информационных свойств, называются информационными объектами.
- Информацию можно:
- создавать; передавать; воспринимать; использовать; запоминать; принимать; копировать; формализовать; распространять; преобразовывать; комбинировать; обрабатывать; делить на части; упрощать; собирать; хранить; искать; измерять; разрушать и др.

Все эти процессы, связанные с определенными операциями над информацией, называются информационными процессами.

Сообщения и сигналы

- **Сигналы и данные.**
- 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 энергетического обмена 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.
- **Данные и методы.**
- 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.
- Прослушивая передачу радиостанции на незнакомом языке, мы получаем данные, но не получаем информацию в связи с тем, что не владеем 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.
- **Сигналы.**
- Информация передается в виде сообщений от некоторого источника информации к её приёмнику посредством канала связи между ними. Источник посылает передаваемое сообщение, которое кодируется в передаваемый сигнал. Этот сигнал посылается по каналу связи. В результате в приёмнике появляется принимаемый сигнал, который декодируется и становится принимаемым сообщением.

- Примеры:
- Живое существо своими органами чувств (глаз, ухо, кожа, язык и т.д.):
- 1) воспринимает информацию из внешнего мира;
- 2) перерабатывает её в определенную последовательность нервных импульсов;
- 3) передает импульсы по нервным волокнам;
- 4) хранит в памяти в виде состояния нейронных структур мозга;
- 5) воспроизводит в виде звуковых сигналов;
- 6) движений и т.п.;
- 7) использует в процессе своей жизнедеятельности.
- Передача информации по каналам связи часто сопровождается воздействием помех, вызывающих искажение и потерю информации.
- **Данные**
- 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.
- 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). В качестве носителей, использующих изменение магнитных свойств, можно назвать магнитные ленты и диски.
- В ходе информационного процесса данные преобразуются из одного вида в другой с помощью методов. Обработка данных включает в себя множество различных операций.
- **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 микрофотопленку.

- Приведенный здесь список типовых операций с данными далеко не полон.
- Измерение количества информации, единицы измерения информации
- **Какое количество информации содержится, к примеру, в тексте романа Война и мир, во фресках Рафаэля или в генетическом коде человека?** Ответа на эти вопросы наука не даёт и, по всей вероятности, даст не скоро.
- **А возможно ли объективно измерить количество информации?** Важнейшим результатом теории информации является вывод:
- В определенных, весьма широких условиях можно пренебречь качественными особенностями информации, выразить её количество числом, а также сравнить количество информации, содержащейся в различных группах данных. В настоящее время получили распространение подходы к определению понятия количества информации, основанные на том, что информацию, содержащуюся в сообщении, можно нестрого трактовать в смысле её новизны или, иначе, уменьшения неопределённости наших знаний об объекте.
- Так, американский инженер Р. Хартли (1928 г.) процесс получения информации рассматривает как выбор одного сообщения из конечного, наперёд заданного множества, из N равновероятных сообщений, а количество информации I , содержащееся в выбранном сообщении, определяет как двоичный логарифм N .
- **Формула Хартли: $I = \log_2 N$.**
- Допустим, нужно угадать одно число из набора чисел от единицы до ста. По формуле Хартли можно вычислить, какое количество информации для этого требуется: $I = \log_2 100 \approx 6,644$. То есть сообщение о верно угаданном числе содержит количество информации, приблизительно равное 6,644 единиц информации.
- Приведем другие примеры равновероятных сообщений:
- 1) при бросании монеты: выпала решка, выпал орел;
- 2) на странице книги: количество букв чётное, количество букв нечётное.
- Определим теперь, являются ли равновероятными сообщения первая выйдет из дверей здания женщина и первым выйдет из дверей здания мужчина. Однозначно ответить на этот вопрос нельзя. Все зависит от того, о каком именно здании идет речь. Если это, например, станция метро, то вероятность выйти из дверей первым одинакова для мужчины и женщины, а если это военная казарма, то для мужчины эта вероятность значительно выше, чем для женщины.
- Для задач такого рода американский учёный Клод Шеннон предложил в 1948 г. другую формулу определения количества информации, учитывающую возможную неодинаковую вероятность сообщений в наборе.
- **Формула Шеннона: $I = - (p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_N \log_2 p_N)$,**
- где p_i — вероятность того, что именно i -е сообщение выделено в наборе из N сообщений. Легко заметить, что если вероятности p_1, \dots, p_N равны, то каждая из них равна $1/N$, и формула Шеннона превращается в формулу Хартли.
- Помимо двух рассмотренных подходов к определению количества информации, существуют и другие. Важно помнить, что любые теоретические результаты применимы лишь к определённому кругу случаев, очерченному первоначальными допущениями.
- В качестве единицы информации условились принять один бит (англ. bit — binary, digit — двоичная цифра).
- Бит в теории информации — количество информации, необходимое для различения двух равновероятных сообщений. А в вычислительной технике битом называют наименьшую порцию памяти, необходимую для хранения одного из двух знаков 0 и 1, используемых для внутримашинного представления данных и команд. Бит — слишком мелкая единица измерения. На практике чаще применяется более крупная единица — байт, равная восьми битам. Именно восемь битов требуется для того, чтобы закодировать любой из 256 символов алфавита клавиатуры компьютера ($256=2^8$).
- Широко используются также ещё более крупные производные единицы информации:
- 1 бит (binary digit – двоичное число) = 0 или 1,
- 1 байт 8 бит,
- 1 килобайт (1Кб) = 213 бит, 1 мегабайт (1Мб) = 223 бит, 1 гигабайт (1Гб) = 233 бит, 1 терабайт (1Тб) = 243 бит, 1 петабайт (1Пб) = 253 бит,
- 1 эксабайт (1Эб) = 263 бит.
- За единицу информации можно было бы выбрать количество информации, необходимое для различения, например, десяти равновероятных сообщений. Это будет не двоичная (бит), а десятичная (дит) единица информации.

Качество 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 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, непосредственно наблюдавших события или получивших информацию косвенным путем (посредством умозаключений или со слов третьих лиц).
- **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 приводят 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.
- **Информация и энтропия**
- Энтропия (от греч. entropia – поворот, превращение) – мера неупорядоченности больших систем. Впервые понятие энтропия введено в XIX в. в результате анализа работы тепловых машин, где энтропия характеризует ту часть энергии, которая рассеивается в пространстве, не совершая полезной работы (отсюда определение: энтропия – мера обесценивания энергии). В XX в. понятие «энтропия» оказалось плодотворным для исследования биосистем, а также процессов передачи и обработки информации. Эволюция в целом и развитие каждого организма происходит благодаря тому, что биосистемы, будучи открытыми, питаются энергией из окружающего мира. Но при этом биопроцессы протекают таким образом, что связанные с ними «производство энтропии» минимально. Это служит важным руководящим принципом и при разработке современных технологических процессов, при проектировании технических систем. Количественная мера информации формально совпадает с «отрицательно определенной» энтропией. Но глубокое понимание соответствия энтропии физической и информационной остается одной из кардинальных недостаточно исследованных проблем современной науки. Ее решение послужит одним из важных факторов становления нового научно-технического мышления.
- Впервые понятия энтропия и информация связал К.Шеннон в 1948 г. С его подачи энтропия стала использоваться как мера полезной информации в процессах передачи сигналов по проводам. Следует подчеркнуть, что под информацией Шеннон понимал сигналы нужные, полезные для получателя. Неполезные сигналы, с точки зрения Шеннона, это шум, помехи. Если сигнал на выходе канала связи является точной копией сигнала на входе то, с точки зрения теории информации, это означает отсутствие энтропии. Отсутствие шума означает максимум информации. Взаимосвязь энтропии и информации нашло отражение в формуле:
- $H + I = 0$,
- где H – энтропия, I – информация.