

Информатика и информационные технологии

Лекция 1. Информация, информационные процессы и системы

1.1. Введение в информатику и информационные технологии

Слово *informatik* (информатика) впервые было введено в немецкий язык немецким ученым, в области кибернетики, профессором университета в г. Карлсруэ К. Штейнбухом в 1957 году. Позднее, в 1962 году слово *informatique* было введено и во французский язык Ф. Дрейфузом, французским инженером-физиком и программистом, профессором Гарвардского университета (США). Это слово является акронимом, поскольку образовано было путем слияния начала слова *information* (информация) и конца слова *automatique* (автоматика) и означало дословно на русском языке информационная автоматика (автоматизация). В англоязычных странах французскому слову *informatique* соответствует синоним *computer science* (наука о компьютерной технике). В СССР слово информатика было впервые введено советскими учеными Михайловым А.И., Черным А.И. Гиляровским Р.С. в 1968 году. В нашей стране слово информатика утвердилось с момента принятия соответствующего решения в 1983 году на сессии годовичного собрания Академии наук СССР об организации нового отделения информатики, вычислительной техники и автоматизации.

Словом информатика стала называться комплексная наука, изучающая все аспекты разработки, проектирования, создания, оценки, функционирования основанных на ЭВМ систем переработки информации, их применения и воздействия на различные области социальной практики, а также учебная дисциплина, которая стала преподаваться в школах СССР с 1-го сентября 1985 года, а затем и в вузах. Преподавание дисциплины «Информатика» в школах совпало с выходом учебника академика Ершова А.П. «Основы информатики и вычислительной техники». Здесь также можно отметить, что 4 декабря в нашей стране отмечается День российской информатики, поскольку в этот день в 1948 году Государственный комитет Совета министров СССР по внедрению передовой техники в народное хозяйство зарегистрировал изобретение № 10475 принадлежащее Бруку И.С. и Рамееву Б.И. под названием цифровая электронная вычислительная машина.

Между информатикой наукой и информатикой учебной дисциплиной существует сходство и различие, которое лежит в плоскости решаемых ими задач. Для информатики как науки основной задачей является разработка теоретических положений на основе принятой аксиоматики, т.е. системы начальных положений и аксиом, и экспериментальная проверка этих положений. Для информатики являющейся учебной дисциплины основная задача состоит в разработке методик доведения до учащихся этих положений

в доступной форме и формирования у учащихся научного мировоззрения, позволяющего реализовать информационный подход к исследованию мира.

В процессе развития информатики как науки изменялись и ее определения, что связано с большой трудностью ее однозначного определения, и в конечном итоге это привело к тому, что определения стали заменяться соответствующими понятиями. Можно привести несколько понятий информатики, которые отражают определенные этапы ее развития.

Информатика – техническая наука о системах и средствах обработки информации при помощи ЭВМ.

Информатика – фундаментальная естественная наука, изучающая процессы передачи и обработки информации (академик Ершов А.П., 1986 г.).

Информатика – общенаучная дисциплина, изучающая законы, процессы, методы и средства преобразования и распространения информации в природе и обществе, в том числе – при помощи технических систем (профессор Колин К.К., 1990 г.).

Информатика – фундаментальная наука и общеобразовательная дисциплина (Конгресс Юнеско, Москва 1996 г.).

Информатика – фундаментальная наука об информационных процессах в природе и обществе, комплексное междисциплинарное направление (ИПИ РАН, 2006 г.).

Каждое из представленных понятий имеет место на существование, поскольку все они являются следствием общей проблемы терминологий, которая была и остается одной из самых актуальных в любой науке, в том числе и в информатике. Это объясняется во многом тем, что в настоящее время эволюция информатики как науки связана в значительной мере с развитием понятийного аппарата, в частности с попытками уточнения основных понятий информации и информатики и с разработкой методов и средств, для количественной оценки информации. Как отмечено в [1] весьма перспективным может оказаться, в частности, понимание информатики как смежной науки об информации и информационном взаимодействии в различных предметных областях.

В этом отношении информатика связана с рядом наук: электроникой, радиотехникой, семиотикой, кибернетикой и т.д. Однако наиболее близко информатика связана с кибернетикой в плане формирования основных понятий. Это, во-первых, объясняется тем, что информатика развивалась в значительной мере в недрах кибернетики, более ранней науки, фактически на единой технической базе - вычислительной технике, средствах связи и передачи данных. Во-вторых, кибернетика, являясь наукой об общих законах и закономерностях управления и связи, объективно была вынуждена заниматься вопросами использования информации в интересах управления [2]. Исходя из этого дадим определение основных понятий информатики с учетом понятий в области кибернетики, которые сведем в таблице 1.1.

Таблица 1.1

Основные понятия	Кибернетика	Информатика
Определения	Наука об общих законах и закономерностях управления и связи в сложных системах различной природы	Наука об информации и реализации информационных процессов с помощью соответствующих технических средств
Объекты исследования	Управление, процессы управления	Информация, информационные процессы
Предметы исследования	Системы и технологии управления	Информационные системы и технологии
Понятия	Управление, процессы управления, система управления, обратная связь, модель, информация, технологии управления...	Информация, информационные процессы, системы, технологии, модель...
Основные задачи	Анализ и синтез технологий и систем управления	Создание информационных технологий и информационных систем

Таким образом, информатика является наукой, у которой нет единого определения, поэтому его заменяют понятиями. Эти понятия формулируют ученые, исходя из собственных взглядов на информатику.

Основными составляющими современной информатики являются:

- теоретическая информатика, часть информатики, которая посвящена разработке теоретических положений, связанных с созданием и анализом различных информационных моделей, описывающих состояние информационного поля с использованием различных инструментов познания, в основном математических;

- компьютерная техника, часть информатики в которой разрабатываются общие принципы построения компьютерных систем – архитектура, принципы функционирования и взаимодействия технических устройств;

- программирование часть информатики, в которой разрабатываются принципы алгоритмизации и компьютерные программные средства.

Объектами приложений информатики являются самые различные области практической деятельности человека, где она является теоретической основой для разнообразных информационных технологий, применяемых обществом в экономике, медицине, телевидении, образовании и т.д. При этом под информационными технологиями будем понимать совокупность методов, средств и производственных процессов, используемых обществом для реализации информационных процессов. Основная задача, решаемая с помощью информационных технологий –

повышение производительности труда в любой сфере человеческой деятельности.

Информационные технологии, применяемые в различных областях практической деятельности человека, имеют сходства и различия. Общим для них является то, что все они базируются на науке информатики, а различия между ними определяются предметной областью, где они используются.

1.2. Понятия информации, информационного поля, информационного процесса и информационной системы

Понятия информация, информационный процесс и информационная система являются основополагающими в курсе информатики, поэтому формирование этих понятий было и остается сложной проблемой, связанной в первую очередь с различными мировоззренческими подходами (религиозными, научными и т.д.) к решению этой проблемы, как отдельными авторами, так и представителями различных научных направлений и школ. Информация как научная категория является предметом изучения многих научных дисциплин: философии, информатики, кибернетики, физики, биологии и т.д. Информация, материя, энергия являются базовыми научными категориями, для которых не существует строгих научных определений, поэтому их заменяют соответствующими понятиями и попытки выразить одно базовое понятие через другое нельзя считать успешными. Слово «информация» происходит от латинского слова *informatio*, что в переводе означает сведение, ознакомление, разъяснение. Дословная трактовка этого слова означает то, что присуще человеческому сознанию и общению и не отражает того многообразия свойств информации, которое проявляется не только на уровне общения между людьми, но и на уровне живой и неживой природы, включая и технические устройства, созданные человеком. Существует достаточно много различных определений понятия информации, но ни одно из них не может быть принято в качестве основного, поскольку все они отражают только отдельные свойства информации, к которым можно отнести:

- дуализм, свойство информации, которое характеризует ее двойственность, с одной стороны она объективна т.к. не зависит от воспринимаемого ее объекта, с другой субъективна, т.к. может трактоваться субъектами по-разному;
- достоверность, свойство информации, которое характеризует степень соответствия информации реальному объекту с заданной точностью;
- адекватность, свойство информации, которое устанавливает соответствие создаваемого с помощью информации образа реальному объекту, процессу или явлению;
- и т.д.

Этот ряд свойств может быть продолжен. Однако здесь наиболее важным представляется отметить следующее: совокупность свойств

информации, характеризующих степень ее соответствия потребностям пользователей, называется качеством информации.

Поэтому целесообразным представляется при определении понятия информации использовать подход, нашедший применение в теории информации и основанный на введении соответствующих ограничений на количество свойств информации, проявляющихся при взаимодействии объектов, организованных в систему.

Введем общее ограничение, при котором будем рассматривать свойства информации, проявляющиеся только на уровне человеческого сознания.

Если не принимать во внимание такие свойства информации, как смысловое содержание, полезность, актуальность и т.д., то при определении понятия информации можно ограничиться только вероятностным (энтропийным) уровнем рассмотрения этого понятия. Этот уровень рассмотрения называют также синтаксическим. На данном уровне показываются внешние характеристики информации независимо от ее смыслового содержания, характера использования и т.д. При таком рассмотрении, под информацией, следует понимать меру устранения неопределенности в отношении интересующего нас исхода, или другими словами, информация есть снятая неопределенность при осуществлении выбора из множества альтернатив.

Если принять во внимание только свойство информации, связанное с ее смысловым содержанием, то при определении понятия информации можно ограничиться смысловым или семантическим уровнем рассмотрения этого понятия. В этом случае под информацией следует понимать меру, связывающую ее смысловое содержание со способностью человека ее осмыслить.

Если принять во внимание только свойство информации, связанное с ее полезностью, то при определении понятия информации можно ограничиться прагматическим уровнем рассмотрения этого понятия. В этом случае под информацией можно понимать меру достижения потребителем информации (человеком) поставленной практической цели.

Таким образом, принципиально может существовать бесконечно много уровней ограничений и соответственно понятий информации. Данный вывод не содержит противоречий, поскольку учитывает одно из основных свойств информации быть всеобъемлющей. Если постулировать данное свойство, т.е. принять его как всеобщее свойство материи, связанное с отражением (материалистический подход), то можно говорить о существовании информационного поля, для которого необходимо наличие соответствующих условий.

К таким условиям можно отнести:

- а) наличие источника информации;
- б) наличие среды распространения информации (канала передачи);
- в) наличие материальных посредников распространения информации;
- г) наличие приемника информации.

В качестве источника информации выступать естественные природные источники, т.е. явления живой и неживой природы и искусственные источники, созданные человеком. Наличие источника информации является необходимым условием для возникновения и существования информационного поля.

В качестве среды распространения информации можно выделить среды с распределенными параметрами (естественные природные среды – воздушные, водные и др. среды) и среды с сосредоточенными параметрами (искусственные каналы передачи) созданные человеком. Наличие среды распространения информации является необходимым условием для возникновения и существования информационного поля.

В качестве материальных посредников распространения информации можно выделить естественные (природные) и искусственные посредники, к которым можно отнести знаки, сигналы и сообщения. Существует достаточно много определений этих понятий, однако остановимся здесь, на трактовке этих понятиях исходя из представлений, выработанных семиологией, наукой предметом изучения которой являются знаковые системы.

Знак или символ можно рассматривать как некий чувственно-воспринимаемый объект, который условно представляет и отсылает к обозначаемому им предмету, явлению, событию или действию. Если придерживаться этого определения, то между знаком (символом) и обозначаемым им предметом (референтом) устанавливается условная связь в виде понятия (референции) об этом предмете, которую можно представить с помощью известного треугольника Огдена и Ричардса (рисунок 1.1).

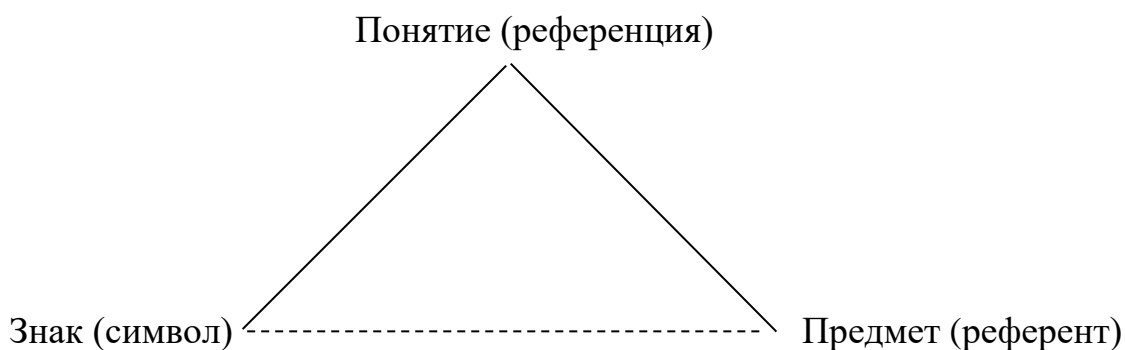


Рис. 1.1. Треугольник Огдена и Ричардса.

Эта связь условна (на рис.1.1 она показана в виде пунктирной линии), поскольку нет прямой связи между знаком и предметом. В действительности существуют знаки (символы) у которых есть понятия, но нет предмета. Одним из примеров этого могут служить некоторые знаки зодиака (водолей, стрелец), мифологические персонажи (единорог) и т.д.

Сигнал можно рассматривать как материальный носитель, с помощью которого передается информация. В общем случае под сигналом следует понимать физический процесс, представленный в той или иной материально-

энергетической форме (электрической, звуковой и т.д.) и распространяющийся в определенной среде. Сигнал можно рассматривать и как некоторый процесс, которому свойственна непрерывность и дискретность во времени, т.е. сигналы, могут быть непрерывными или дискретными (прерывными). Для передачи информации сигнал, как материальный носитель, подвергается преобразованию (модуляции, кодированию и т.д.) т.е. определенному воздействию со стороны передаваемой информации. Кроме этого, на сигнал может воздействовать и сама среда распространения сигнала в виде неблагоприятных воздействий – помех. Свойства сигналов изучаются в теории сигналов и цепей.

Сообщение можно отнести к основной характеристике передаваемой информации. Сообщение является формой представления информации в виде совокупности знаков, которая передается с помощью сигналов от источника к приемнику. Сообщение как форма представления информации подвергается соответствующему кодированию с помощью различных знаковых систем. Примерами сообщений могут быть: телеграммы, речь, музыка, телевизионное изображение, данные на выходе вычислительных и измерительных систем и т.д.

В качестве приемника информации могут выступать естественные природные приемники, т.е. объекты живой и неживой природы и искусственные приемники, созданные человеком и соответствующим образом настроенные на прием. В качестве приемника передаваемой информации может выступать непосредственно и сам человек. Наличие приемника информации не является необходимым условием для возникновения и существования информационного поля, поскольку информационное поле существует объективно, т.е. независимо от нас и необязательно подлежит осмыслению.

Таким образом, в общем случае коммуникативный процесс между источником и приемником сообщений можно представить в виде структурной схемы (рисунок 1.2)

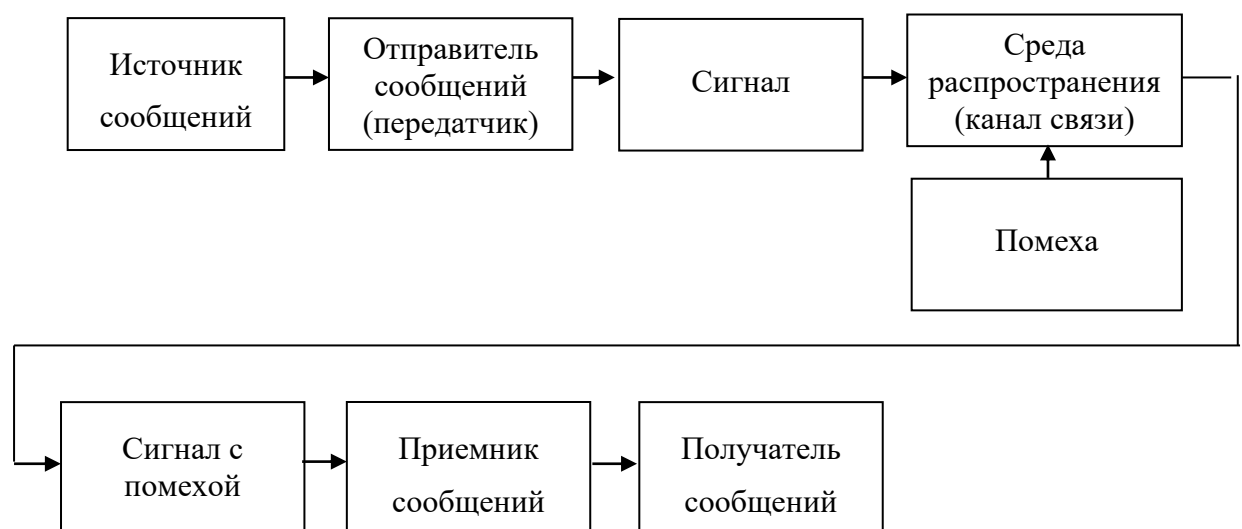


Рис. 1.2. Коммуникативный процесс между источником и приемником сообщений

В подтверждение тому, что может существовать много различных понятий информации, приведем еще одно, сформулированное на основе условий существования информационного поля: информация есть свойство информационного поля, которое проявляется при обязательном выполнении всех его условий возникновения и существования.

Говоря об информационном поле, необходимо также заметить, что ему присуще: многообразие видов и форм представления, наличие энергетических, геометрических (пространственных), временных и других характеристик, а также изменение этих характеристик, под воздействием неблагоприятных факторов (помех).

Многообразие видов информационного поля приводит и к многообразию видов самой информации. Для выявления определенных видов информации необходимо согласно существующим методам классификации применить тот или иной классификационный признак (критерий) для их разграничения. Таких классификационных признаков может быть достаточно много (n). В табл. 1.2 приведены некоторые из них и соответствующие им виды информации.

Таблица 1.2

<i>Классификационный признак</i>	<i>Вид информации</i>
1. Область возникновения	Механическая, биологическая, социальная
2. Тип материального носителя	Документальная, речевая, телекоммуникационная
3. Форма представления	Аналоговая, дискретная
4. Форма проявления в задачах управления	Осведомляющая, управляющая и т.д.
.....
.....
n.

Приведенные в табл. 1.2 виды информации могут образовывать иерархические структуры, для которых характерно расположение видов информации в порядке от высокого уровня к низкому уровню. Например, по области возникновения (классификационный признак первого уровня) различают: механическую (отражающую процессы и явления неживой природы), биологическую (отражающую процессы и явления живой природы) и социальную (отражающую процессы и явления в человеческом обществе). Социальную информацию в свою очередь по назначению (классификационный признак второго уровня) можно разделить на три вида: массовую, специальную и личную. Если использовать классификационный признак третьего уровня (специфика проявления), то массовый и специальный вид информации можно разделить на подвиды. Так, массовую информацию можно представить в виде общественно-политической и научно-популярной, а специальную - в виде научной, технической, экономической и управляющей.

Таким образом, существует иерархическая структура классификационных признаков и соответствующих им видов информации, примеров этому можно привести достаточно много. Приведем здесь в качестве примера, одно из определений понятия экономической информации. Экономическая информация – это информация, которая используется при осуществлении функций управления экономическими процессами, т.е. она отражает и обслуживает процессы производства, распределения, обмена и потребления материальных благ. Экономическая информация служит основой для осуществления управления процессами в области экономики и обладает рядом характерных свойств: многоцелевым назначением, цикличностью, простотой математических операций, сложностью логических процедур и т.д.

С понятием информация тесно связаны такие понятия как информационный процесс и информационная система. Это также достаточно сложные понятия, трактовка которых далеко неоднозначна.

Под информационным процессом будем понимать последовательность действий, выполняемых с информацией. К таким действиям можно отнести: отбор (восприятие) информации, преобразование информации, передачу информации, обработку информации, хранение и поиск информации, принятие решения.

Кратко охарактеризуем каждое из этих действий, более подробное их описание можно найти в соответствующей литературе.

Отбор информации связан с целенаправленным извлечением и анализом информации об объекте, процессе или явлении. Результатом этого отбора является формирование модели (образа) объекта, процесса или явления их опознание и оценка. Отбор информации выполняет несколько функций, среди которых можно выделить две важных функции: оценка информации на значимость и сжатие информации. При оценке информации на значимость происходит ее селекция или информационная фильтрация, результатом которой является отделение полезной информации от информационных шумов (помех). Сжатие информации необходимо для ее представления в компактном виде, который позволяет более эффективно использовать каналы, по которым передается информация и средства ее хранения.

Преобразование информации включает в себя операции, связанные с нормализацией (приведение к определенным нормам), аналого-цифровым преобразованием, кодированием и т.д. Результатом преобразования информации является материальный носитель (сигнал) представленный в форме, удобной для передачи, обработки и хранения.

Передача информации является важной составляющей информационного процесса, в результате которого происходит транспортировка информации от источника информации к приемнику. Передача осуществляется в среде распространения по соответствующим каналам, имеющим различную физическую природу: электромагнитную, электрическую, оптическую и т.д.

Обработка информации также является важной составляющей информационного процесса. Под обработкой можно понимать преобразование информации, осуществляемое по определенным законам (математики, логики, физики и т.д.) и неформальным правилам (опыт, интуиция, «здравый смысл» и т.д.). Результатом обработки является тоже информация, которая может быть представлена в иных формах (например, упорядоченная по каким-то признакам) или содержащая ответы на поставленные вопросы. В случае если процесс обработки подлежит формализации, то он может осуществляться техническими средствами. Примером может служить обработка информации на компьютере.

Хранение информации связано с ее записью на некоторый материальный носитель с целью дальнейшего поиска, извлечения и использования. В качестве таких носителей могут использоваться: бумага, магнитная лента, накопители на жестких магнитных дисках, накопители на оптических и магнитооптических дисках и т.д.

Важной составляющей информационного процесса является принятие решения, которое может осуществляться человеком или технической системой в зависимости от поставленной задачи.

В заключении, говоря об информационном процессе, необходимо отметить, что это не простая последовательность действий, выполняемых с информацией, а сложная их взаимосвязь, взаимопроникновение и развитие. В каждом из действий (составляющих процесса информации) можно найти признаки других составляющих: в отборе – преобразование, в хранении – передачу, в передаче – обработку и т.д.

Реализуются информационные процессы в системах, которые называются информационными (ИС), при этом под системой в общем случае понимается совокупность объектов и отношений между ними существующая как единое целое.

ИС также можно классифицировать по различным признакам: по сфере применения, по организации информационных процессов, по уровню автоматизации, по территориальному признаку и т.д. Например, по сфере применения, ИС можно подразделить: на административные, экономические, производственные, медицинские и т.д. По уровню автоматизации: на автоматизированные ИС, в которых человек принимает существенное участие и автоматические ИС, без участия человека или при несущественном его участии.

На рисунке 1.3 представлена обобщенная структура ИС. Представленная структура ИС является открытой, поскольку в зависимости от конкретной задачи решаемой ИС будет изменяться ее конфигурация и состав. Необходимо также отметить, что почти всегда течение информационных процессов в ИС определяется двумя взаимно противоречивыми тенденциями. Это, во-первых, стремление представить потребителю наиболее полную информацию, что, естественно, удлиняет продолжительность информационного процесса и, во-вторых, стремление представить эту информацию в наиболее короткие сроки, что зачастую ведет

к уменьшению объема информации. Недостаточно полный объем или запаздывание в общем случае имеют один и тот же результат - обесценивание информации

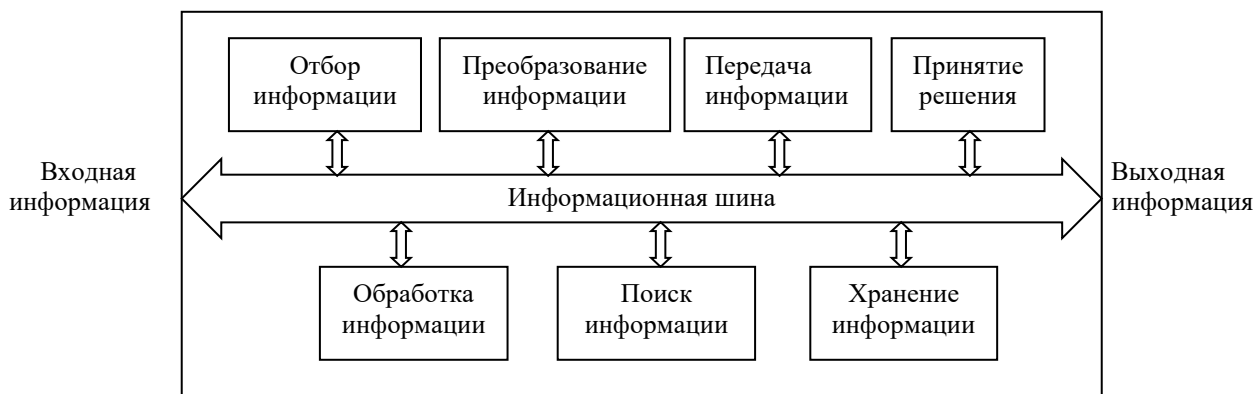


Рис. 1.3 Обобщенная структура ИС

1.3. Определение количество информации. Единицы измерения количества информации

Как уже отмечалось, понятие информации можно рассматривать с учетом различных ограничений, накладываемых на свойства информации, т.е. различных уровнях рассмотрения. В основном выделяют три уровня: синтаксический, семантический и прагматический. Соответственно на каждом из этих уровней применяют различные оценки для определения количества информации.

На синтаксическом уровне для измерения количества информации принят энтропийный подход. Этот подход основан на том, что получение информации всегда связано с уменьшением неопределенности (энтропии) системы [1, 10, 13, 14]. Таким образом, количество информации в сообщении определяется как мера уменьшения неопределенности состояния системы после получения сообщения. Количество информации при таком подходе определяется следующим образом:

$$I = H_{apr.} - H_{aps.}$$

где $H_{apr.}$ – априорная энтропия (до получения сообщения) о состоянии системы или процесса.

$H_{aps.}$ – апостериорная энтропия (после получения сообщения) о состоянии системы или процесса.

Из формулы видно, что в случае, когда после полученного сообщения неопределенность снята ($H_{aps.} = 0$) количество полученной информации равно

$$I = H_{apr.}$$

Разработанные в середине XX века математические и в, частности, вероятностные методы позволили реализовать энтропийный подход к оценке количества получаемой информации. Выдающийся американский ученый К. Шеннон обобщил понятие меры неопределенности, которая зависит не только от числа состояний системы, но и от вероятностей этих состояний.

Такой подход постулирует принцип: если некоторое сообщение приводит к уменьшению неопределенности наших знаний о некоторой системе, то можно утверждать, что такое сообщение содержит информацию. При этом сообщения содержат информацию о каких-либо событиях, которые могут реализоваться с различными вероятностями. Формулу для определения количества информации, получаемой после принятия сообщений с различными вероятностями от дискретного источника информации предложил К. Шеннон в 1948 году. Согласно этой формуле, количество информации, может быть определено следующим образом:

$$I = - \sum_{i=1}^N p_i \log_2 p_i \quad (1.1)$$

где I – количество информации,
 N – количество возможных сообщений (событий),
 p_i – вероятность отдельных сообщений (событий).

В данной формуле количество информации принимает только положительные значения. Поскольку вероятность отдельных событий в формуле 1.1 меньше единицы, то соответственно выражение $\log_2 p_i$ является отрицательной величиной и для получения положительного значения количества информации в формуле 1.1 перед знаком суммы стоит знак минус.

Если вероятности появления отдельных событий равны между собой и они образуют полную группу событий, т.е.

$$\sum_{i=1}^N p_i = 1$$

то формула 1.1 преобразуется в формулу Р. Хартли:

$$I = \log_2 N \quad (1.2)$$

В формулах 1.1 и 1.2 отношение между количеством информации и соответственно между вероятностью отдельных сообщений (событий) или количеством отдельных сообщений (событий) выражается с помощью логарифма. Применение логарифма можно объяснить следующим образом. Будем последовательно подставлять в формулу 1.2 значения N , которым для простоты вычисления I присвоим значения 1, 2, 4, 8, 16, 32, 64 и т.д., при этом получим следующие значения I : 0, 1, 2, 3, 4, 5, 6 и т.д. Значения чисел N представляют собой ряд, известный в математике как ряд чисел образующих геометрическую прогрессию, значения найденных чисел I будет являться рядом образующий арифметическую прогрессию. Таким образом, логарифм в формулах 1.1 и 1.2 устанавливает соотношение, между рядами представляющих геометрическую и арифметическую прогрессии, что достаточно хорошо известно в математике.

Для количественного определения любой физической величины необходимо определить единицу измерения, которая в теории измерений носит название меры. Информацию перед обработкой, передачей и хранением необходимо подвергнуть кодированию. Кодирование информации производится с помощью специальных алфавитов (знаковых систем). В информатике, изучающей процессы отбора, преобразования, обработки, передачи, хранения и поиска информации с помощью вычислительных (компьютерных) систем в основном используется двоичное кодирование, которое использует два символа 0 и 1. При этом, исходя из вероятностного подхода к определению количества информации, эти два символа двоичной знаковой системы можно рассматривать как два различных возможных сообщения или события. Исходя из этого за единицу количества информации принято такое количество информации, которое содержит сообщение, уменьшающее неопределенность в два раза. Такая единица называется bit (от английского Binary digit – двоичная цифра) или на русском языке бит. Это название также было предложено К. Шенноном. Таким образом, минимальный объем информации, который может быть передан с помощью двоичного кодирования равен одному биту. Следующей по величине единицей количества информации является байт, представляющий собой последовательность двоичных цифр, составленный из восьми бит, т.е.

$$1 \text{ байт} = 2^3 \text{ бит} = 8 \text{ бит.}$$

Широко используются также кратные байту единицы измерения количества информации, однако в отличие от метрической системы мер, где в качестве множителей кратных единиц используется коэффициент 10^n , где $n=3,6,9$ и т.д. в кратных единицах измерения количества информации используется коэффициент 2^n . Выбор этот объясняется тем, что компьютер оперирует числами не в десятичной, а в двоичной системе счисления.

Согласно ГОСТ 8.417-2002 Государственная система обеспечения единства измерений (ГСИ). Единицы величин (с Поправками), кратные байту единицы измерения количества информации вводятся следующим образом:

$$\begin{aligned} 1 \text{ Килобайт (Кбайт)} &= 2^{10} \text{ байт} = 1024 \text{ байт} \\ 1 \text{ Мегабайт (Мбайт)} &= 2^{10} \text{ Кбайт} = 1024 \text{ Кбайт} \\ 1 \text{ Гигабайт (Гбайт)} &= 2^{10} \text{ Мбайт} = 1024 \text{ Мбайт} \\ 1 \text{ Терабайт (Тбайт)} &= 2^{10} \text{ Гбайт} = 1024 \text{ Гбайт} \\ 1 \text{ Петабайт (Пбайт)} &= 2^{10} \text{ Тбайт} = 1024 \text{ Тбайт} \\ 1 \text{ Экзабайт (Эбайт)} &= 2^{10} \text{ Пбайт} = 1024 \text{ Пбайт} \\ 1 \text{ Зеттабайт (Збайт)} &= 2^{10} \text{ Эбайт} = 1024 \text{ Эбайт} \\ 1 \text{ Йоттабайт (Йбайт)} &= 2^{10} \text{ Збайт} = 1024 \text{ Збайт} \end{aligned}$$

Необходимо отметить, что исторически сложилась такая ситуация, что с наименованием «байт» некорректно (вместо $1000 = 10^3$, принято $1024 = 2^{10}$) использовали (и используют) приставки СИ: 1 Кбайт = 1024 байт, 1 Мбайт = 1024 Кбайт и т.д. При этом обозначение Кбайт начинают с прописной буквы в отличие от строчной буквы «к» для обозначения множителя 10^3 .

Весьма интересным представляется определение количества информации, представленной с помощью знаковых систем. Если

рассматривать символы алфавита как множество возможных сообщений N , то количество информации, которое несет один знак можно определить по формуле К. Шеннона. При равновероятном появлении каждого знака алфавита в тексте для определения количества информации можно воспользоваться формулой:

$$I = \log_2 N.$$

Количество информации, которое несет один знак алфавита, тем больше, чем больше знаков входят в этот алфавит и тем больше его мощность.

Поясним вышесказанное на примерах.

Пример 1.1

Рассчитаем, какое количество информации можно получить после реализации одного из шести событий. Вероятность первого события составляет 0,15, второго – 0,25, третьего – 0,2, четвертого – 0,12, пятого – 0,12, шестого – 0,1, т. е. $P_1 = 0,15$, $P_2 = 0,25$, $P_3 = 0,2$, $P_4 = 0,18$, $P_5 = 0,12$, $P_6 = 0,1$.

Решение.

Для расчета количества информации применим формулу (1.1)

$$I = -(P_1 \cdot \log_2 P_1 + P_2 \cdot \log_2 P_2 + P_3 \cdot \log_2 P_3 + P_4 \cdot \log_2 P_4 + P_5 \cdot \log_2 P_5 + P_6 \cdot \log_2 P_6) \text{ бит.} = -(0,15 \cdot \log_2 0,15 + 0,25 \cdot \log_2 0,25 + 0,2 \cdot \log_2 0,2 + 0,18 \cdot \log_2 0,18 + 0,12 \cdot \log_2 0,12 + 0,1 \cdot \log_2 0,1) \text{ бит.}$$

Для вычисления этого выражения, содержащего логарифмы, воспользуемся сначала компьютерным калькулятором, а затем табличным процессором Excel, входящим в интегрированный пакет программ Microsoft (MS) Office 2007, или его более поздней версией, например MS Office 2010, MS Office 2016 и т.д.

Для вычисления с помощью компьютерного калькулятора выполним следующие действия:

1. Запустим программу Калькулятор. Программа Калькулятор в операционной системе Windows 7 запускается с помощью команды [Кнопка «Пуск» - Все программы – Стандартные – Калькулятор]. После запуска программ выполним команду [Вид – Инженерный]. На рисунке 1.4 показан внешний вид инженерного калькулятора.

Кнопка «log» калькулятора производит вычисление десятичного (по основанию 10) логарифма отображаемого числа. Поскольку в нашем случае необходимо производить вычисления логарифмов по основанию 2, а данный калькулятор не вычисляет логарифм по основанию 2, то необходимо воспользоваться известной формулой:

$$\log_b N = M \cdot \log_a N, \text{ где } M = \frac{1}{\log_a b} \text{ (модуль перевода).}$$



Рис 1.4. Инженерный калькулятор

В нашем случае соотношение примет вид:

$$\log_2 N = M \cdot \log_{10} N, \text{ где } M = \frac{1}{\log_{10} 2} = 3,322, \text{ т.е.}$$

$\log_2 N = 3,322 \cdot \log_{10} N$, и выражение для вычисления количества информации примет вид:

$$I = -3,322(0,15 \cdot \log_{10} 0,15 + 0,25 \cdot \log_{10} 0,25 + 0,2 \cdot \log_{10} 0,2 + 0,18 \cdot \log_{10} 0,18 + 0,12 \cdot \log_{10} 0,12 + 0,1 \cdot \log_{10} 0,1) \text{ бит.}$$

При вычислении на калькуляторе используем кнопки: +/- (изменение знака отображаемого числа), () (открывающие и закрывающие скобки), log (логарифм числа по основанию 10) и т.д. Результат вычисления показан на рисунке 1.4. Таким образом, $I = 2,519552$ бит.

Воспользуемся теперь табличным процессором MS Office Excel 2007. Для запуска программы Excel выполним команду [Кнопка «Пуск» - Все программы – MS Office – MS Office Excel 2007]. В ячейки A1, B1, C1, D1, E1, F1 открывшегося окна Excel заносим буквенные обозначения вероятностей $P_1, P_2, P_3, P_4, P_5, P_6$ а в ячейку G1 количество информации I , которое необходимо определить. Для написания нижних индексов у вероятностей $P_1 \div P_6$ в ячейках A1, B1, C1, D1, E1, F1 выполним следующую команду: [Формат – Ячейки – Шрифт – Видоизменение (поставим флажок напротив нижнего индекса)]. В ячейки A2, B2, C2, D2, E2, F2 заносим соответствующие значения вероятностей. После занесения значений в этих ячейках необходимо установить формат чисел. Для этого необходимо выполнить следующую команду: [Формат – Ячейки – Число – Числовой (устанавливаем число десятичных знаков равное двум)]. Устанавливаем в ячейке G2 тот же числовой формат. В ячейку G2 записываем выражение $= -(A2*LOG(A2;2) + B2*LOG(B2;2) + C2*LOG(C2;2) + D2*LOG(D2;2) + E2*LOG(E2;2) + F2*LOG(F2;2))$. После нажатия на клавиатуре компьютера клавиши «Enter», в ячейке G2 получим искомый результат, $I = 2,52$ бит. На рисунке 1.5 представлен конечный результат вычисления количества информации для данного примера.

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇
1	0,15	0,20	0,28	0,28	0,32	0,30	0,32

Рис 1.5. Результат вычисления количества информации

Пример 1.2

Определим, какое количество байт и бит информации содержится в сообщении, если его объем составляет 0,25 Кбайт.

Решение.

С помощью калькулятора, определим количество байт и бит информации, которое содержится в данном сообщении.

$$I = 0,25 \text{ Кбайт} \cdot 1024 \text{ байт/1Кбайт} = 256 \text{ байт.}$$

$$I = 256 \text{ байт} \cdot 8 \text{ бит/1байт} = 2048 \text{ бит.}$$

Пример 1.3

Определим мощность алфавита, с помощью которого передано сообщение, содержащее 4096 символов, если объем сообщения составляет 2 Кбайта.

Решение.

С помощью калькулятора переведем информационный объем сообщения в биты:

$$V = 2 \text{ Кбайт} \cdot 1024 \text{ байт/1 Кбайт} = 2048 \text{ байт} \cdot 8 \text{ бит/1байт} = 16384 \text{ бит.}$$

Определим количество бит, приходящееся на один символ в алфавите:

$$I = 16384 \text{ бит} : 4096 = 4 \text{ бит.}$$

Используя формулу (1.2) определим мощность алфавита (количество символов в алфавите):

$$N = 2^I = 2^4 = 16.$$

Как уже отмечалось, если принять во внимание только свойство информации, связанное с ее смысловым содержанием, то при определении понятия информации можно ограничиться смысловым или семантическим уровнем рассмотрения этого понятия.

На семантическом уровне информация рассматривается по ее содержанию, отражающему состояние отдельного объекта или системы в целом. При этом не учитывается ее полезность для получателя информации.

На данном уровне изучаются отношения между знаками их предметными и смысловыми значениями (см. рис. 1.1), что позволяет осуществить выбор смысловых единиц измерения информации. Поскольку смысловое содержание информации передается с помощью сообщения, т.е. в виде совокупности знаков (символов) передаваемых с помощью сигналов от источника информации к приемнику, то широкое распространение для измерения смыслового содержания информации получил подход, основанный на использовании тезаурусной меры. При этом под тезаурусом понимается совокупность априорной информации (сведений), которой располагает приемник информации. Данный подход предполагает, что для понимания (осмысливания) и использования полученной информации приемник (получатель) должен обладать априорной информацией (тезаурусом), т.е. определенным запасом знаков наполненных смыслом, слов, понятий, названий явлений и объектов между которыми установлены связи на смысловом уровне. Таким образом, если принять знания о данном объекте или явлении за тезаурус, то количество информации, содержащееся в новом сообщении о данном предмете, можно оценить по изменению индивидуального тезауруса под воздействием данного сообщения. В зависимости от соотношений между смысловым содержанием сообщения и тезаурусом пользователя изменяется количество семантической информации, при этом характер такой зависимости не поддается строгому математическому описанию и сводится к рассмотрению трех основных условий:

- тезаурус пользователя стремится к нулю, т.е. пользователь не воспринимает поступающее сообщение;
- тезаурус пользователя стремится к бесконечности, т.е. пользователь досконально знает все об объекте или явлении и поступившее сообщение его не интересует;
- тезаурус пользователя согласован со смысловым содержанием сообщения, т.е. поступившее сообщение понятно пользователю и несет новые сведения.

Два первых предельных случая соответствуют состоянию, при котором количество семантической информации, получаемое пользователем минимально. Третий случай связан с получением максимального количества семантической информации. Таким образом, количество семантической информации, получаемой пользователем, является величиной относительной, поскольку одно и то же сообщение может иметь смысловое содержание для компетентного и быть бессмысленным для некомпетентного пользователя. Отсюда вытекает и сложность получения объективной оценки количества информации на семантическом уровне ее рассмотрения, поэтому для получения такой оценки используют различные единицы измерения количества информации: абсолютные или относительные. В качестве абсолютных единиц измерения могут использоваться символы, реквизиты, записи и т.д. В качестве относительной единицы измерения (меры) может использоваться коэффициент содержательности, который определяется как

отношение семантической информации к ее объему. Например, для определения количества информации на семантическом уровне, полученного учащимися на занятиях, в качестве единицы измерения может быть принят исходный балл (символ) характеризующий степень усвояемости ими нового учебного материала, на основе которого можно косвенно определить количество информации, полученное каждым учащимся. Это количество информации будет выражено через соответствующий оценочный балл, в принятом диапазоне оценок.

При семантическом подходе к оценке количества информации и выборе единицы измерения существенным является вид получаемой информации (сообщения). Так данный подход к оценке количества экономической информации позволяет выявить составную единицу экономической информации, состоящую из совокупности других единиц информации, связанных между собой по смыслу. Элементарной составляющей единицей экономической информации является реквизит, т.е. это информационная, совокупность, которая не поддается дальнейшему делению на единицы информации на смысловом уровне. Деление реквизитов на символы приводит к потере их смыслового содержания. Каждый реквизит характеризуется именем, значением и типом. При этом под именем реквизита понимается его условное обозначение, под значением – величина, характеризующая свойства объекта или явления в определенных обстоятельствах, под типом – множество значений реквизита, объединенных определенными признаками и совокупностью допустимых преобразований.

Реквизиты принято делить на реквизиты – основания и реквизиты – признаки.

Реквизиты - основания характеризуют количественную сторону экономического объекта, процесса или явления, которые могут быть получены в результате совершения отдельных операций – вычислений, измерений, подсчета натуральных единиц и т.д. В экономических документах к ним можно отнести, например, цену товара, его количество, сумму и т.д. Реквизиты – основания в основном выражаются в цифрах, над которыми могут выполняться математические операции.

Реквизиты – признаки отражают качественные свойства экономического объекта, процесса или явления. С помощью реквизитов – признаков сообщения приобретают индивидуальный характер. В экономических документах к ним можно отнести, например, номер документа, имя отправителя, дату составления документа, вид операции и т.д. Реквизиты – признаки позволяют осуществлять логическую обработку единиц количества информации на семантическом уровне: поиск, выборку, группировку, сортировку и т.д.

Отдельный реквизит – основание вместе с относящимися к нему реквизитами – признаками образуют следующую в иерархическом отношении составную единицу экономической информации – показатель. Показатель имеет наименование, в состав которого входят термины, обозначающие измеряемый объект: себестоимость, затраты, мощность,

прибыль и т.д. Кроме того, показатель содержит формальную характеристику и дополнительные признаки. К формальной характеристике относится способ его получения (объем, сумма, прирост, процент, среднее значение и т.д.) а к дополнительным признакам относятся пространственно-временные (где находится измеряемый объект, время к которому относится данный показатель) и метрологические (единицы измерения).

Таким образом, с помощью совокупности реквизитов и соответствующих им показателей можно оценить количество экономической информации, получаемой от исследуемого объекта (источника информации).

Кроме подхода, основанного на использовании тезаурусной меры при определении количества информации на семантическом уровне, находят применение и другие подходы. Например, один из подходов связанный с семантической оценкой количества информации заключается в том, что в качестве основного критерия семантической ценности информации, содержащейся в сообщении, принимается количество ссылок на него в других сообщениях. Количество получаемой информации определяется на основе статистической обработки ссылок в различных выборках.

Подводя итог сказанному, относящемуся к определению количества информации и выбора единиц измерения на семантическом уровне, можно утверждать, что существовала и существуют проблема формирования единого системного подхода к определению информации на семантическом уровне. Это подтверждается и тем, что в свое время для создания строгой научной теории информации К. Шеннон вынужден был отбросить важную сторону, связанную со смысловым содержанием информации.

Кроме перечисленных уровней рассмотрения понятия информации достаточно широко используется прагматический уровень. На данном уровне информация рассматривается с точки зрения ее полезности (ценности) для достижения потребителем информации (человеком) поставленной практической цели. Данный подход при определении полезности информации основан на расчете приращения вероятности достижения цели до получения информации, и после ее получения. Количество информации, определяющее ее ценность (полезность) находится из соотношения:

$$I = \log_2 P_1 - \log_2 P_o = \log_2 \frac{P_1}{P_o} \quad (1.3)$$

где P_o – вероятность достижения цели до получения информации,

P_1 – вероятность достижения цели после получения информации.

В качестве единицы измерения (меры) количества информации, определяющей ее ценность, может быть принят один бит (при основании логарифма равно двум), т.е. это такое количество полученной информации при котором отношение вероятностей достижения цели равно двум.

Рассмотрим три случая, когда количество информации, определяющее ее ценность, равно нулю и когда она принимает положительное и отрицательное значение.

Количество информации равно нулю при $P_o = P_I$, т.е. полученная информация не увеличивает и не уменьшает вероятность достижения цели.

Значение информации является положительной величиной при $P_I > P_o$, т.е. полученная информация уменьшает исходную неопределенность и увеличивает вероятность достижения цели.

Значение информации является отрицательной величиной при $P_I < P_o$, т.е. полученная информация увеличивает исходную неопределенность и уменьшает вероятность достижения цели. Такую информацию называют дезинформацией.

Дальнейшее развитие данного подхода базируется на статистической теории информации и теории решений. При этом кроме вероятностных характеристик достижения цели, после получения информации вводятся функции потерь, и оценка полезности информации, производится в результате минимизации функции потерь. Максимальной ценностью обладает то количество информации, которое уменьшает потери до нуля при достижении поставленной цели.

Контрольные вопросы

1. В чем состоит сходство и различие между информатикой наукой и информатикой учебной дисциплиной.
2. С какой наукой у информатики наиболее родственные отношения.
3. Перечислите основные составляющие современной науки информатика.
4. Существует ли единственное верное определение информации.
5. Перечислите основные свойства информации.
6. Какие среды распространения информации существуют.
7. Укажите основные виды информации.
8. Дайте определение информационного процесса и системы.
9. Перечислите основные составляющие информационного процесса.
10. Существует ли ограничения на использование формулы К. Шеннона для определения количества информации.
11. В чем отличие синтаксической информации от семантической информации.
12. Перечислите единицы оценки синтаксической и семантической информации.