

Тема №1

"ОСНОВНЫЕ ПОНЯТИЯ ИНФОРМАТИКИ И ИНФОРМАЦИИ"

УЧЕБНЫЕ ВОПРОСЫ:

- 1. ОСНОВНЫЕ ПОНЯТИЯ ИНФОРМАТИКИ.**
- 2. ХАРАКТЕРИСТИКИ ИНФОРМАЦИИ.**
- 3. ХРАНЕНИЕ И ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ. ЕДИНИЦЫ
ИЗМЕРЕНИЯ**

ОСНОВНЫЕ ПОНЯТИЯ ИНФОРМАТИКИ

Деятельность по сбору, хранению и распространению документов называется **документалистика**. С годами ее поле деятельности расширилось. Она превратилась в научную дисциплину по сбору, распространению, переработки и использованию информации.

С появлением ЭВМ и ростом масштабов их применения из документалистики возникла новая отрасль знаний, получившая в настоящее время наименование **“Информатика”**.

Что же такое **информатика**? В отечественной и научной литературе пока еще не сложилось окончательное определение этого понятия. Авторы дают различные формулировки. Необходимо указать на два важных момента, лежащих в основе этого термина.

Во - первых, понятие **“информатика”** применяется для характеристики человеческой деятельности, которая охватывает процессы сбора, получения, хранения, поиска, переработки, передачи и использования информации.

Информатика является самостоятельной отраслью народного хозяйства со своим обслуживающим персоналом и производственными фондами, прежде всего - средствами вычислительной техники, необходимыми для обработки информации. Дальнейшее развитие **информатики**, как отрасли во многом будет определяться совершенствованием элементарной базы и программного обеспечения используемых технических средств.

Во - вторых, под термином **“информатика”** понимается фундаментальная научная дисциплина, которая призвана изучать структуру и свойство информации, а также закономерности ее обработки.

Предметом исследования науки информатики являются **процессы сбора, передачи, обработки, хранения и использование информации**.

В последние годы быстро увеличивается число электронно - вычислительных машин и систем, возрастает их производительность и объемы хранимой информации.

Информатика ныне становится основой современного общества. Она тесно связана с техникой, экономикой, политикой и оказывает непосредственное влияние на их развитие.

Правовая информатика вызвана к жизни потребностями юридической деятельности. Поэтому, по своей природе она является смежной отраслью знаний, в структуре которой имеются элементы юридических наук и наук, с ними взаимодействующих.

От традиционных юридических наук правовая информатика отличается характером своих междисциплинарных связей и функциями. Дело в том, что в рамках этой науки, разрабатываются **общие методологические основы математизации** применительно ко всем юридическим наукам и юридической деятельности в целом. В

силу этого информатика приобретает черты **общей теории решения правовых задач с использованием математических методов** и в этом смысле выполняет функции, схожие с функциями общей теории государства и права в отношении конкретных юридических наук.

Основными задачами правовой информатики являются:

а) исследование и разработка теории правовой информации и методологических основ ее подготовки к машинной обработке. В рамках этого направления изучаются:

- сущность правовой информации и ее значение;
- виды информации, источники и носители;
- методы сбора и анализа правовой информации;
- формализация информации и иные способы её подготовки к машинной обработке;

б) исследование и разработка технологии обработки правовой информации и решения правовых задач с использованием средств вычислительной техники.

Существенными здесь являются следующие вопросы:

- ЭВМ как средство обработки и хранения правовой информации (понятия и виды ЭВМ, их устройство и классификация, принципы создания и использования, сети ЭВМ и др.);
- программное обеспечение работы ЭВМ;
- особенности постановки и решения правовых задач с использованием ЭВМ;

в) исследование и разработка частных методик решения правовых задач на основе использования идей, средств и методов математики, логики, информатики, кибернетики и сопряженных с ними наук;

г) исследование и разработка теоретических основ, принципов построения и практики использования автоматизированных информационных систем и комплексов для решения правовых задач в различных отраслях юридической деятельности.

Предметом изучения и разработки здесь являются следующие классы автоматизированных информационных систем:

- информационно-поисковые;
- информационно-логические;
- системы управления;
- обучающие системы.

ВЫВОДЫ:

Таким образом, информатику можно определить как междисциплинарную отрасль знания о закономерностях и особенностях информационных процессов в сфере юридической деятельности, об их автоматизации, о принципах построения и методиках использования автоматизированных информационных систем, создаваемых для совершенствования и повышения эффективности юридической деятельности и решения частных правовых задач на базе комплексного использования теории и методологии правовых наук, средств и методов математики, информатики и кибернетики.

ОСНОВНЫЕ ПОНЯТИЯ ИНФОРМАЦИИ

Информация (от латинского слова - informatio - разъяснение, изложение) - это сведения или совокупность сведений о предметах, явлениях и процессах окружающего нас мира.

Информация - понятие многократное, поэтому одни ученые выделяют **семантический** (содержательный) аспект, другие - **прагматический** (ценностный), третий дает определение информации по средствам различных свойств материи.

Имеется точка зрения, по которой информацию необходимо отнести к общим научным понятиям, поскольку она используется большинством современных наук и отраслей знаний.

Приведем два определения информации из советского энциклопедического словаря:

- информация это “сведения, передаваемые людьми устным, письменным или др. способом”;
- информация это “общенаучное понятие, включающее обмен сведениями между людьми, человеком и автоматом, автоматом и автоматом; обмен сигналами в животном и растительном мире; передачу признаков от клетке к клетке, от организма к организму”.

В информационной сфере заняты люди самых различных профессий это - служащие учреждений, научные работники и специалисты, руководители всех рангов.

Известному ученому академику А.И.Бергу принадлежит образное выражение о том, что информация проникает во все поры жизни людей и общества, а жизнь невозможна в информационном вакууме. “Прямым опытом доказано, писал А.И.Берг, - что человек может нормально мыслить длительное время только при условии не прекращающемся информационного общением с внешним миром. Полная информационная изоляция от внешнего мира - это начало безумия.

Информационная, стимулирующая мышление, связь с внешним миром также необходима как пища и тепло, мало того как наличие энергетических полей, в которых происходит вся жизненная деятельность людей”.

Информация - основа управленческой деятельности недаром одно из определений понятия “управление” формулируется как “обмен между субъектом и объектом управления”.

Для функционирования любой системы необходимо:

- собрать информацию о состоянии системы и окружающей ее среды;
- передать эту информацию по каналам связи на место ее переработки;
- обработать информацию и сформировать команды управления;
- реализовать команду управления, т.е. передать исполнительным органам информацию, содержащую команду управления;

- проконтролировать выполнение данных операций.

Ныне информация, подобно полезным ископаемым, таким как уголь, нефть, газ, становится одним из важнейших национальных ресурсов.

По оценкам специалистов, “в конце этого столетия информационные ресурсы станут основным национальным богатством, а эффективность их промышленной эксплуатации во все большей степени будет определять экономическую мощь страны в целом”.

Информация состоит из **сообщений**, которые могут быть в виде речи и текста, графиков и таблиц и т.д.

Лиц, заинтересованных в содержании сообщений, принято называть **потребителями информации**.

Сообщения целенаправленно передаются от источника (создателя) информации к потребителю через **информационные коммуникации**.

Информационные связи принято делить на **формальные** и **неформальные**.

Первые осуществляются посредством специально созданных организаций: информационных служб, библиотек, печати, теле и радио вещания.

Вторые - это личные встречи телефонные разговоры переписка, т.е. непосредственно возникающие связи между создателями и потребителями информации.

Существует несколько способов классификации сообщений. Например, **по форме представления сообщения могут быть документальными и не документальными**.

Исторически сначала возникли не документальные сообщения. Они передаются с помощью речи, жестов и даже запахов по неформальным каналам.

В наши дни для распространения недокументальных сообщений часто используются такие технические средства как радио и телефон. Их достоинствами является прежде всего оперативность, избирательность, быстрая и надежная обратная связь. К недостаткам относится невозможность объективной проверки достоверности и долговременного хранения ограниченный круг распространения информации.

Перечисленные недостатки отсутствуют в документальных сообщениях, представленных в виде текста, набор цифровых данных, магнитных записей и т.д.

Сообщения бывают непрерывные и дискретные.

Непрерывность - отсутствие промежутков. Например, непрерывно изменяется значение силы тока температуры воздуха.

Дискретными являются величины с фиксированным набором элементов. Например, число работников милиции, количество патрульных автомобилей и т.д.

Если дискретное сообщение состоит из цифр, то информацию называют цифровой. Существуют специальные способы, позволяющие представить в цифровой форме как дискретные так и непрерывные сообщения.

Информацию классифицируют в зависимости от сферы возникновения, способов передачи и восприятия, по социальному назначению.

Принято различать три вида информации в зависимости от сферы возникновения:

- **элементарная** (неживая природа);
- **биологическая** (мир животных и растений);
- **социальная** (человеческое общество).

Эстетическая информация обязана своим происхождением возникающим в природе сочетанием звуков запахов, цвета, света и теней. Произведения искусства (музыка, живопись, литература) также относят к эстетической информации.

Семантическая информация возникает в результате деятельности людей. В правоохранительной сфере примером семантической информации может служить сообщение о совершенном преступлении.

Информация, собранная на месте преступления, - результат криминалистической деятельности работников ОВД, также является **семантической**.

В зависимости от **способа передачи и восприятия** можно выделить следующие виды информации:

- **визуальная** (передается и воспринимается зрительными образами);
- **аудиальная** (звуками)
- **тактильная** (ощущениями);
- **вкуссовая** (запахами);
- **машинно-ориентированная** (воспринимаемая и обрабатываемая ЭВМ).

По **социальной ориентации** выделяют **массовую, личную и специальную** информацию.

Если **массовая** информация ориентирована на самый широкий круг потребителя, то личная адресуется строго определенному лицу или группе лиц.

Специальная информация рассчитана на специалистов и может быть научной или художественной, технической или гуманитарной и т.д. Специальную информацию часто подразделяют по отраслевому принципу.

Например: машиностроительная, приборостроительная, энергетическая, юридическая.

В настоящее время наиболее распространенным носителем информации является **бумага**, на которой печатаются газеты, журналы, книги.

Для **фиксирования** информации применяются хорошо всем знакомые карандаши и ручки, а также такие технические средства, как пишущие машинки, магнитофоны, диктофоны.

Информация может быть записана посредством пробивки отверстий по специальному правилу (коду) на бумажных перфорационных картах и лентах.

Носителями информации являются также чертежи, схемы, карты, электрифицированные табло. Для наглядного отображения информации используются и другие технические средства: магнито - светопланы, световые планшеты, проекторы и диапроекторы, автоматические справочные установки, средства для чтения микрофильмов.

В последнюю четверть века в связи с широким распространением ЭВМ носителями информации служат магнитные ленты, жесткие и гибкие магнитные диски, оптические диски.

Передача информации осуществляется как путем пересылки носителя информации потребителю (например, доставка газет почтой), так и с помощью проводной и радиосвязи (телефон, радио, телевидение). В перспективе для этих целей будут широко использоваться спутниковые и оптоволоконные линии связи.

В правоохранительных органах для получения и передачи информации используются телефоны, телетайпы, радиостанции, почтовые и фельдъегерская связь, промышленное телевидение, каналы связи.

Информация, зафиксированная на бумаге и снабженная необходимым реквизитом (гриф, исходящий номер и др.), становится документом.

Для копирования и размножения документов применяются современные технические средства (типа “ксерокс”), а для их уничтожения специальные устройства.

Источниками информации, поступающей в правоохранительные органы, могут быть отдельные граждане, трудовые коллективы, учреждения (больницы, суды и т.п.) и т.д.

Информация принимается в **устном** виде и в виде **документов**: заявлений, протоколов, сообщений и т.п.

В правоохранительных органах информацией пользуются различные службы и подразделения, информационные центры, ИТУ, учебные заведения и научно - исследовательские учреждения МВД России.

Внутри правоохранительных органов информация хранится и перерабатывается (т.е. создается новая) в виде приказов и указаний, наставлений и инструкций, оперативных, розыскных и других дел, статистических карточек.

Выводы:

- * Информация - это сведения или совокупность сведений о предметах, явлениях и процессах окружающего нас мира.
- * Информация является основой управленческой деятельности.
- * Информация состоит из сообщений, которые могут быть представлены в виде речи и текста, графиков и таблиц и т.д.
- * Информацию классифицируют в зависимости от сферы возникновения, способов передачи и восприятия, по социальному назначению.
- * В последнюю четверть века в связи с широким распространением ЭВМ носителями информации служат магнитные ленты, жесткие и гибкие магнитные диски.
- * В ОВД для получения и передачи информации используются телефоны, телетайпы, радиостанции, почтовые и фельдъегерская связь, промышленное телевидение.
- * Внутри ОВД информация хранится и перерабатывается (т.е. создается новая) в виде приказов и указаний, наставлений и инструкций, оперативных, розыскных и других дел, статистических карточек.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ИНФОРМАЦИИ

К числу основных характеристик информации относятся: *целевое назначение, объем, ценность, полнота, надежность, достоверность, избыточность, скорость передачи и обработки.*

Целевое назначение информации является одной из важнейших ее характеристик, ибо она часто применяется в различных целях. Например, одни и те же данные могут быть использованы как для анализа сложившейся оперативной обстановки, так и для проведения оперативно-розыскных мероприятий.

Для передачи и обработки информации важное значение имеет ее **объем**, который зависит от количества передаваемых знаков (символов).

Ценность информации во многом определяется как своевременностью ее передачи, степенью влияния на принимаемое решение, так и важностью самого решения. Ценность информации зависит от ряда ее характеристик: полнота, надежность, достоверность.

Характеристика, называемая **полнотой**, используется для определения содержания наиболее существенных параметров передаваемой информации.

Информация считается **полной**, если она соответствует требуемому объему. Несоответствие между затребованной и полученной информацией свидетельствует либо о неполноте, либо о ее излишестве.

Надежность информации характеризуется наличием ошибок или их отсутствием. Она во многом зависит от используемых технических средств.

Для определения степени соответствия информации тому объекту, явлению или процессу, которые она отражает, используется характеристика, называемая **достоверность**.

Под **избыточностью** информации понимается увеличение объема передаваемых данных, не влекущих за собой получение дополнительных сведений.

И последняя характеристика - **скорость передачи и обработки** информации. Она зависит от быстродействия используемых технических средств и систем.

Выводы:

К числу основных характеристик информации относятся: *целевое назначение, объем, ценность, полнота, надежность, достоверность, избыточность, скорость передачи и обработки.*

ХРАНЕНИЕ И ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ.

ЕДИНИЦЫ ИЗМЕРЕНИЯ

Как уже отмечалось ранее, первостепенным объектом исследования информатики, так же как и правовой информатики, является информация и информационные процессы.

Свойства информации и правила ее обработки определяют устройство и порядок работы ЭВМ. Поэтому, рассмотрение принципов работы ЭВМ целесообразно начать с изложения сущности понятия "информация".

ЭВМ применяется для обработки очень специфичного "материала" - информации, правила обработки которой задаются программами.

Информация - это сведения или совокупность сведений об объектах и явлениях реального мира снимающие неопределенность существующую до их получения.

Все многообразие окружающей нас информации можно сгруппировать по различным признакам.

По способу **восприятия и передачи:**

- | | |
|---------------------|---------------|
| - видимыми образами | - визуальная; |
| - звуками | - аудиальная; |
| - ощущениями | - тактильная; |
| - средствами ЭВТ | - машинная |

По области **возникновения**, информация отражающая :

- процессы и явления животного и растительного мира - биологическая;
- процессы и явления неодушевленной природы - элементарная;
- процессы в человеческом обществе - социальная и т.д.

Общественные процессы относятся к самой высокой форме движения материи, поэтому социальная информация является наиболее сложным и многообразным видом информации. Она отражает отношения людей и их взаимодействие и потребности, интересы, мотивы деятельности, эмоциональное состояние и психическое отношение.

Социальная информация может быть разделена на:

- экономическую;
- политическую;
- научно-техническую;
- правовую и др.

Одной из разновидностей социальной информации является информация правоохранительных органов.

Требования предъявляемые к информации:

Достоверность информации - означает соответствие поступающих сведений объективной реальности, действительному состоянию дел. На достоверность информации оказывает влияние ряд факторов:

- надежность и степень совершенства технических средств ее получения и передачи;
- психологические закономерности восприятия информации (человек, получая даже достоверную информацию может внести в нее элемент субъективного - на это оказывает влияние такие факторы как настроение, отношение к делу). Действительность может быть искажена умышленно (дезинформация).

В случаях когда возникает необходимость убедиться в достоверности полученной информации и обеспечить ее объективность, прибегают к проверке информации - получению дополнительных сведений через другие источники. **Достоверность** - качественный критерий информации.

Однако здесь наряду с качественной характеристикой следует обратить внимание на такой показатель как **количество** информации.

Современная теория информации под количеством информации понимает степень снятия неопределенности в результате полученного сообщения.

Полнота - одно из центральных и наиболее сложных требований предъявляемых к информации.

Полнота как требование к информации наиболее полно раскрывается еще в римской формулировке раскрытия преступления, обстоятельств подлежащих доказыванию: кто, что, где, с чьей помощью, почему, каким образом, когда.

С требованием полноты самым непосредственным образом связан вопрос о времени хранения информации. Преждевременное уничтожение информации может привести к безвозвратной утере важных сведений.

Своевременность получение информации - связана с организацией четкой системы передачи; получения информации; обнаружением и поиском необходимой информации.

Непрерывность и **систематичность** поступление информации - связаны с непрерывным характером самого процесса управления.

Объекты и явления реального мира, а точнее их состояние в пространстве и времени, используя физические величины:

- массу;
- температуру;
- силу тока;
- напряжение и т.д.

При этом, природа некоторых физических величин такова, что их величина может принимать любые значения в некотором диапазоне. Такие величины называют **непрерывными** /аналоговыми/.

Пример непрерывной величины:

- температура тела человека;
- скорость движения автомобиля и т.д.

Вместе с тем, существуют физические величины, не являющиеся непрерывными. Например - число деталей в станке, число букв на странице книги.

Такие величины принимают только определенные значения и называются **дискретными**.

Для представления информации в непрерывной форме используют непрерывные величины, связанные с реальными процессами /ток, напряжение/.

Для представления дискретной информации используют последовательность символов, характеризующую изменяющуюся величину.

Например: Дом № 35. Цифры 3 и 5 являются символами, дающие информацию о месте положения дома на улице.

Однако, в реальной жизни информация об объектах представляется не только в виде цифр, а обычно в виде слов и их последовательностей. Слова в свою очередь состоят из букв определенного алфавита /русского, латинского, греческого/.

Произвольный фиксированный конечный набор символов любой природы называется алфавитом.

Возникает вопрос, если всю совокупность информации мы можем представить в виде букв и цифр /символов/, то как можно представить буквы, цифры /символы/, чтобы без потери информации их можно было бы удобно обрабатывать на ЭВМ.

Оказывается что для представления любой числовой информации можно ограничиться алфавитом из девяти цифр /1,2,3.....9/.

Совокупность десятичных цифр вместе с запятой для отделения дробной части числа можно рассматривать как частный случай абстрактного (придуманного нами) алфавита с 11 символами - буквами алфавита. Другой пример - алфавит русского языка. Понятие буквы здесь несколько отличается от общепринятого в быту, т.е. цифра может быть

буквой алфавита, точка, вопросительный знак, тире - это тоже буквы алфавита в том смысле, о котором идет речь. Как мы теперь видим, понятие алфавита, с помощью которого представляется информация, здесь несколько шире, чем в общепринятом бытовом смысле.

Имея дело с информацией, нужно уметь не только ее представлять в каком-то виде, но и измерять ее количество.

Единицей количества информации является **бит** (от английского binary digit - двоичная цифра). **Один бит - это минимальное количество информации, которое показывает, что произошло одно из двух равновероятных событий.** Случай неравновероятных событий не рассматривается. В памяти ЭВМ информация кодируется в виде последовательности цифр 0 и 1. Например, выпадению "орла" пусть соответствует 0, а выпадению "решки" - 1. Тогда, записывая 0, мы точно знаем, что речь идет о факте выпадения "орла", а записывая 1, мы точно знаем, что речь идет о факте выпадения "решки".

Из всех возможных форм записи информации для вычислительной техники в большинстве случаев используются только два символа - 0 и 1. Это могут быть какие-то два различающихся между собой физических состояния (свет - темнота, холодно -

горячо, наличие электрического напряжения - отсутствие его, сухо - влажно и т.д.). Если использовать для записи этих состояний символы, то достаточно всего двух разных символов.

Теперь перейдем к объяснению двоичной системы счисления. В нашей повседневной жизни мы привыкли пользоваться десятичной системой счисления и не можем представить себе более удобной, хотя этот выбор обусловлен всего-навсего наличием 10 пальцев на руках. В десятичной системе счисления принято пользоваться 10 цифрами для представления чисел: 0,1,2,3,4,5,6,7,8 и 9. Если имеется число меньше 10, то хватает одной цифры. Если число от 10 до 99, то надо использовать уже две цифры или два разряда: один для единиц, а другой для десятков. От 100 до 999 требуется три разряда и т.д. К такой системе все привыкли. Но то, что удобно для человека, оказывается е очень удобно для машины, в которой нет пальцев, а есть электронные приборы. А электронному прибору гораздо легче различить наличие напряжения или его отсутствие, чем распознавать десятки различных значений этого напряжения. Поэтому, пожертвовав удобством представления чисел в десятичной системе исчисления, решили воспользоваться самой простой из возможных - двоичной системой счисления, что позволило преодолеть массу технических проблем при создании вычислительных машин.

Двумя символами можно записать только 0 или 1. Если требуется записать число два, то надо использовать следующий, старший разряд, аналогично десятичной системе счисления, где после достижения последней цифры - числа 9 надо записывать 0 в данном разряде и 1 в старшем разряде. Аналогично в двоичной системе счисления число 2 записывается как 10, а число 3 как 11. Для числа 4 потребуется еще один разряд и оно записывается как 100. Число 5 записывается в виде 101, 6 - в виде 110 и 7 - в виде 111. Для записи десятичных чисел 8 и 9 потребуется использование уже четвертого двоичного разряда. Иными словами, говоря о разрядности, мы имеем дело со степенями двойки: 2 в степени 0 = 1, 2 в степени 1 = 2, 2 в степени 3 = 8, 2 в степени 4 = 16 и т.д. Показатель степени говорит о том, сколько двоичных разрядов мы имеем, а результат возведения в степень - максимальное число (0 тоже считается) состояний, которое можно записать в таком виде двоичных разрядов.

Небольшая таблица показывает соответствие некоторых десятичных и двоичных чисел:

0	0	
1	1	2 в степени 0
2	10	2 в степени 1
3	11	
4	100	2 в степени 2
5	101	
6	110	
7	111	
8	1000	2 в степени 3
9	1001	

10	1010	
11	1011	
12	1100	
13	1101	
14	1110	
15	1111	
16	10000	2 в степени 4

Для представления информации в ЭВМ используется **двухсимвольный** алфавит, состоящий всего из двух символов "0" и "1".

Это связано с тем, что современные ЭВМ создаются на базе электронных элементов, каждый из которых может находиться только в двух устойчивых состояниях, соответствующих низкому и высокому уровню напряжений.

Например: 0 - 0,5 Вольта - "0"; 2,5-5,0 Вольта - "1".

Таким образом мы подошли к тому, что всю информацию можно представить в виде последовательности символов, причем как текстовую так и числовую.

В качестве таких символов в ЭВМ используют "1" и "0". **Представление информации двоичными символами, а цифры 0 и 1 называются битами информации.**

При двоичном кодировании информации каждому символу сопоставляется его последовательность из фиксированного количества нулей и единиц.

В большинстве ЭВМ, каждому символу соответствует последовательность из 8 "0" и "1", называется **байтом**. **Один бит - это минимальное количество информации, которое может принимать одно из двух значений - "0" или "1".**

Из **восьми** "0" и "1" можно сформировать **256** неповторяющихся последовательностей, что позволяет закодировать **256** разных символов.

Например: большие и малые буквы русского и латинского алфавитов, цифры, знаки препинания и т.д.

Соответствие байтов и символов задается с помощью специальной стандартизированной кодовой таблицы.

Кодированием, называется процесс построения данных из элементов конечного множества по установленным правилам.

Кодом данных, или просто кодом, называется система, образуемая кодовым набором и правилами, по которым из элементов этого кодового набора строят данные при кодировании.

Таким образом закодированные буквы, цифры, графические знаки занимают в памяти ЭВМ по **одному байту**.

Два байта соседних составляют машинное **полуслово**, четыре соседних байта - **машинное слово**.

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. Второй фактор, также вызывающий 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

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 непременно 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}.

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:

$N = 2^m$, 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.

$$19:2 = 9+1 \quad 9:2 = 4 + 1 \quad 4:2 = 2 + 0 \quad 2:2=1$$

Таким образом, $19_{10} = 1011_2$.

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 65535, а 24 бита — уже более 16,5 миллионов разных значений.

Для кодирования действительных чисел используют 80-разрядное кодирование. При этом число предварительно преобразуется в нормализованную форму:

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

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

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

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 приведена в таблице 1.1.

Таблица 1.1. 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

Таблица 1.1. Базовая таблица кодировки ASCII

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.

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 закрепилась и нашла широкое распространение (таблица 1.2). Эта кодировка

используется на большинстве локальных компьютеров, работающих на платформе Windows. ,

Таблица 1.2. Кодировка Windows 1251

Таблица 1.2. Кодировка Windows 1251

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

Другая распространенная кодировка носит название КОИ-8 {код обмена информацией восьмизначный) — ее происхождение относится ко временам действия Совета Экономической Взаимопомощи государств Восточной Европы (таблица 1.3). Сегодня кодировка КОИ-8 имеет широкое распространение в компьютерных сетях на территории России и в российском секторе Интернета.

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).

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 (см. таблицу 1.5).

Таблица 1.3. Кодировка КОИ-8

Таблица 1.2. Кодировка Windows 1251

128 Ъ	144 ђ	160	176 °	192 А	208 Р	224 а	240 р
129 Ҁ	145 '	161 Ҁ	177 ±	193 Б	209 С	225 б	241 с
130	146	162 Ҁ	178 Ҁ	194 В	210 Г	226 в	242 г
131	147	163 Ҁ	179 Ҁ	195 Г	211 Д	227 д	243 д
132	148	164 Ҁ	180 Ҁ	196 Д	212 Е	228 е	244 е
133	149	165 Ҁ	181 Ҁ	197 Е	213 Ж	229 ж	245 ж
134	150	166 Ҁ	182 Ҁ	198 Ж	214 З	230 з	246 з
135	151	167 Ҁ	183 Ҁ	199 З	215 И	231 и	247 и
136	152	168 Ҁ	184 Ҁ	200 И	216 Й	232 й	248 й
137 ‰	153 ™	169 ©	185 №	201 Й	217 К	233 к	249 к
138 Љ	154 ъ	170 €	186 с	202 К	218 Ь	234 л	250 ъ
139	155	171 €	187 »	203 Л	219 Ъ	235 л	251 ъ
140 Њ	156 ъ	172 -	188 Ј	204 М	220 Ь	236 м	252 ъ
141 К	157 к	173 -	189 S	205 Н	221 Э	237 н	253 э
142 ҂	158 ҂	174 ©	190 s	206 О	222 Ю	238 о	254 ю
143 ҃	159 ҃	175 і	191 і	207 П	223 Я	239 п	255 я

В связи с избытием 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.

Единицы измерения количества информации

Для измерения массы, длины, времени и т.д. придуманы приборы и процедуры измерения.

Например, достаточно приложить линейку и мы узнаем длину предмета.

А как узнать количество информации в сообщении и в каких единицах эту информацию измерять?

Для двоичных сообщений в качестве меры информации используют количество бит в сообщении. Это количество называют информационным объемом.

Биты и байты используются также для измерения "емкости" памяти ЭВМ и для измерения скорости передачи двоичных сообщений / **бит/сек** /.

Наряду с битами и байтами для измерения количества информации используются и более крупные единицы:

Байт	1 байт=8 бит.
Кб	1Кбайт=1024 байт= 2^{10} байт.
Мб	1Мбайт=1024 байт= 2^{20} байт.
Гб	1Гбайт=1024 байт= 2^{30} байт.
Тб	1Тбайт=1024 байт= 2^{40} байт.

Байт служит для образования более крупных единиц информации.

Какой объем информации в байтах получится на странице учебника, если:

- каждая буква, знак, пробел -1 байт;
- на странице = 50 строк;
- в каждой строке = 60 знаков /байт/;

Тогда информационный объем страницы = 3000 байт или 3 Кб.