



Введение.

Понятие информации (*informatio* - разъяснение, осведомление, изложение) является одним из основных, ключевых понятий не только в информатике (в информологии - области знаний, изучающей проявление информации, её представление, измерение и т.д.), но и в математике, в физике и др. Понятие «информация» - плохо формализуемое и структурируемое понятие. В силу его всеобщности, объёмности, расплывчатости оно часто понимается неточно и неполно не только обучаемыми. Как правило, это понятие в курсе информатики не определяется, принимается как исходное базовое понятие, неопределяемый терм.

Информация трактуется по разному, например, как:

- любая сущность, которая вызывает изменения в некоторой информационно-логической (инфологической - состоящей из данных, знаний, абстракций и т.д.) модели системы (математика, системный анализ);
- сообщения, полученные системой от внешнего мира в процессе адаптивного управления, приспособления (теория управления, кибернетика);
- отрицание энтропии, отражение меры хаоса в системе (термодинамика);
- связи, устраняющие неопределённость в системе (теория информации);
- вероятность выбора в системе (теория вероятностей);
- отражение разнообразия в системе (физиология, биокибернетика);
- отражение материи, атрибут сознания, «интеллекта» системы (философия).

В соответствии с парадигмой фундаментальности информатики как образовательной и научной дисциплины, её исключительной роли в усилении междисциплинарных связей, познании системно-информационной картины мира, необходимо введение в это фундаментальное, первичное понятие информатики на достаточно научном, строго-понятийном и в то же время доступном, содержательном уровне, например, с помощью содержательных примеров и задач. Эта проблема связана с трудностями методологического и методического плана. Данная статья (адаптированные к школе выдержки из рукописи авторского учебника «Информатика» для студентов, второе переработанное и дополненное издание [9]) ставит своей целью облегчение этой трудной, но обязательно нуждающейся в рассмотрении задачи.

1. Понятие информации, виды информации

Информация - это некоторая последовательность (налицо упорядоченность) сведений, знаний, которые актуализируемы (получаемы, передаваемы, преобразуемы, сжимаемы или регистрируемы) с помощью некоторых знаков (символьного, образного, жестового, звукового, сенсомоторного типа). Это приращение, развитие, актуализация знаний, возникающее в процессе целеполагающей интеллектуальной деятельности человека. Никакая информация, никакое знание не появляется сразу - этому предшествует этап накопления, осмысления, систематизации опытных данных, взглядов. Знание - продукт такого процесса. Мышление - необходимый атрибут такого процесса.

Информация может существовать в пассивной (не актуализированной) и активной (актуализированной) форме.

Пример. Информация актуализируется сообщениями, при этом формы облачения информации в сообщения различны, например, для живых существ - сигналы, жесты, для технических устройств - сигналы. Информация передаваемая от одного человека другому, может передаваться символами (письмо), жестами (сигнальщик на боевом корабле), звуками (диктор), геометрическими фигурами (чертёжник), художественными образами (балерина). Информация передающаяся животными может быть передана звуками (лай, вой, писк), ситуационным поведением (образами). Информация в технических устройствах, автоматах может быть передана электрическими, магнитными, световыми импульсами, как это происходит в ЭВМ.

Информация по отношению к окружающей среде (или к использующей ее среде) бывает трех типов: входная, выходная и внутренняя.

Входная информация (по отношению к окружающей среде) - информация, которую система воспринимает от окружающей среды.

Выходная информация (по отношению к окружающей среде) - информация, которую система выдает в окружающую среду.



Внутренняя, внутрисистемная информация (по отношению к системе) - информация, которая хранится, перерабатывается, используется только внутри системы т.е. актуализируемая лишь только подсистемами системы. Это несколько идеализированное (особенно с точки зрения физики открытых систем) понятие.

Пример. Человек воспринимает, обрабатывает входную информацию, например, данные о погоде на улице, формирует выходную реакцию - ту или иную форму одежды. При этом используется внутренняя информация, например, это генетически заложенная (или приобретённая) физиологическая информация о реакции, например, о "морозостойкости" человека.

Пример. Генетически заложенная в молекулах ДНК информация и приобретённая информация (в памяти) влияют на поведение, на адаптацию человека в окружающей среде. В машинах первого поколения внутренняя структура определялась тысячами ламп, причем каждая из них отдельно была невысокой надёжности, т.е. вся система была ненадежной в работе. Это влияло на входную информацию, например, такие ЭВМ не были способны на работу в многозадачном режиме, в режиме реального времени (обработки сообщений по мере получения входных данных).

Информация по отношению к конечному результату проблемы бывает:

- **исходная** (на начало актуализации этой информации);
- **промежуточная** (от начала до завершения актуализации информации);
- **результатирующая** (после завершения её актуализации).

Пример. При решении системы линейных алгебраических уравнений информация от методах решения, среде реализации, входных данных (источники, точность и т.д.), размерности системы и т.д. является исходной информацией; информация о совместности системы уравнений, численных значениях корня и т.д. - результирующая; информация о текущих состояниях коэффициентов уравнений реализации схемы Гаусса - промежуточная.

Информация по изменчивости при её актуализации бывает:

- **постоянная** (не изменяемая никогда при её актуализации);
- **переменная** (изменяемая при актуализации);
- **смешанная** - условно - постоянная (или условно-переменная).

Пример. В известной физической задаче определения дальности полёта снаряда артиллерийского орудия, информация об угле наклона орудия может быть переменной, информация о начальной скорости вылета снаряда - постоянной, а информация о координатах цели - условно-постоянной.

Возможна также **классификация информации** и по другим признакам:

- **по стадии использования** (первичная, вторичная);
- **по полноте** (избыточная, достаточная, недостаточная);
- **по отношению к цели системы** (синтаксическая, семантическая, прагматическая);
- **по отношению к элементам системы** (статическая, динамическая);
- **по отношению к структуре системы** (структурная, относительная);
- **по отношению к управлению системой** (управляющая, советующая, преобразующая, смешанная);
- **по отношению к территории, территориально** (федеральная, региональная, местная, относящая к юридическому лицу, относящаяся к физическому лицу, смешанная);
- **по доступу** (открытая или общедоступная, закрытая или конфиденциальная, смешанная);
- **по предметной области, по характеру использования** (статистическая, коммерческая, нормативная, справочная, научная, учебная, методическая и т.д., смешанная) и другие.

Информация в философском аспекте бывает, в основном: *мировоззренческая; эстетическая; религиозная; научная; бытовая; техническая; экономическая; технологическая.*

Все это (с человеком) составляет **ноосферу** общества - более высокое состояние биосферы, возникшее в результате эволюции, структурирования, упорядочивания и гармонизации связей в природе и обществе под воздействием целеполагающей деятельности человечества. Это понятие введено впервые В. И. Вернадским в качестве отражения эволюции общества и природы т.е. системы, в рамках которой потенциально может быть реализовано гармоническое, устойчивое развитие (эволюция) систем "Общество" и "Природа", а также постепенное слияние, интеграция и гармонизация наук о природе, познании и об обществе. Без этого невозможно построение информационного общества.



Основные свойства информации (и сообщений):

- *полнота* (содержание всего необходимого для понимания информации);
- *актуальность (необходимость) и значимость* (сведений, знаний);
- *ясность* (выразительность сообщений на языке интерпретатора);
- *адекватность, точность, корректность* (актуализации знаний);
- *интерпретируемость и понятность* (интерпретатору информации);
- *достоверность* (отображаемых сообщениями информации);
- *информативность, значимость* (сообщений, отображающих информацию);
- *массовость* (применимость ко всем проявлениям);
- *кодируемость и экономичность* (кодирования, актуализации сообщений);
- *сжимаемость и компактность* (сообщений);
- *защищённость и помехоустойчивость* (актуализации информации);
- *доступность* (интерпретатору);
- *ценность* (значимость при достаточном уровне потребителя).

Пример. Рекламный щит - простой красочный кусок дерева (железа), но информация заложенная в сообщениях на этом щите должна обладать всеми вышеперечисленными свойствами и только тогда этот щит будет ассоциироваться у интерпретатора (человека) с рекламируемым товаром (услугами) и актуализировать информацию. При этом вся форма представления рекламы должна строиться с учетом понятности интерпретатору, быть информативной. Пока символы не организованы определенным образом, не используются для некоторой определённой цели, они не отражают информацию.

Информация может оказаться и вредной, влияющей негативно на сознание, например, воспитывающей восприятие мира от безразличного или же не критического - до негативного, "обозлённого", неадекватного. Информационный поток - достаточно сильный раздражитель.

Пример. Негативной информацией - раздражителем может быть информация о крахе коммерческого банка, о резком росте (спаде) валютного курса, об изменении налоговой политики и др.

Информация не существует без других типов ресурсов - энергии, вещества, организации, как и они не могут существовать без информации. Любые взаимодействия систем (подсистем) - взаимодействия всегда материально-энерго-информационные. Выявление (структурирование, упорядочивание, установление отношений), формализация (описание формальными средствами, языками), изучение (разработка моделей, методов, алгоритмов), применение (разработка и актуализация технологий) этих взаимодействий и составляет основную задачу информатики - как науки, как человеческой деятельности.

2. Методы получения, использования информации

Методы получения и использования информации можно разделить на три группы, условно разграничиваемые и часто перекрываемые друг другом.

1. **Эмпирические методы** или методы получения эмпирической информации (эмпирических данных).
2. **Теоретические методы** или методы получения теоретической информации (построения теорий).
3. **Эмпирико - теоретические методы** (смешанные, полуэмпирические) или методы получения эмпирико-теоретической информации.

Охарактеризуем кратко эмпирические методы.

1. **Наблюдение** - сбор первичной информации о системе (в системе).
2. **Сравнение** - установление общего и различного в системе (системах).
3. **Измерение** - нахождение эмпирических законов, фактов для системы.
4. **Эксперимент** - целенаправленное преобразование системы (систем).

Кроме этих классических форм их реализации в последнее время используются и такие формы как *опрос, интервью, тестирование* и др.

Охарактеризуем кратко эмпирико - теоретические методы.

1. **Абстрагирование** - установление общих свойств объекта (объектов), замещение системы ее моделью.
2. **Анализ** - разделение системы на подсистемы с целью выявления их взаимосвязей.
3. **Синтез** - соединение подсистем в систему с целью выявления их взаимосвязей.



4. **Индукция** - получение знания о системе по знаниям о подсистемах.
5. **Дедукция** - получение знания о подсистемах по знаниям о системе.
6. **Эвристики, использование эвристических процедур** - получение знания о системе по знаниям о подсистемах и наблюдениям, опыту.
7. **Моделирование**, использование приборов - получение знания об объекте с помощью модели и/или приборов.
8. **Исторический метод** - нахождение знаний о системе путем использования его предыстории.
9. **Логический метод** - метод нахождения знаний о системе путём воспроизведения его подсистем, связей или элементов в мышлении, в сознании.
10. **Макетирование** - получение информации по макету системы, т.е. с помощью представления подсистем в упрощенном виде, сохраняющем информацию, необходимую для понимания взаимодействия и связей этих подсистем.
11. **Актуализация** - получение информации с помощью активизации, инициализации её, т.е. переводом из статического (неактуального) состояния в динамическое (актуальное) состояние; при этом все необходимые связи и отношения (открытой) системы с внешней средой должны быть сохранены.
12. **Визуализация** - получение информации с помощью визуального представления состояний актуализированной системы; визуализация предполагает возможность выполнения операции типа “передвинуть”, “повернуть”, “укрупнить”, “уменьшить”, “удалить”, “добавить” и т.д. (как по отношению к отдельным элементам, так и к подсистемам системы).

В последнее время часто используются и такие формы как *мониторинг* (система наблюдений и анализа состояний системы), *деловые игры и ситуации*, *экспертные оценки* (экспертное оценивание), *имитация* (имитационная процедура, эксперимент) и др.

Охарактеризуем кратко теоретические методы.

1. **Восхождение от абстрактного к конкретному** - получение знаний о системе на основе знаний о его проявлениях в сознании, в мышлении.
2. **Идеализация** - получение знаний о системе или о ее подсистемах путём мысленного конструирования, представления в мышлении систем и/или подсистем, не существующих в действительности.
3. **Формализация** - получение знаний о системе с помощью знаков или же формул, т.е. языков искусственного происхождения, например, языка математики (или математическое, формальное описание, представление).
4. **Аксиоматизация** - получение знаний о системе или процессе с помощью некоторых, специально для этого сформулированных аксиом и правил вывода из этой системы аксиом.
5. **Виртуализация** - получение знаний о системе созданием особой среды, обстановки, ситуации, которую реально, без этой среды невозможно реализовать и получить соответствующие знания.

Все эти методы получения информации обычно применяются многоуровневым комплексным образом и можно предложить схему (рис. 1).

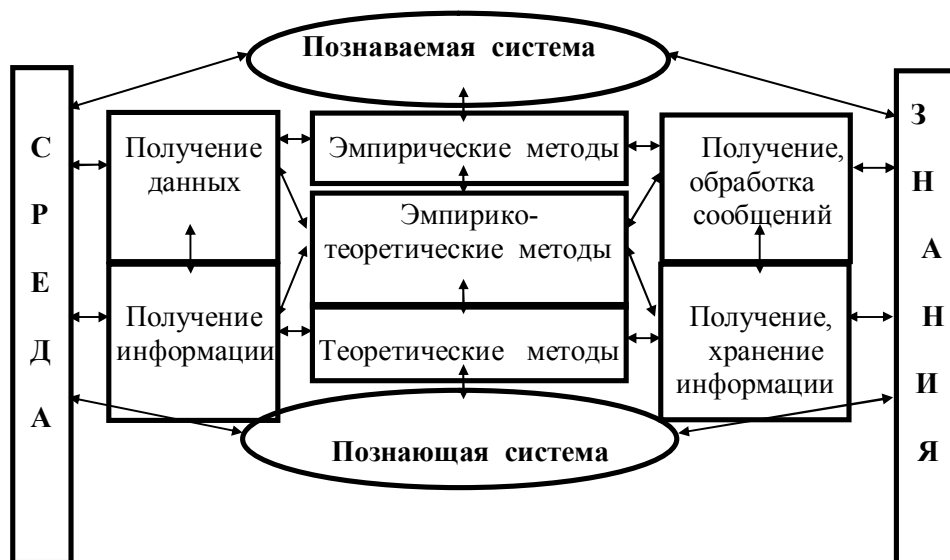




Рис. 1. Структура познания системы

3. Измерение сообщений и информации

Если отвлечься от конкретного смыслового содержания информации и рассматривать сообщения информации как последовательности знаков, сигналов, то их можно *представлять* битами, а *измерять* в байтах, килобайтах, мегабайтах, гигабайтах, терабайтах и петабайтах.

Выше было отмечено, что информация может пониматься и интерпретироваться по разному. Вследствие этого имеются различные подходы к определению методов измерения информации, меры количества информации. Раздел информатики (теории информации) изучающий методы измерения информации называется *информметрией*.

Количество информации - числовая величина, адекватно характеризующая актуализируемую информацию по разнообразию, сложности, структурированности, определённости, выбору (вероятности) состояний отображаемой системы.

Если рассматривается система, которая может принимать одно из n возможных состояний, то актуальна задача оценки такого выбора, исхода. Такой оценкой может стать мера информации (или события). Мера - это некоторая непрерывная действительная неотрицательная функция, определённая на множестве событий и являющаяся аддитивной т.е. мера конечного объединения событий (множеств) равна сумме мер каждого события.

1. Мера Р. Хартли. Пусть имеется N состояний системы S или N опытов с различными, равновероятными последовательными состояниями системы. Если каждое состояние системы закодировать, например, двоичными кодами определённой длины d , то эту длину необходимо выбрать так, чтобы число всех различных комбинаций было бы не меньше, чем N . Наименьшее число, при котором это возможно или мера разнообразия множества состояний системы задаётся формулой Р. Хартли: $H = k \log_a N$, где k - коэффициент пропорциональности (масштабирования, в зависимости от выбранной единицы измерения меры), a - основание системы меры.

Если измерение ведётся в экспоненциальной системе, то $k=1$, $H = \ln N$ (*нат*); если измерение - в двоичной системе, то $k=1/\ln 2$, $H = \log_2 N$ (*бит*); если измерение - в десятичной системе, то $k=1/\ln 10$, $H = \lg N$ (*дит*).

Пример. Чтобы узнать положение точки в системе из двух клеток т.е. получить некоторую информацию, необходимо задать 1 вопрос ("Левая или правая клетка?"). Узнав положение точки, мы увеличиваем суммарную информацию о системе на 1 бит ($I = \log_2 2$). Для системы из четырех клеток необходимо задать 2 аналогичных вопроса, а информация равна 2 битам ($I = \log_2 4$). Если система имеет n различных состояний, то максимальное количество информации равно $I = \log_2 n$.

Справедливо **утверждение Хартли**: если во множестве $X = \{x_1, x_2, \dots, x_n\}$ выделить произвольный элемент $x_i \in X$, то для того, чтобы найти его, необходимо получить не менее $\log_a n$ (единиц) информации.

По Хартли, для того, чтобы мера информации имела практическую ценность - она должна быть такова, чтобы отражала количество информации пропорционально числу выборов.

Пример. Имеются 192 монеты из которых одна фальшивая. Определим сколько взвешиваний нужно произвести, чтобы определить ее. Если положить на весы равное количество монет, то получим 2 возможности (мы сейчас отвлекаемся от того, что в случае фальшивой монеты таких состояний будет два - состояния независимы): а) левая чашка ниже; б) правая чашка ниже. Таким образом, каждое взвешивание дает количество информации $I = \log_2 2 = 1$ и, следовательно, для определения фальшивой монеты нужно сделать не менее k взвешиваний, где k удовлетворяет условию $\log_2 2^k \geq \log_2 192$. Отсюда, $k \geq 7$ или, $k = 7$. Следовательно, нам необходимо сделать не менее 7 взвешиваний (достаточно семи).

Пример. ДНК человека можно представить себе как некоторое слово в четырехбуквенном алфавите, где каждой буквой помечается звено цепи ДНК или нуклеотид. Определим сколько информации (в битах) содержит ДНК, если в нем содержится примерно $1,5 \times 10^{23}$ нуклеотидов. На один нуклеотид приходится $\log_2(4) = 2$ (бит) информации. Следовательно, структуры ДНК в организме человека позволяет хранить 3×10^{23} бит информации. Это вся информация, куда входит и избыточная. Реально используемой, - структурированной в памяти человека информации, - гораздо меньше. В этой связи, заметим, что человек за



среднюю продолжительность жизни использует около 5 — 6 % нейронов (нервных клеток мозга - “ячеек ОЗУ человека”). Генетический код - чрезвычайно сложная и упорядоченная система записи информации. Информация заложенная в генетическом коде (по учению Дарвина) накапливалась многие тысячелетия. Хромосомные структуры - своеобразный шифровальный код и при клеточном делении создаются копии шифра, каждая хромосома - удваивается, в каждой клетке имеется шифровальный код, при этом каждый человек получает, как правило, свой набор хромосом (код) от матери и от отца. Шифровальный код разворачивает процесс эволюции человека. Вся жизнь, как отмечал Э. Шредингер, “упорядоченное и закономерное поведение материи, основанное ... на существовании упорядоченности, которая поддерживается всё время”.

Формула Хартли отвлечена от семантических и качественных, индивидуальных свойств рассматриваемой системы (качества информации, содержащейся в системе, в проявлениях системы с помощью рассматриваемых N состояний системы). Это основная положительная сторона этой формулы. Но имеется и основная отрицательная сторона: формула не учитывает различимость и различность рассматриваемых N состояний системы.

Уменьшение (увеличение) H может свидетельствовать об уменьшении (увеличении) разнообразия состояний N системы.

Обратное, как это следует из формулы Хартли (основание логарифма берётся больше 1!), - также верно.

2. Мера К. Шеннона. Формула Шеннона дает оценку информации независимо, отвлеченно от ее смысла:

$$I = - \sum_{i=1}^n p_i \log_2 p_i.$$

где n - число состояний системы; p_i - вероятность (или относительная частота) перехода системы в i -ое состояние, причем сумма всех p_i равна 1.

Если все состояния равновероятны (т.е. $p_i = 1/n$), то $I = \log_2 n$.

К. Шенноном доказана теорема о единственности меры количества информации. Для случая равномерного закона распределения плотности вероятности мера Шеннона совпадает с мерой Хартли. Справедливость и достаточная универсальность формул Хартли и Шеннона подтверждается и данными нейробиологии.

Пример. Время t реакции испытуемого на выбор предмета из имеющихся N предметов линейно зависит от $\log_2 N$: $t = 200 + 180 \log_2 N$ (мс). По аналогичному закону изменяется и время передачи информации в живом организме. В частности, один из опытов по определению психофизиологических реакций человека состоял в том, что перед испытуемым большое количество раз зажигалась одна из n лампочек, которую он должен указать. Оказалось, что среднее время, необходимое для правильного ответа испытуемого, пропорционально не числу n лампочек, а именно величине I определяемой по формуле Шеннона, где p_i - вероятность зажечь лампочку номер i .

Легко видеть, что в общем случае:

$$I = - \sum_{i=1}^n p_i \log_2 p_i \leq \log_2 n.$$

Если выбор i -го варианта predetermined заранее (выбора, собственно говоря, нет, $p_i = 1$), то $I = 0$.

Сообщение о наступлении события с меньшей вероятностью несёт в себе больше информации, чем сообщение о наступлении события с большей вероятностью. Сообщение о наступлении достоверно наступающего события несёт в себе нулевую информацию (и это вполне ясно, - событие всё равно произойдёт когда-либо).

Пример. Если положение точки в системе известно, в частности, она - в k -ой клетке, т.е. $p_k = 1$, кроме $p_k = 1$, то тогда $I = \log_2 1 = 0$ и мы здесь новой информации не получаем.

Пример. Выясним, сколько бит информации несет каждое двузначное число со всеми значащими цифрами (отвлекаясь при этом от его конкретного числового значения). Так как таких чисел может быть всего 90 (10 - 99), то информации будет количество $I = \log_2 90$ или приблизительно $I = 6.5$. Так как в таких



числах значащая первая цифра имеет 9 значений (1- 9), а вторая - 10 значений (0-9), то $I = \log_2 90 = \log_2 9 + \log_2 10$. Приблизительное значение $\log_2 10$ равно 3.32. Итак, сообщение в одну десятичную единицу несет в себе в 3.32 больше информации, чем в одну двоичную единицу (чем $\log_2 2 = 1$), а вторая цифра в, например, числе aa несёт в себе больше информации, чем первая (если цифры разряда a неизвестны; если же эти цифры a известны, то выбора нет - информация равна нулю).

Если в формуле Шеннона обозначить $f_i = -n \log_2 p_i$, то получим, что I можно понимать как среднеарифметическое величин f_i .

Отсюда, f_i можно интерпретировать как информационное содержание символа алфавита с индексом i и величиной p_i вероятности появления этого символа в сообщении, передающем информацию.

Пусть сообщение состоит из n различных символов, m_i - количество символов номер $i=1, 2, \dots, n$ в этом сообщении, а N - длина сообщения в символах. Тогда вероятность появления i -го символа в сообщении равна $p_i = m_i/N$. Число всех различных сообщений длины n будет равно

$$p = N! / \prod_{i=1}^n m_i!$$

Информация в одном таком сообщении равна

$$I = \log_2 p = \ln p / \ln 2 = \ln (N! / \prod_{i=1}^n m_i!) / \ln 2.$$

Используя формулу Стирлинга (достаточно точна, например, при $N > 100$) - $N! \approx (N/e)^N$, а точнее, её следствие - $\ln N! \approx N(\ln N - 1)$ получаем (в битах):

$$I \approx (N \ln N - \sum_{i=1}^n m_i \ln m_i) / \ln 2 = - (N / \ln 2) \sum_{i=1}^n p_i \ln p_i.$$

Пример. Пусть рассматривается алфавит из двух символов русского языка - «к» и «а». Относительные частоты встречаемости этих букв в частотном словаре русского языка равны соответственно $p_1 = 0.028$, $p_2 = 0.062$. Возьмем произвольное слово p длины N из k букв «к» и m ($k+m=N$) букв «а» над этим алфавитом. Число всех таких возможных слов, как это следует из комбинаторики, равно $n = N!/(k! m!)$. Оценим количество информации в таком слове: $I = \log_2 n = \ln n / \ln 2 = \log_2 e [\ln N! - \ln k! - \ln m!]$. Воспользовавшись следствием приведённой выше формулой Стирлинга получаем оценку количества информации (в битах) на 1 символ любого слова:

$$\begin{aligned} I_1 &= I/N \approx (\log_2 e/N) [(k+m)(\ln N - 1) - k(\ln k - 1) - m(\ln m - 1)] = \\ &= (\log_2 e/N) [k \ln(N/k) - m \ln(N/m)] = \\ &= -\log_2 e [(k/N) \ln(k/N) + (m/N) \ln(m/N)] \leq \end{aligned}$$

$$\leq -\log_2 e [p_1 \ln p_1 + p_2 \ln p_2] = -\log_2 e [0.028 \ln 0.028 + 0.062 \ln 0.062] \approx 0.235.$$

Пример. В сообщении 4 буквы «а», 2 буквы «б», 1 буква «у», 6 букв «р». Определим количество информации в одном таком (из всех возможных) сообщении. Число N различных возможных сообщений длиной в 13 букв будет равно величине: $N = 13!/(4! \times 2! \times 1! \times 6!) = 180180$. Количество информации I в одном сообщении будет равно величине: $I = \log_2(N) = \log_2 180180 \approx 17.5$ (бит).

Если k - коэффициент Больцмана, известный в физике как $k = 1.38 \times 10^{-16}$ эрг/град, то выражение

$$S = -k \sum_{i=1}^n p_i \ln p_i$$

в термодинамике известно как **энтропия** или **мера хаоса**, беспорядка в системе. Сравнивая выражения I и S видим, что I можно понимать как информационную энтропию (энтропию из-за нехватки информации о/в системе).

Нулевой энтропии соответствует максимальная информация. Основное соотношение между энтропией и информацией:

$$I + S(\log_2 e)/k = \text{const}$$

или в дифференциальной форме

$$dI/dt = -((\log_2 e)/k) dS/dt.$$



Основными положительными сторонами формулы Шеннона является её отвлечённость от семантических и качественных, индивидуальных свойств системы, а также то, что в отличие от формулы Хартли она учитывает различность, разноречивость состояний - формула имеет статистический характер (учитывает структуру сообщений), делающий эту формулу удобной для практических вычислений. Основные отрицательные стороны формулы Шеннона: она не различает состояния (с одинаковой вероятностью достижения, например), не может оценивать состояния сложных и открытых систем и применима лишь для замкнутых систем, отвлекаясь от смысла информации.

Увеличение (уменьшение) меры Шеннона свидетельствует об уменьшении (увеличении) энтропии (организованности) системы. При этом энтропия может являться мерой дезорганизации систем от полного хаоса ($S=S^{max}$) и полной информационной неопределённости ($I=I^{min}$) до полного порядка ($S=S^{min}$) и полной информационной определённости ($I=I^{max}$) в системе.

Пример. Чем ближе движущийся объект к нам, тем полнее информация обрабатываемая нашими органами чувств, тем чётче и более структурирован (упорядочен) объект. Чем больше информации мы имеем о компьютерной технике, тем меньше психологический барьер перед ним (согласно основному соотношению между энтропией и информацией).

3. Термодинамическая мера. Информационно-термодинамический подход связывает величину энтропии системы с недостатком информации о её внутренней структуре (не восполняемым принципиально, а не нерегистрируемым). При этом число состояний определяет, по существу, степень неполноты наших сведений о системе.

Пусть дана термодинамическая система (процесс) S , а H_0, H_1 - термодинамические энтропии системы S в начальном (равновесном) и конечном состояниях термодинамического процесса, соответственно. Тогда термодинамическая мера информации (негэнтропия) определяется формулой:

$$H(H_0, H_1) = H_0 - H_1.$$

Эта формула универсальна для любых термодинамических систем. Уменьшение $H(H_0, H_1)$ свидетельствует о приближении термодинамической системы S к состоянию статического равновесия (при данных доступных ей ресурсах), а увеличение - об удалении.

Поставим некоторый вопрос о состоянии некоторой термодинамической системы. Пусть до начала процесса можно дать p_1 равновероятных ответов на этот вопрос (ни один из которых не является предпочтительным другому), а после окончания процесса - p_2 ответов. Изменение информации при этом:

$$\Delta I = k \ln(p_1 / p_2) = k (\ln p_1 - \ln p_2).$$

Если $p_1 > p_2$ ($\Delta I > 0$) - прирост информации, т.е. сведения о системе стали более определёнными, а при $p_1 < p_2$ ($\Delta I < 0$) - менее определёнными. Универсально то, что мы не использовали явно структуру системы (механизм протекания процесса).

Пример. Предположим, что имеется термодинамическая система - газ в объёме $10 \text{ (м}^3\text{)}$, который расширяется до объёма $20 \text{ (м}^3\text{)}$. Нас интересует вопрос о координате некоторой молекулы газа. В начале мы знали ответ на вопрос и поэтому $p_1 = 1$ ($\ln p_1 = 0$). Число ответов было пропорционально $[\ln 10]$. После поднятия заслонки мы знаем координату, микросостояние, т.е. изменение информации о состоянии системы равно $\Delta I = -k \ln(20/10) = -k \ln 2$ (нат). Это известное в термодинамике выражение для прироста энтропии в расчёте на одну молекулу и оно подтверждает второе начало термодинамики. Энтропия - мера недостатка информации о микросостоянии статической системы.

Величина ΔI может быть интерпретирована как количество информации, необходимой для перехода от одного уровня организации системы к другой (при $\Delta I > 0$ - более высокой, а при $\Delta I < 0$ - более низкой организации).

Термодинамическая мера (энтропия) применима к системам, находящимся в тепловом равновесии. Для систем, далёких от теплового равновесия, например, живых биосистем, мера - энтропия - менее подходящая.

4. Энергоинформационная (квантово-механическая) мера. Энергия (ресурс) и информация (структура) - две фундаментальные характеристики систем реального мира, связывающие их вещественные, пространственные, временные характеристики. Если A - множество "энергетического происхождения", а B - множество "информационного происхождения", то можно определить энергоинформационную меру (системы $A \leftrightarrow B$) как отношение, связь между этими множествами вида $f: A \rightarrow B$. Она должна отражать механизм взаимосвязей физико-информационных и вещественно-энергетических структур и процессов в



системе. Сейчас актуально говорить о биоэнергоинформационных мерах, отражающих механизм взаимосвязей биофизикоинформационных и вещественно-энергетических процессов в системе, в ноосфере.

Пример. Процесс деления клеток сопровождается излучением квантов энергии с частотами приблизительно до $N=1.5 \times 10^{15}$ гц. Этот спектр можно воспринимать как спектр функционирования словарного запаса клетки - как биоинформационной системы. С помощью этого спектра можно закодировать до 10^{15} различных биохимических реакций, что примерно в 10^7 раз больше количества реакций реально протекающих в клетке (их примерно 10^8), т.е. словарный запас клетки избыточен для эффективного распознавания, классификации, регулирования этих реакций в клетке. Количество информации на 1 квант энергии: $I = \log_2 10^{15} \approx 50$ бит. При делении клеток, количество энергии, расходуемой на передачу 50 бит информации равно энергии кванта (h - постоянная Планка, ν - частота излучения): $E = h\nu = 6.62 \times 10^{-27} (\text{эрг/сек}) \times 0.5 \times 10^{15} (\text{сек}^{-1}) = 3.3 \times 10^{-12} (\text{эрг})$. При этом, на 1 Вт мощности "передатчика" или на $\mu = 10^7$ эрг/сек. может быть передано количество квантов: $n = \mu/E = 10^7 (\text{эрг/сек}) / (3.3 \times 10^{-12} (\text{эрг})) \approx 3.3 \times 10^{18} (\text{квант})$. Общая скорость передачи информации на 1 Вт затрачиваемой клеткой мощности определяется по числу различных состояний клетки N и числу квантов (излучений) m : $V = n \log_2 N = 3.3 \times 10^{18} \times 50 \approx 1.6 \times 10^{20} (\text{бит/сек})$.

4. Понятие о связи информации и самоорганизации

Любая открытая информационная система эволюционирует так, что начиная с состояния наибольшей энтропии (неопределённости) стремится спиралеобразно к новым связям и отношениям, к организованности и порядку в системе в процессе взаимоотношений со средой и перестройки структуры с целью уменьшения энтропии.

Пример. На телевизионной игре "Что? Где? Когда?" обсуждение вопроса часто начинается хаотично, спонтанно, независимо и в конце обсуждения может organizоваться в единодушное принятие правильного решения.

Самоорганизация может наблюдаться и в неживых системах.

Пример. Эволюция ЭВМ - пример самоорганизации: от 1-го поколения ЭВМ (40-50-ые годы 19 века) с ненадёжными электронными лампами и быстродействием порядка 10^4 операций в сек. до 1-го поколения оптических вычислительных нейророботных структур (конец 90-ых годов) с голографической памятью, с логикой на потоках фотонов, быстродействием порядка 10^{12} операций в сек. и высокой надёжностью.

Сформулируем основные аксиомы теории информационных динамических процессов (информационной синергетики).

Аксиома 1. Развитие системы определяется некоторой целью и информационными ресурсами системы.

Аксиома 2. При стремлении к цели система воспринимает входную информацию, которая используется и для изменения внутренней структуры самой системы, внутрисистемной информации.

Аксиома 3. Изменение внутрисистемной информации происходит таким образом, чтобы уменьшалась энтропия (мера беспорядка) в системе.

Аксиома 4. Любое изменение внутрисистемной информации оказывает воздействие на выходную информацию системы (на окружающую среду).

Аксиома 5. Процесс актуализации информации структурирует окружающий нас мир. Все, что не познано в данный момент времени, образует "хаос", который заставляет актуализировать новую информацию, новые формы представления и описания знаний, приводит к появлению новых ветвей знания; этот хаос развивает при этом и исследователя.

Информация - это знание, которое используется для развития, совершенствования системы и её взаимодействий с окружающей средой.

Информация сама развивается вслед за развитием системы. Новые формы, принципы, подсистемы, взаимосвязи и отношения вызывают изменения в информации, ее содержании, формах получения, переработки, передачи и использования. Благодаря потокам информации система осуществляет целесообразное взаимодействие с окружающей средой, т.е. управляет или управляема. Своевременная и оперативная информация может позволить стабилизировать систему, адаптироваться, восстанавливаться при нарушениях структуры и/или подсистем. От степени информированности системы, от взаимодействия системы и среды зависит развитие и устойчивость системы.



Наблюдаемая математизация и математическая информатизация разделов современной науки показывает, что их эффективность зависит как от сложности и возможности описания её законов и принципов адекватными математическими и логико-информационными моделями, так и от используемого математического аппарата, выбранных мер информации, интеллектуальной деятельности, знаний, конструктивных методов и алгоритмов их измерения, оценивания информационных ресурсов.

Кроме указанных выше подходов к определению меры информации, есть и множество других (меры Винера, Колмогорова, Шрейдера и др.), но основными методами (в образовательной информатике) являются указанные. Рекомендуются для чтения [1-11] (приведены в хронологическом порядке).

5. Задачи для самостоятельного решения

В заключение предложим некоторые задачи для закрепления материала и усиления интереса при обучении рассмотренной выше проблеме.

Задача 1. ДНК человека можно представить себе как некоторое слово в четырехбуквенном алфавите $X=\{A, B, C, D\}$, где буквой помечается звено цепи ДНК (нуклеотид). Среднее число их равно примерно 1.5×10^{23} нуклеотид. Считая, что ядро каждой из примерно 10^{13} клеток человеческого тела является хранителем генетической информации, оценить объем информации в них (в теле человека).

Задача 2. Ресурсы человеческого мозга рассчитаны на переработку информации в 16 бит в сек. Какое количество информации перерабатывает человек за свою жизнь, если предположить что переработка информации идёт непрерывно в течении средней продолжительности жизни, которую принять равной 70 лет (заметим, что и во время сна человек перерабатывает информацию, более того, сон - продукт такой переработки информации).

Задача 3. Некоторая система может находиться в четырёх состояниях: в первом - с вероятностью 0.1, во втором и третьем - с вероятностью 0.5, в четвёртом - с вероятностью 0.4. Чему равно количество информации (или неопределённость выбора) в системе? Если система может находиться только в состоянии номер 2, то чему оно равно?

Задача 4. Система может принимать 128 различных равновероятных состояний. если состояние системы неизвестно, то каково количество информации в системе (неопределённость выбора)? Если известно, что система находится в состоянии номер 8, то чему равно количество информации?

Задача 5. Определить количество информации (в битах) в каждой букве русского алфавита, отождествив вероятность появления этой буквы в словах с частотой по частотному словарю русского языка приведённому ниже. Найти какое количество информации определяется двухбуквенными сочетаниями (всем алфавитом)? Оценить количество информации в каждом слове из букв «а», «б», «с» над алфавитом русского языка. Частотный словарь русского языка имеет вид:

Буква	Частота	Буква	Частота	Буква	Частота
о	0.090	к	0.028	ь, ъ, ѓ	0.014
е, е	0.072	м	0.026	ч	0.013
а, и	0.062	д	0.025	й	0.012
т, н	0.053	п	0.023	х	0.009
с	0.045	у	0.021	ж, ю, ш	0.006
р	0.040	я	0.018	ц, щ, э	0.003
в	0.035	ы, з	0.016	ф	0.002

Задача 6. Увеличиться или уменьшиться количество информации о(в) системе «Сосуд с водой» до и после замораживания воды. Как изменится энтропия этой системы. Ответ обоснуйте.



Задача 7. Передатчик генерирует независимо друг от друга последовательности букв из двух букв «а», одной буквы «б», трёх букв «в». Сколько таких всевозможных различных слов может генерировать передатчик? Какое количество информации несёт в себе каждое такое слово?

Задача 8. Прodelайте (достаточно грубый) эксперимент по определению вашего числа Страуда (числа мысленных различий в сек): напарник показывает или называет двадцать букв в определённой (случайной для Вас) последовательности, а Вы воспроизводите их максимально в этой же последовательности; число правильно воспроизведённых букв и есть Ваше число Страуда. Проверьте подчиняются ли эти числа (для нескольких испытуемых) закону типа Хартли.

Задача 9. Оцените грубо быстродействие (гипотетического) компьютера построенного на биочипах (молекулярном способе запоминания информации, например, с помощью выращенных специально молекул, способных окрашиваться в два устойчивых цвета - 0 и 1), если количество таких чипов считается заданным.

Задача 10. На стандартной клавиатуре компьютера имеется 46 клавиш (без управляющих, цифровых и функциональных). Каждая клавиша имеет два регистра и два значения (латинского алфавита и кириллицы). Оцените примерно объём информации, который позволяет вводить этот набор.

6. Наука **ИНФОРМАТИКА**

История информатики в нашей стране (сначала СССР, а затем России) насыщена драматическими коллизиями и резкими изменениями приоритетов. Это ощущается даже в терминологии. Термин "информатика" для обозначения совокупности научных направлений, тесно связанных с появлением компьютеров и их стремительным вхождением в ноосферу, определяемую жизнедеятельностью людей, у нас относительно новый. Он получил "права гражданства" в начале 80-х годов, а до этого, согласно определению, данному в Большой Советской энциклопедии, информатика рассматривалась как "дисциплина, изучающая структуру и общие свойства научной информации, а также закономерности ее создания, преобразования, передачи и использования в различных сферах человеческой деятельности" [1, с. 1031].

Подобное определение связывало информатику с библиотековедением, библиографией, методами поиска информации в массивах документов. Когда в 1952 году был создан Институт научной информации АН СССР, позже преобразованный в ВИНТИ - Всесоюзный институт научной и технической информации, то он должен был стать головным академическим учреждением в области информатики.

То, что стало называться информатикой в начале 80-х в нашей стране, было совершенно иным. Ближе всего содержание этого понятия подходит к тому, что в США и большинстве других стран называется "computer science", т. е. "компьютерные науки".

В [2] говорится, что "компьютерные науки" концентрируют свое внимание на различных аспектах, связанных с протеканием и использованием информационных процессов, с теми структурами, в которых представляется информация, и теми процедурами, которые используются при её переработке. Последнее связывает область "компьютерных наук" с теорией машин для переработки информации - компьютеров - и методами их использования в системах переработки информации.

Для термина "информатика" в [2] нет отдельной статьи, а есть лишь ссылки на термины: "компьютерные науки", "компьютерное обучение" и "информационные науки". Содержание понятия "информационные науки" в [2] полностью совпадает с толкованием термина "информатика" в БСЭ [1].

С начала 80-х содержание того, что скрывается за термином "информатика", ближе всего к тому, что понимают французы, когда говорят о науке, носящей название *informatique*.

До этого совокупность научных направлений, называемых теперь информатикой, именовалась по-разному. Сначала объединяющим названием был термин "кибернетика", затем на роль общего названия той же области исследований стала претендовать "прикладная математика". Следы этой разногласицы хорошо видны в наименовании высших учебных заведений и научных институтов. Факультет в МГУ, готовящий специалистов в области информатики, носит название "Вычислительная математика и кибернетика", а институты, ведущие исследования в данной области, могут называться и "Институт кибернетики".



Национальной АН Украины", и "Институт прикладной информатики РАН", и "Институт прикладной математики РАН".

Поэтому, говоря об истории информатики в бывшем СССР и теперешней России, по сути, надо излагать историю отечественной кибернетики и частично прикладной математики и вычислительной техники. Именно так построена эта вводная статья.

7. Структура информатики.

На протяжении полувековой истории информатики в ней неоднократно возникали и исчезали те или иные направления. В настоящее время ее структура, по-видимому, определилась. В нее входят следующие основные области исследования:

- теория алгоритмов (формальные модели алгоритмов, проблемы вычислимости, сложность вычислений и т. п.);
- логические модели (дедуктивные системы, сложность вывода, нетрадиционные исчисления: индуктивный и абдуктивный вывод, вывод по аналогии, правдоподобный вывод, немонотонные рассуждения и т. п.);
- базы данных (структуры данных, поиск ответов на запросы, логический вывод в базах данных, активные базы и т. п.);
- искусственный интеллект (представление знаний, вывод на знаниях, обучение, экспертные системы и т. п.);
- бионика (математические модели в биологии, модели поведения, генетические системы и алгоритмы и т. п.);
- распознавание образов и обработка зрительных сцен (статистические методы распознавания, использование призначных пространств, теория распознающих алгоритмов, трехмерные сцены и т. п.);
- теория роботов (автономные роботы, представление знаний о мире, децентрализованное управление, планирование целесообразного поведения и т. п.);
- инженерия математического обеспечения (языки программирования, технологии создания программных систем, инструментальные системы и т. п.);
- теория компьютеров и вычислительных сетей (архитектурные решения, многоагентные системы, новые принципы переработки информации и т. п.);
- компьютерная лингвистика (модели языка, анализ и синтез текстов, машинный перевод и т. п.);
- числовые и символьные вычисления (компьютерно-ориентированные методы вычислений, модели переработки информации в различных прикладных областях, работа с естественно-языковыми текстами и т. п.);
- системы человеко-машинного взаимодействия (модели дискурса, распределение работ в смешанных системах, организация коллективных процедур, деятельность в телекоммуникационных системах и т. п.);
- нейроматематика и нейросистемы (теория формальных нейронных сетей, использование нейронных сетей для обучения, нейрокомпьютеры и т. п.);



- использование компьютеров в замкнутых системах (модели реального времени, интеллектуальное управление, системы мониторинга и т. п.).

Эти области информатики возникли не одновременно. История информатики связана с постепенным расширением области ее интересов. Возможность расширения диктовалась развитием компьютеров и накоплением моделей и методов их применения при решении задач различного типа.