

1.1. Информатика и информация

1.1.1. Информатика

Задачи, связанные с хранением, передачей и обработкой информации человеку приходилось решать во все времена: требовалось передавать знания из поколения в поколения, искать нужные книги в хранилищах, шифровать секретную переписку. К концу XIX века количество документов в библиотеках стало настолько велико, что появилась необходимость как-то систематизировать накопленную информацию, для того чтобы было удобно ее хранить и искать нужные данные. В конце XIX века зародилось новое научное направление, в котором изучалась *документальная* информация, то есть информации в виде документов (книг, журналов, статей и т.п.). В английском языке оно получило название *«information science»* («информационная наука», «наука об информации»).

Применение компьютерной техники значительно увеличило возможности людей в области работы с информацией. Слово *«информатика»*¹ в современном значении образовано в результате объединения двух слов: «информация» и «автоматика». Таким образом, получается «автоматическая работа с информацией». В английском языке существует близкое по значению выражение *«computer science»* (наука о компьютерах).

Современная информатика, которая стала самостоятельной наукой в 1970-х годах, включает следующие научные направления:

- **теоретическую информатику** (теорию информации, теорию кодирования, математическую логику, теорию систем и др.);
- **кибернетику** (теорию управления в природе, технике и обществе);
- **искусственный интеллект** (распознавание образов, понимание речи, машинный перевод, логические выводы, алгоритмы самообучения);
- **вычислительную технику** (устройство компьютеров и компьютерных сетей);
- **программирование** (методы создания новых программ);
- **прикладную информатику** (персональные компьютеры, прикладные программы, информационные системы и т.д.).

Раньше эти вопросы частично рассматривались в других науках – математике, лингвистике (науке о языке), электронике и др.

В понятие «информатика» часто включают и *новые информационные технологии*, связанные с применением компьютеров. Они используются практически во всех областях современной жизни: для оформления документов; для подготовки книг и журналов к печати; для расчета заработной платы; для продажи билетов на поезда и самолеты; для автоматизации производства; при проектировании зданий, кораблей, станков и т.д. Кажется, что все это – совершенно разные сферы деятельности, однако они «связаны» в единое целое понятием *«информация»*.

1.1.2. Что такое информация?

Латинское слово *«informatio»* переводится как *«разъяснение», «сведения»*. В быту под информацией мы обычно понимаем любые сведения или данные об окружающем нас мире и о нас

¹ Впервые это слово использовал немецкий ученый К. Штейнбух в 1957 году (в немецком языке – *Informatik*). Затем оно было введено во французский язык (фр. *informatique*) и переведено на английский (англ. *informatics*).

самих. Однако дать определение информации весьма непросто. Более того, в каждой области знаний слово «информация» имеет свой смысл.

Философы говорят о том, что информация, как зеркало, *отражает* мир (реальный или вымышленный). Биологи чаще всего связывают понятие «информация» только с живой природой. Социологи изучают ценность и полезность информации в человеческом обществе. Специалистов по компьютерной технике в первую очередь интересует представление информации в виде знаков. Мы попробуем посмотреть на информацию с разных сторон и попытаться выявить некоторые ее свойства.

Давайте сравним два изображения одинакового размера (рисунок справа). На первом из них пусто, а на втором справа мы видим репродукцию с картины Рафаэля «Сикстинская мадонна». Вряд ли кто-то способен долго разглядывать чистый лист, в то же время можно подолгу смотреть на картину, открывая все новые и новые детали. Почему так?



Можно считать, что на первом рисунке нет информации, поэтому разглядывать его неинтересно. Там все одинаково – везде белый цвет. На втором рисунке есть *разнообразие*, на холст нанесены краски, и он перестал быть однородным.

Информация характеризует разнообразие (неоднородность) в окружающем мире.

Зачем вообще нам нужна информация? Дело в том, что наши знания всегда в чем-то неполны, в них есть *неопределенность*. Например, вы стоите на остановке автобуса и не знаете, на каком именно автобусе вам нужно ехать в гости к другу (его адрес известен). Неопределенность мешает вам решить свою задачу. Нужный номер автобуса можно определить, например, по карте с маршрутами транспорта. Очевидно, что при этом вы получите новую информацию, которая увеличила знание (уменьшила неопределенность).

Информация уменьшает неопределенность знаний.

Мы только что посмотрели на информацию с разных точек зрения. Многие выдающиеся ученые XX века (Н. Винер, У. Эшби, К. Шеннон, А. Урсул, А. Моль, В. Глушков) давали свое определение информации, но ни одно из них не стало общепринятым. И этому есть свое объяснение.

Дело в том, что мы всегда определяем новые понятия через уже известные. И, в конечном счете, приходим к понятиям, которые определить невозможно. Например, все термины в геометрии определяются через *базовые* понятия «*прямая*», «*точка*» и «*плоскость*».

В современной науке информация считается одним из базовых понятий, так же, как *материя* и *энергия*. Поэтому дать строгое определение информации не удастся, можно только объяснить значение этого слова на примерах и сравнить с другими понятиями. Норберт Винер, создатель *кибернетики* – науки об управлении и связи – писал: «Информация есть информация, а не материя и не энергия».

1.1.3. Виды информации

Человек получает информацию через свои органы чувств: глаза, уши, рот, нос и кожу. Поэтому всю получаемую нами информацию можно разделить на следующие виды:

- *зрительная информация* (визуальная, от англ. *visual*), которая поступает через глаза (по разным оценкам, 80-90% всей получаемой нами информации);
- *звуковая информация* (аудиальная, от англ. *audio*);
- *вкусовая информация* (вкус);
- *запахи* (обонятельная информация);

- *тактильная информация*, которую мы получаем с помощью осязания, «на ощупь».

Зафиксированная (закодированная) каким-то способом информация может быть представлена в различных формах:

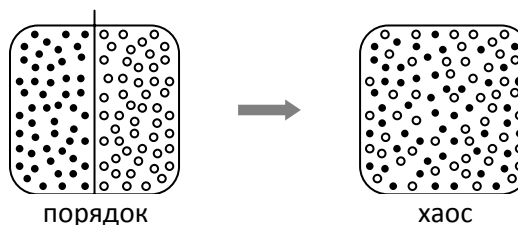
- *символ* (буква, цифра, знак) – самый простой вид информации;
- *текст*, который строится из символов; в отличие от набора символов, в тексте важен порядок их расположения, например, КОТ и ТОК – два разных текста, хотя они состоят из одинаковых символов;
- *числовая информация* (иногда ее не считают отдельным видом информации, полагая, что число – это текст специального вида, состоящий из цифр);
- *графическая информация* (рисунки, картины, чертежи, карты, схемы, фотографии);
- *звуковая информация* (звучание голоса, мелодии, шум, стук, шорох и т.п.);
- *комбинированная информация*, которая объединяет несколько видов информации (например, видеоинформация).

Обратим внимание, что одна и та же информация может быть представлена по-разному. Например, результаты измерения температуры в течение недели можно сохранить в виде текста, таблицы, графика, диаграммы, видеофильма и т.д.

1.1.4. Информация в природе

В науках, изучающих **неживую природу** (прежде всего, в физике), информацию связывают со сложностью объекта. Чем разнообразнее и сложнее объект, тем больше информации он содержит и тем большее количество знаков необходимо для того, чтобы его описать.

В замкнутых системах, которые не обмениваются веществом и энергией с окружающей средой, постепенно устанавливается состояние равновесия или хаоса. Все становится однородным, поэтому количество информации уменьшается. Например, если соединить две части сосуда, заполненного разными газами, эти газы постепенно перемешаются, и произойдет переход от порядка к хаосу².



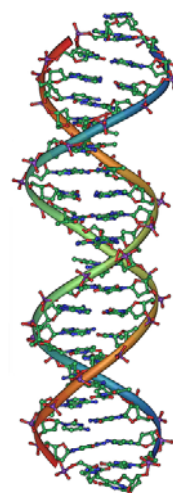
Порядок (и количество информации) не может увеличиваться «просто так», для этого необходим приток энергии и вещества. Системы, обладающие этим свойством, называются *открытыми*. Например, планета Земля – это открытая система, она получает энергию (и вместе с ней информацию) от Солнца. Согласно одной из теорий³, молекулы вещества – это природные ячейки памяти. Под действием солнечного света молекулы переходят в возбужденное состояние (в ячейки памяти записывается информация), это и служит причиной всех процессов, происходящих на Земле. При этом увеличивается сложность строения вещества (например, из атомов водорода и кислорода образуются молекулы воды). Это значит, что увеличивается также и количество накопленной информации.

² Если считать, что наша Вселенная – замкнутая система, можно сделать вывод о том, что ее ждет «тепловая смерть» – молекулы вещества равномерно распределятся в пространстве, и наступит полный хаос, развитие закончится. Однако современная теоретическая физика считает, что такого не произойдет благодаря действующим силам тяготения, которые заставляют Вселенную расширяться или сжиматься.

³ <http://thd.pnpi.spb.ru/~makariev/ec02.pdf>

Еще бóльшую роль играет информация в **живой природе**. Даже простейшие растения (сине водоросли) и животные (амебы) могут обрабатывать информацию о химическом составе и температуре окружающей среды и приспосабливаться к изменяющимся условиям. Более высокоразвитые животные обмениваются звуковой информацией (например, токование глухаря), зрительной (позы собаки, кошки), обонятельной (запахи, в том числе на больших расстояниях). Животные используют информацию на уровне инстинктов для того, чтобы выжить, избежать опасности, продолжить род.

Наследственная информация растений и животных, определяющая строение, внешний вид, предрасположенность к болезням, хранится и передается из поколения в поколение с помощью молекул ДНК (дезоксирибонуклеиновой кислоты). В этих молекулах информация закодирована в виде цепочек, составленных из четырех химических веществ (они называются *аденин*, *гуанин*, *тимин* и *цитозин*). Современный уровень развития биологии позволяет с помощью ДНК клонировать организмы, то есть создавать точные копии из одной клетки-образца.



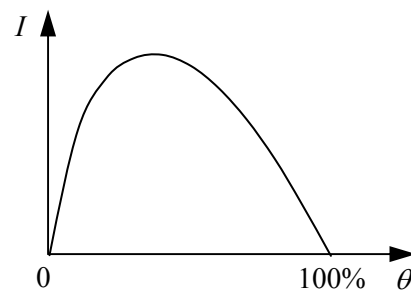
Молекула ДНК

1.1.5. Человек, информация, знания

Обо всех изменениях в окружающем мире человек узнает с помощью своих органов чувств: сигналы от них («первичная» информация) постоянно поступают в мозг. Там они обрабатываются с учетом уже имеющегося у этого человека опыта; он дополняет уже имеющиеся у него представления о мире, накапливает *знания* – свои представления о природе, обществе, самом себе. Знания позволяют человеку принимать решения, определяют его поведение и отношения с другими людьми.

Всегда ли полученная информация увеличивает наши знания? Очевидно, что нет. Например, информация о том, что $2 \cdot 2 = 4$ вряд ли увеличит ваши знания, потому что вы это уже знаете, эта информация не нова. Однако, она будет новой для тех, кто изучает таблицу умножения. Это значит, что изменение знаний при получении сообщения зависит от того, что человек знал до этого момента. Если он знает все, что было в полученном сообщении, знания не изменяются.

С другой стороны, сообщение о том, что «учет вибрационных взаимодействий континуализирует моделирование диссипативных структур» (или сообщение на неизвестном языке) также не увеличивает знания, потому что эта фраза вам непонятна. Иначе говоря, имеющихся знаний не хватило для того, чтобы воспринять новую информацию.



Эти идеи послужили основой *семантической* (смысловой) теории информации, предложенной в 1960-х годах советским математиком Ю.А. Шрейдером. На рисунке показано, как зависит количество полученных знаний I от того, какая доля информации θ в сообщении уже известна получателю.

Сообщение увеличивает знания человека, если оно понятно и содержит новые сведения.

К сожалению, «измерить» смысл информации, оценить его числом, довольно сложно. Поэтому для оценки количества информации используют другие подходы, о которых вы узнаете чуть позже.

Когда человек хочет поделиться с кем-то своим знанием, он может сказать «Я знаю, что...» или «Я знаю, как...». Это говорит о том, что есть два разных вида знаний. В первом случае знания –

это некоторый известный факт, например, «я знаю, что Луна вращается вокруг Земли». Такие знания называются **декларативными**, человек выражает их словами (*декларирует*). Декларативные знания – это факты, законы, принципы.

Второй тип знаний («Я знаю, как...») называют **процедурными**. Они выражаются в том, что человек *умеет* выполнять некоторые действия. Самое интересное, что объяснить эти знания бывает довольно сложно (например, попробуйте объяснить, как кататься на велосипеде).

Для того, чтобы сохранить знания и передать другим людям, нужно выразить их на каком-то языке (например, рассказать, записать, нарисовать и т.п.). Такие сведения можно хранить, обрабатывать, передавать, причем с этим может справиться и компьютер. В научной литературе информацию, зафиксированную (закодированную) в какой-то форме, называют *данными*, имея в виду, что компьютер может выполнять с ними какие-то операции, но не способен понимать смысл.

Данные – это просто какие-то зафиксированные сигналы, которые никак и никем не используются, не помогают решать какие-то задачи. Для того, чтобы данные стали информацией, их нужно понять и осмыслить, а на это способен пока только человек. Если человек, получающий сообщение, знает язык, на котором оно записано, он может понять смысл этого сообщения, то есть, получить информацию. Обработывая и упорядочивая информацию, он получает знания.

Мы увидели, что в науке существуют достаточно тонкие различия между понятиями «данные», «информация», «знания». Тем не менее, на практике чаще всего все это называется общим термином «информация».

1.1.6. Информация в обществе

Очень велика роль информации в человеческом обществе. Информация, получаемая нами из разных источников, позволяет принимать решения и во многом определяет всю нашу жизнь. Огромно влияние на общество средств массовой информации (СМИ) – газет, телевидения, изданий в Интернете. Их часто называют «четвертой властью», сравнивая с остальными тремя ветвями власти: законодательной, исполнительной и судебной.

В идеале информация должна быть

- *объективной* (не зависящей от чьего-либо мнения);
- *понятной* для получателя;
- *полезной* (позволяющей получателю решать свои задачи) ;
- *достоверной* (истинной, соответствующей реальной ситуации);
- *актуальной* (значимой в данный момент);
- *полной* (достаточной для принятия решения).

К сожалению, информация не всегда обладает всеми этими свойствами. Сообщение «В стакане мало молока» необъективно (для пессимиста полстакана – это мало, а для оптимиста – много). Сообщение 私は散歩に行った. непонятно для нас (оно означает «Я пошел гулять», только по-японски).

Полезность информации определяется для каждого человека в конкретной ситуации. Например, информация о том, как древние люди добывали огонь, для большинства городских жителей бесполезна, поскольку она никак не помогает им решать свои жизненные проблемы. С другой стороны, в экстремальной ситуации, когда человек оказывается один на один с природой, такие знания очень полезны, потому что сильно увеличивают шансы на выживание, то есть, помогают достичь цели.

Слухи, байки, искаженная информация (в том числе дезинформация) – это примеры недостоверной информации. Сообщение «10 лет назад тут был ларек с мороженым» неактуально, эта информация устарела. Информация «Сегодня будет концерт» неполна, потому что не указано

время и участники концерта, поэтому мы не можем принять решение (идти или не идти?). Необъективная, неполная и недостоверная информация нередко используется в СМИ для продвижения идей и интересов отдельных личностей и групп.

В будущем, по-видимому, нас ждет переход к информационному обществу, где большая часть населения будет заниматься сбором, обработкой и распространением информации, поэтому высказывание немецкого банкира Н. Ротшильда «Кто владеет информацией, тот владеет миром»⁴ становится актуально как никогда.

Развитие глобальной сети Интернет, в которую ежеминутно вносится огромное количество самых разнообразных данных, во многом перевернуло привычные представления о работе с информацией. Например, основным источником для поиска учебных материалов теперь фактически является Интернет, а не библиотеки. Однако при использовании информации из Интернета необходимо относиться к ней критически, так как ее достоверность никто не гарантирует.

1.1.7. Информация в технике

Практически все современные технические устройства (телевизоры, телефоны, стиральные машины, системы управления самолетами и судами) строятся на микропроцессорах, которые обрабатывают информацию: анализируют сигналы с датчиков, выбирают нужный режим работы. Широко используются *системы программного управления*, например, станки, обрабатывающие детали по программе, заложенной в памяти. Эту программу очень легко поменять и настроить станок на изготовление другой детали.

Во многих ситуациях, особенно при выполнении опасных, тяжелых и утомительных работ, человека могут заменить роботы, которые имеют датчики, заменяющие органы чувств. Одним из наиболее совершенных считается человекоподобный робот (*андрoид*) Asimo, разработанный фирмой *Honda*. Он умеет распознавать предметы, жесты, звуки, узнавать лица, разговаривать через домофон, передавать информацию через Интернет.



Робот Asimo

Наиболее универсальным устройством для обработки информации можно считать компьютер. Хотя современные компьютеры пока не умеют работать с вкусовой и обонятельной информацией (запахом), работы в этом направлении ведутся. Уже существуют экспериментальные приборы, названные «электронный нос»⁵ и «электронный язык»⁶; они построены на основе химических датчиков.

Сейчас в теоретической информатике считается, что компьютер может хранить и обрабатывать только *данные*, но не информацию. Многие ученые считают, что машина принципиально не может научиться понимать смысл информации и делать выводы. В пользу такой точки зрения свидетельствуют фактический провал проекта «компьютеров пятого поколения»⁷ (Япония, 1980-е гг.), в ходе которого планировалось создать машины, общающиеся с человеком на естественном языке. Тем не менее, ученые уделяют этим проблемам огромное внимание. Например, возникло целое научное направление *data mining* («добыча данных»), в котором изучаются методы извлечения информации («смысла», закономерностей, связей, знаний) из огромных наборов данных. В некоторых случаях действительно удается использовать огромные вычислительные

⁴ http://www.newsfril.com/internet/page_3.html

⁵ <http://www.rg.ru/2007/06/26/nos.html>

⁶ <http://www.business-magazine.ru/ideas/tech/pub282851>

⁷ http://ru.wikipedia.org/wiki/Компьютеры_пятого_поколения

мощности компьютеров для того, чтобы найти неизвестные ранее закономерности, которые можно использовать на практике.

Вопросы и задачи

1. Что изучает информатика?
2. Какие научные направления обычно включают в информатику?
3. Чем занимается кибернетика?
4. Что такое искусственный интеллект?
5. Как связана неопределенность наших знаний с получением информации?
6. Как связана информация и сложность объекта?
7. Объясните, почему термин «информация» трудно определить?
8. Какие еще базовые понятия науки вы знаете?
9. Основоположителем какого научного направления был Норберт Винер?
10. Согласны ли вы с определением информации, которое дал Н. Винер?
11. Как человек воспринимает информацию?
12. Как можно классифицировать информацию? Какие существуют подходы?
13. Чем отличается текст от набора символов?
14. Почему числовую информацию иногда не выделяют как отдельный вид?
15. К какому виду информации относятся видеофильмы?
16. Что такое тактильная информация?
17. Что такое «информация» в неживой природе?
18. В чем разница между замкнутыми и открытыми системами?
19. Что такое «тепловая смерть Вселенной»? Что думают современные физики по этому поводу?
20. Какую роль играет информация в живой природе?
21. Как передается наследственная информация растений и животных?
22. Как кодируется наследственная информация в молекуле ДНК?
23. Что такое «клонирование»?
24. Как вы думаете, почему клонирование человека запрещено во многих странах, в том числе и в России?
25. Всякая ли информация увеличивает знания? Почему?
26. Что такое семантическая теория информации?
27. Приведите примеры своих декларативных и процедурных знаний.
28. В чем, на ваш взгляд, разница между понятиями «данные», «информация», «знания»?
29. Почему считают, что компьютер может работать только с данными?
30. Какова роль информации в жизни общества?
31. Какими свойствами должна обладать «идеальная» информация?
32. Приведите примеры необъективной, непонятной, бесполезной, недостоверной, неактуальной и неполной информации.
33. Может ли информация быть достоверной, но бесполезной? достоверной, но необъективной? объективной, но недостоверной? актуальной, но непонятной?
34. Какое общество называют информационным?
35. Какие изменения произошли в жизни общества в результате широкого распространения Интернета?
36. Приведите примеры анализа и обработки информации в технических устройствах.
37. Что умеет робот *Asimo*? Какую информацию он обрабатывает?
38. Что такое «электронный нос» и «электронный язык»?

39. Как вы считаете, смогут ли компьютеры научиться понимать смысл данных?

1.2. Информационные процессы

Можно выделить **три основных информационных процесса**:

- *хранение информации,*
- *обработка информации,*
- *передача информации.*

Иногда информационными процессами называют также и многие другие операции, которые можно сделать с информацией (например, получение, представление, копирование, удаление и др.), но они, в конечном счете, сводятся к трем названным процессам.

1.2.1. Хранение информации

Для **хранения** информации человек, прежде всего, использует свою память. Можно считать, что мозг – это одно из самых совершенных хранилищ информации, во многом превосходящее компьютерные средства. Для запоминания и поиска информации используются нервные клетки мозга – *нейроны*, их более ста миллиардов.

К сожалению, человек многое забывает. Поэтому в древности он записывал информацию на камне, папирусе, бересте, пергаменте, а сейчас – на бумаге, магнитной ленте, электронных носителях. Это нужно еще и для того, чтобы передать знания другим людям, в том числе и следующим поколениям.

В XX веке появились новые средства хранения информации: перфокарты и перфоленты, магнитные ленты и магнитные диски, лазерные диски, флэш-память.

1.2.2. Обработка информации

Обработка – это любое изменение информации, причем изменяться может как содержание информации, так и ее форма.

Можно выделить четыре важнейших вида обработки:

- *создание новой информации*, например, решение задачи с помощью вычислений или логических рассуждений;
- *кодирование*, когда меняется *форма* (внешний вид), но не содержание информации; например, перевод текста на другой язык; один из видов кодирования – шифрование, цель которого – скрыть смысл (содержание) информации от посторонних;
- *поиск информации*, например, в книге, в библиотечном каталоге, на схеме или в Интернете;
- *сортировка* – расстановка элементов списка в заданном порядке, например, расстановка чисел по возрастанию или убыванию, расстановка слов или фамилий по алфавиту; одна из задач сортировки – облегчить поиск информации.

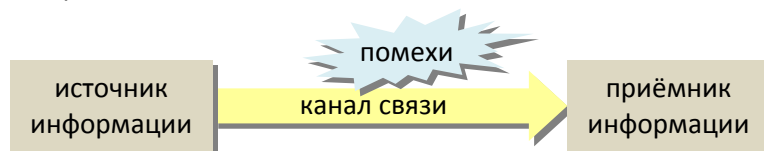
Для обработки информации человек использует, в первую очередь, свой мозг. Нейроны коры головного мозга «переключаются» примерно 200 раз в секунду – значительно медленнее, чем элементы памяти компьютеров. Однако человек практически безошибочно отличает собаку от кошки, а для компьютеров эта задача пока не разрешима. Дело, по-видимому, в том, что мозг решает такие задачи не «в лоб», не путем сложных вычислений, а как-то иначе (как – пока никто до конца не знает).

Компьютер позволяет «усилить» возможности человека в тех задачах обработки информации, решение которых требует длительных расчетов по известным алгоритмам. Однако, в отличие от человека, компьютер не может «мыслить» образами, поэтому для него недоступно фантазии, размышления, творчество.

1.2.3. Передача информации

При **передаче информации** всегда есть два объекта – источник и приемник информации. Эти роли могут меняться, например, во время диалога каждый из участников выступает то в роли источника, то в роли приемника информации.

Информация проходит от источника к приемнику через *канал связи*. При разговоре людей – это воздух, в котором распространяются звуковые волны. В компьютерах информация передается с помощью электрических сигналов или радиоволн (в беспроводных устройствах). Информация может передаваться с помощью света, лазерного луча, системы телефонной или почтовой связи, компьютерной сети и др.



Информация поступает по каналу связи в виде сигналов, которые приемник может обнаружить с помощью своих органов чувств (или датчиков) и «понять» (раскодировать). Сигнал – это любое изменение в окружающей среде, которое можно как-то зафиксировать, например, звуковые колебания, радиоволны, вспышки света, изменение напряжения на контактах.

К сожалению, в реальном канале связи всегда действуют помехи: посторонние звуки при разговоре, шумы радиоэфира, электрические и магнитные поля. Помехи могут полностью или частично исказить информацию, вплоть до полной потери (вспомните телефонные разговоры при перегрузке сети).

Чтобы содержание сообщения, искаженного помехами, можно было восстановить, оно должно быть *избыточным*, то есть, в нем должны быть «лишние» элементы, без которых смысл все равно восстанавливается. Например, в сообщении «Влг впдт в Кспск мр» многие угадают фразу «Волга впадает в Каспийское море», из которой убрали все гласные. Этот пример говорит о том, что естественные языки содержат много «лишнего», их избыточность оценивается в 60-80%.

? Вопросы и задачи

1. К каким из основных информационных процессов можно, на ваш взгляд, отнести получение, представление, копирование, удаление информации?
2. Зачем человек записывает информацию?
3. В чем преимущества и недостатки человеческой памяти в сравнении с компьютерной?
4. В каких задачах компьютер не может соревноваться с человеком? Почему? В каких ситуациях человек явно уступает компьютеру?
5. Какие средства хранения информации используются в компьютерной технике? Какие из них уже вышли или выходят из употребления? Почему?
6. Какие основные виды обработки информации вы знаете?
7. При каких видах обработки информации меняется ее содержание? форма?
8. При каких видах обработки информации меняется только ее форма представления?
9. К какому виду обработки можно отнести шифрование? Почему?
10. Работники удаленной метеостанции каждый 3 часа измеряют температуру и влажность воздуха, и передают данные по радию в районный метеоцентр. Там эти данные сводят в таблицу и отправляют по электронной почте в Гидрометцентр, где мощные компьютеры составляют прогноз погоды. Выделите здесь информационные процессы.

11. Вася Пупкин нашел в старой книге сведения о населении Москвы в XIX веке, составил таблицу по этим данным, построил диаграмму и сделал доклад на школьной конференции. Выделите здесь информационные процессы.
12. Кто (что) может быть источником (приемником) информации? Приведите примеры.
13. Какие каналы связи могут использоваться для передачи информации? Приведите примеры из жизни, из литературных произведений.
14. Что такое сигнал? Приведите примеры сигналов.
15. Как приемник «понимает», что информация пришла?
16. Расскажите, как помехи влияют на передачу информации. Приведите примеры.
17. Что такое избыточность? Почему она полезна при передаче информации?
18. Представьте, что придумали язык, в котором нет избыточности. В чем будет его недостаток?
19. Как вы думаете, какой вариант русского языка обладает наибольшей избыточностью: разговорный, литературный, юридический, язык авиадиспетчеров? (Оценки⁸: 72%, 76%, 83%, 96%)

1.3. Измерение информации

Любая наука рано или поздно приходит к необходимости как-то измерять то, что она изучает. Измерение информации – это одна из важнейших задач теоретической информатики.

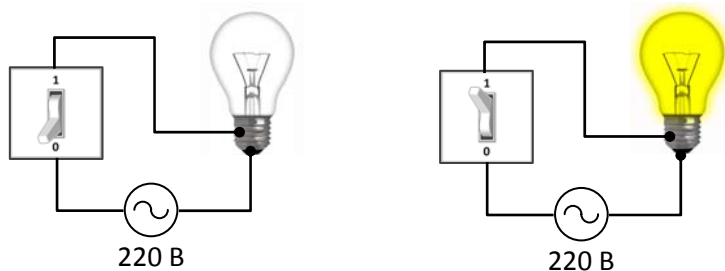
Для человека информация – это, прежде всего, смысл, заключенный в сигналах и данных. Как измерить смысл? На этот вопрос пока нет однозначного ответа.

Вспомним, что компьютеры не могут обрабатывать смысл, они работают только с данными. При этом возникают чисто практические задачи – определить, сколько места займет на диске текст, рисунок или видеофильм; сколько времени потребуется на передачу файла по компьютерной сети и т.п. Поэтому чаще всего используется *объемный* подход к измерению информации. Он заключается в том, что количество информации оценивается просто по числу символов, используемых для ее кодирования. С этой точки зрения стихотворение А.С. Пушкина и случайный набор букв могут содержать одинаковое количество информации. Конечно, этот подход далек от идеала, но он позволяет успешно решать все задачи, связанные с компьютерной обработкой и хранением данных.

1.3.1. Что такое бит?

Чтобы измерить информацию, нужно выбрать какую-то единицу измерения, эталон. В качестве такого эталона принимают информацию, полученную при выборе одного из двух вариантов.

Например, электрическая лампочка может находиться в двух состояниях: «горит» и «не горит». Тогда на вопрос «Горит ли сейчас лампочка» есть два возможных варианта ответа, которые можно обозначить цифрами 1 («горит») и 0 («не горит»). Поэтому ответ на этот вопрос может быть записан как 0 или 1⁹.



⁸ http://habrahabr.ru/blogs/popular_science/50643/

⁹ Конечно, вместо 0 и 1 можно использовать два любых знака.

Цифры 0 и 1 называют *двоичными* (в отличие, например, от набора десятичных цифр), и с этим связано название единицы измерения количества информации – *бит*. Английское слово *bit* – это сокращение от выражения *binary digit*, «двоичная цифра». Впервые слово *бит* в этом значении использовал американский инженер и математик Клод Шеннон в 1948 г.

Бит – это количество информации, соответствующее выбору одного из двух равновероятных вариантов.

Например, в сообщении «подброшенная монета упала гербом» содержится 1 бит информации, потому что до опыта было два возможных варианта: монета могла упасть гербом или «решкой». Ответ на вопрос «Дверь открыта?» тоже содержит 1 бит, если считать, что возможны только два ответа, «да» или «нет». Вот пример диалога, в котором получена информация в 1 бит:

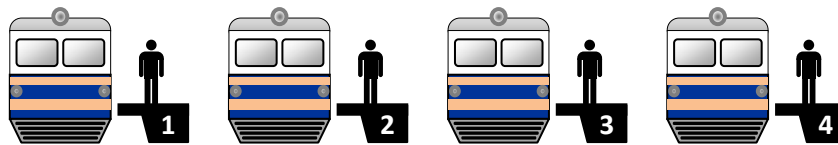
– Вы будете чай или кофе?

– Кофе, пожалуйста.

Слово «равновероятных» означает, что ни одному из этих вариантов заранее нельзя отдать предпочтение (как говорят математики, они *равновероятны*). Это условие будет нарушено, например, когда летом в Воронеже кто-то отвечает на вопрос «Идет ли сейчас снег?» (можно почти с полной уверенностью заранее сказать, что летом в Воронеже снега нет).

1.3.2. 2 бита, 3 бита, ...

А если возможных вариантов не два, а больше? Понятно, что в этом случае информация будет больше, чем 1 бит. Представим себе, что на вокзале стоят 4 одинаковых поезда, причем только один из них идет в Москву. В справочном окне сидит робот, который умеет отвечать на вопросы, но может сказать только «да» или «нет». Как за наименьшее количество вопросов найти поезд в Москву?



Понятно, что можно перебрать все поезда, задавая вопросы типа «Поезд на платформе 1 идет в Москву?» и т.д. В лучшем случае мы угадаем с первого раза, а в худшем (если нужный поезд стоит на платформе 3 или 4) придется задать три вопроса.

Однако есть другой способ, который позволяет за два вопроса гарантированно найти поезд в Москву. Разделим все поезда на две равные группы (например, включим в первую поезда на платформах 1 и 2, а во вторую – все остальные) и спросим: «В первой группе есть поезд в Москву?». В ответ мы получим 1 бит информации, поскольку сделан выбор из двух равновероятных вариантов. При любом ответе («да» или «нет») количество вариантов уменьшается вдвое. Теперь останется одним вопросом выбрать один из двух оставшихся поездов, при этом мы тоже получим 1 бит информации. Таким образом, всего за 2 ответа получено 2 бита информации¹⁰.

Теперь предположим, что в справочном окне работает человек, способный дать полный ответ. Тогда на вопрос «Какой поезд идет в Москву?» он может ответить, например, «В Москву идет поезд с платформы № 2». Сколько информации мы получили в этом сообщении? Очевидно, что мы не узнали ничего нового в сравнении с первым случаем (когда робот отвечал «да» или «нет»). Поэтому количество информации будет равно 2 бита, независимо от того, как мы ее получили.

¹⁰ Важно, что мы каждый раз разбиваем оставшиеся поезда на две равные группы. Если это не так, количество информации в битах не будет совпадать с количеством ответов «да»–«нет», потому что варианты не будут равновероятными. Нужный нам поезд, скорее всего, окажется в той группе, где поездов больше.

Таким образом, при выборе одного из четырех вариантов (любым способом) мы получаем 2 бита информации.

Для большего количества вариантов тоже можно применить принцип «деления на два». При поиске одного из 8 вариантов мы сначала выбираем «четверку», где находится нужный вариант (1 бит информации), потом – пару (еще один бит), а затем находим сам этот вариант (третий бит). Поэтому всего мы получаем 3 бита информации.

Таким образом, при каждом удвоении количества возможных вариантов мы должны задать дополнительный вопрос и, следовательно, получить еще один бит информации. При выборе одного из 16 вариантов мы получаем 4 бита информации, при выборе из 32 вариантов – 5 бит и т.д. (см. таблицу).

<i>I</i> , бит	1	2	3	4	5	6	7	8	9	10
<i>N</i> , вариантов	2	4	8	16	32	64	128	256	512	1024

Наверное, вы заметили, что все числа в нижней строчке таблицы – это степени числа 2.

Осталось выяснить, чему равно количество информации, если выбор делается, скажем, из 6 возможных вариантов (или **вообще** для любого числа, не являющегося степенью числа 2). Несложно догадаться, что это не целое число. Например, при выборе из 6 вариантов количество информации будет между 2 и 3 битами. Более точно мы определим это значение в 11 классе, когда вы изучите логарифмы.

1.3.3. Другие единицы

Считать большие объемы информации в битах неудобно, хотя бы потому, что придется работать с очень большими числами (миллиардами, триллионами, и т.д.). Поэтому стоит ввести более крупные единицы.

Измерение количества информации тесно связано с устройством компьютерной памяти. Память строится из элементов, которые могут находиться в двух состояниях (0 или 1, включено или выключено). Поэтому информация о состоянии такого элемента равна 1 биту. Чтобы обращаться к ячейкам памяти, нужно каждой из них присвоить адрес (номер). Если каждый отдельный бит будет иметь свой адрес, адреса будут очень большие, и для их хранения потребуется много места. Кроме того, реальные данные состоят из нескольких битов, и каждый раз «собирать» число или символ из нескольких отдельных ячеек памяти неудобно. Поэтому группы соседних битов памяти объединяют в ячейки, каждая из которых имеет свой адрес и считывается (или записывается) как единое целое¹¹. Такие ячейки называются **байтами**.

Байт – это группа битов, имеющая собственный адрес в памяти. В современных компьютерах¹²

1 байт = 8 бит

Слово «байт» (англ. *byte*) впервые использовал американский инженер В. Бухгольц в 1956 г. для обозначения наименьшего объема данных, который компьютер может «откусить» (англ. *bite*) за один раз.

¹¹ Компьютеры первых поколений обрабатывали только числа, для хранения которых выделялись ячейки размером от 24 до 60 бит (они часто назывались *машинными словами*). Позднее, когда на компьютерах начали обрабатывать в основном текстовую информацию, большие ячейки стало неудобно использовать, так как приходилось «разбирать» слова на отдельные символы («байты»). Поэтому изменили систему адресации, дав каждому байту свой адрес.

¹² Так было не всегда: см. <http://ru.wikipedia.org/wiki/Байт>.

Байт – тоже очень небольшая единица количества информации. Для того, чтобы не работать с большими числами, часто используют более крупные единицы, образованные с помощью приставок:

1 Кбайт (килобайт) = 1024 байта = 2^{10} байта = 2^{13} бит

1 Мбайт (мегабайт) = 1024 Кбайта = 2^{10} Кбайта = 2^{20} байта = 2^{23} бит

1 Гбайт (гигабайт) = 1024 Мбайта

1 Тбайт (терабайт) = 1024 Гбайта

Так сложилось исторически, что при измерении количества информации приставка «кило-» обозначает, в отличие от международной системы единиц СИ, увеличение не в 1000 раз, а в $1024 = 2^{10}$ раз. Аналогично «мега-» – это увеличение в $1024^2 = 2^{20} = 1048576$ раз, а не в 1 млн = 1000^2 раз.

Строго говоря, нужно называть такие кило-(мега-, гига-, ...)байты *двоичными*, поскольку множитель 1024 – это 2^{10} . Стандарт Международной электротехнической комиссии (МЭК) предлагает называть их «кибибайт», «мебибайт», «гибибайт» и «тебибайт», но эти названия не прижились.

Для перевода количества информации из одних единиц в другие нужно использовать приведенные выше соотношения. При переводе из крупных единиц в мелкие числа умножают на соотношение между единицами. Например,

$$2 \text{ Кбайта} = 2 \cdot (1 \text{ Кбайт}) = 2 \cdot 1024 \text{ байта} = 2048 \text{ байт}$$

$$= 2048 \cdot (1 \text{ байт}) = 2048 \cdot 8 \text{ бит} = 16384 \text{ бита}$$

$$= 2 \cdot 2^{10} \text{ байт} = 2^{11} \text{ байт} = 2^{11} \cdot 2^3 \text{ бит} = 2^{14} \text{ бит}$$

В последней строке все расчеты сделаны через степени числа 2, очень часто так бывает проще.

При переводе из мелких единиц в крупные нужно делить на соотношение между единицами. Например,

$$8192 \text{ бита} = 8192 \cdot (1/8 \text{ байта}) = 8192 : 8 \text{ байт} = 1024 \text{ байта}$$

$$= 1024 \cdot (1/1024 \text{ Кбайта}) = 1024 : 1024 \text{ Кбайт} = 1 \text{ Кбайт}$$

$$= 2^{13} \text{ бита} = 2^{13} \cdot (1/2^3 \text{ байта}) = 2^{10} \text{ байт}$$

$$= 2^{10} \cdot (1/2^{10} \text{ Кбайта}) = 1 \text{ Кбайт}$$

? Вопросы и задачи

1. Какую информацию принимают за единицу измерения количества информации?
2. Приведите примеры информации, равной 1 биту.
3. Что такое двоичные цифры?
4. Что такое равновозможные варианты? Приведите примеры, когда этой свойство нарушено.
5. Вася Пупкин не знает, какой из 8 поездов, стоящих на вокзале, идет в Санкт-Петербург. В справочном бюро он задает 8 вопросов: «Поезд на 1-ой платформе идет в Санкт-Петербург?», «Поезд на 2-ой платформе идет в Санкт-Петербург?» и т.д. На первые 7 вопросов он получает ответ «нет», а на последний – «да». Вася считает, что он получил 8 бит информации. Прав он или нет? Почему? (Ответ: не прав, в самом деле, он получил 3 бита информации)
6. Объясните, почему все числа во второй строке таблицы в разд. 1.3.2 – это степени числа 2.
7. В зоопарке содержится 10 обезьян, причем одна из них – альбинос (вся белая). Обезьяны сидят в двух вольерах, в первом – 8 штук, а во втором – две. Вася Пупкин считает, что сообщение «Обезьяна-альбинос сидит во втором вольере» содержит 1 бит информации. Прав он или нет? (Ответ: нет, варианты не равновероятны.)

8. На лугу пасутся 12 коров. Дедушка говорит Васе Пупкину, приехавшему из города в гости: «Вон та, крайняя справа – это наша Зорька». Сколько информации (в битах) получил Вася? (Ответ: больше 3, но меньше 4 бит)
9. Известно, что дверь открывается двумя из 4-х имеющихся ключей. Оцените количество информации в сообщении «Дверь открывается ключами № 2 и № 4». (Ответ: всего есть 6 вариантов, поэтому количество информации больше 2, но меньше 3 бит)
10. Известно, что дверь открывается двумя из 5 имеющихся ключей. Оцените количество информации в сообщении «Верхний замок открывается ключом № 1, а нижний – ключом № 4». (Ответ: всего есть 20 вариантов, поэтому количество информации больше 4, но меньше 5 бит)
11. Вася Пупкин задумал число от 1 до 100. Нужно отгадать это число за наименьшее число попыток, задавая Васе вопросы, на которые он отвечает только «да» и «нет». За сколько вопросов вы беретесь угадать число? Как нужно задавать вопросы, чтобы их число было минимальным даже в худшем случае?
12. Вася Пупкин задумал число от 20 до 83. Сколько бит информации содержится в сообщении «Вася задумал число 77»? (Ответ: 6 бит)
13. Двое играют в «крестики-нолики» на поле 4 на 4 клетки. Какое количество информации получил второй игрок, узнав ход первого игрока? (Ответ: 4 бита)
14. Что такое байт? Всегда ли в байте 8 бит?
15. Чему равен байт в современных компьютерах?
16. Какие единицы используют для измерения больших объемов информации?
17. Что означают приставки «кило-», «мега-», «гига-» и «тера-» при измерении количества информации?
18. Какие приставки рекомендует МЭК для обозначения двоичных килобайта и мегабайта?
19. Переведите 1 Мбайт во все изученные единицы измерения количества информации.
20. Переведите 2^{26} бит во все изученные единицы измерения количества информации.
21. Сколько килобайт содержится в 32768 битах?
22. Сколько бит в 8 Кбайтах?
23. Сколько бит содержит 1/16 Кбайта?
24. Сколько бит содержит 1/512 Мбайта?

1.4. Структура информации

1.4.1. Зачем структурировать информацию?

Давайте сравним четыре информационных сообщения.

Первое:

«Для того, чтобы добраться до села Васино, нужно сначала долететь на самолете до Ивановска. Затем на электричке доехать до Ореховска. Там на пароме переправиться через реку Слоновую в поселок Ольховка, и оттуда ехать в Васино на попутной машине».

Второе:

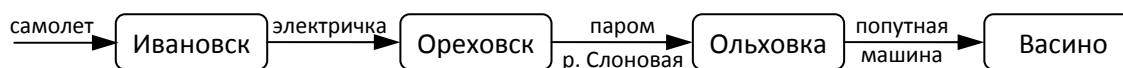
Как ехать в Васино?

- 1) *На самолете до Ивановска.*
- 2) *На электричке до Ореховска.*
- 3) *На пароме через р. Слоновую в пос. Ольховка.*
- 4) *На попутной машине до с. Васино.*

Третье:

Откуда	Куда	Транспорт
Москва	Ивановск	самолет
Ивановск	Ореховск	электричка
Ореховск	пос. Ольховка	паром (р. Слоновая)
пос. Ольховка	с. Васино	попутная машина

Четвертое:



Можно считать, что все эти (такие разные по форме!) сообщения содержат одну и ту же информацию. Какие из них проще воспринимать? Очевидно, что «вытащить» полезную информацию из простого текста (первое сообщение) сложнее всего. Во втором случае мы сразу видим все этапы поездки. Третье сообщение (таблицу) и четвертое (схему) можно понять сразу, с первого взгляда. Второй, третий и четвертый варианты лучше и быстрее воспринимаются, потому что в них выделена *структура* информации, в которой самое главное – этапы поездки в Васино.

Структура (лат. *structura* — строение) – это внутреннее устройство чего-либо.

Почему книгу разбивают на главы и разделы, а не пишут сплошной текст? Зачем в тексте выделяют абзацы? Прежде всего, для того, чтобы подчеркнуть основные мысли – каждая глава, раздел, абзац содержат определенную идею. При таком выделении структуры улучшается передача информации от автора к читателю.

Кроме того, есть еще и другая причина – облегчить **поиск** нужной информации. В книгах чаще всего есть оглавление, позволяющее быстро найти нужный раздел. Слова в словарях всегда расставлены в алфавитном порядке (представьте себе, что было бы, если бы они были расположены произвольно!). В больших книгах используют *индексы* – списки основных терминов с указанием страниц, на которых они встречаются.

Оглавление:

1. Информация	5
1.1 Что такое информация?	6
1.2 Виды информации	8
1.3 Информация в природе ...	10
1.4 Информация в технике	11
2. Измерение информации	12
2.1 Что такое бит?	13
2.2 Байт и другие единицы	14

Словарь:

автомат – <i>automaton</i>
автор – <i>author</i>
адрес – <i>address</i>
алгебра – <i>algebra</i>
алгоритм – <i>algorithm</i>
архив – <i>archive</i>
архитектура – <i>architecture</i>
асимметрия – <i>asymmetry</i>

Индекс:

А
аксиома 45
алгоритм 30, 78
архиватор 125
Б
бит 5, 15, 25, 43
брандмауэр 112
браузер 322

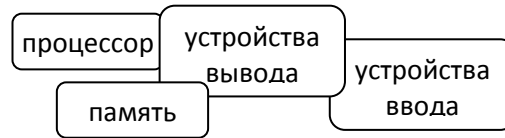
Структурирование – это выделение важных элементов в информационных сообщениях и установление связей между ними. Цели структурирования – облегчение восприятия и поиска информации, выявление закономерностей.

1.4.2. Простые структуры

Простейшая структура – это **множество**, то есть, некоторый набор элементов. Чтобы определить множество, мы должны перечислить все его элементы (например, множество, состоящее из Васи, Пети и Коли) или определить характерный признак, по которому элементы включаются в это множество (множество драконов с пятью зелеными хвостами).

Множество может состоять из конечного числа элементов (множество букв русского алфавита), бесконечного числа элементов (множество натуральных чисел) или вообще быть пустым (множество слонов, живущих на Северном полюсе). В документах конечное множество часто оформляют в виде маркированного списка, например:

- процессор;
- память;
- устройство ввода;
- устройства вывода.

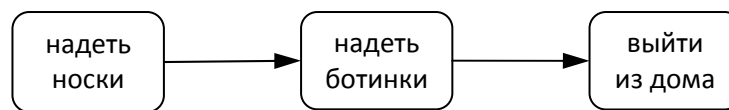


В таком списке порядок элементов не важен, от перестановки элементов множество не меняется.

Если множество состоит из конечного числа элементов и его элементы должны быть расположены в строго определенном порядке, мы получаем **линейный список**. Список обычно упорядочен (отсортирован) по какому-то правилу, например, по алфавиту, по важности, по последовательности действий и т.д. В тексте он оформляется как нумерованный список, например:

- 1) надеть носки;
- 2) надеть ботинки;
- 3) выйти из дома.

Переставить местами элементы такого списка нельзя (это будет уже другой список). Линейный список может быть представлен в виде цепочки связанных элементов:

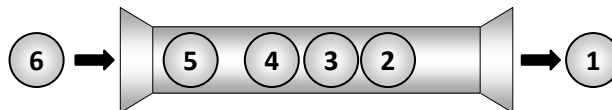


Список можно задать перечислением элементов, с первого до последнего:

(надеть носки, надеть ботинки, выйти из дома)

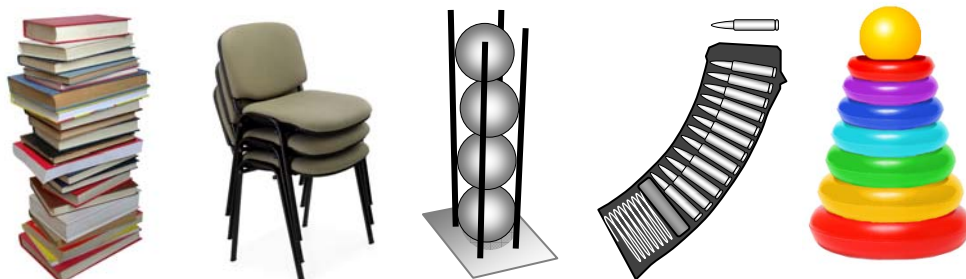
Теперь предположим, что нам нужно добавлять и удалять элементы в линейном списке. Знакомый вам пример – это *очередь* в кассу в магазине. Действительно, очередь – это цепочка, в которой элементы нельзя переставлять местами (если так случается, то кто-то лезет без очереди).

Очередь – это линейный список, в котором элементы добавляются с одного конца, а удаляются с другого («первым пришел – первым ушел»¹³).



Очередь можно представить в виде трубы, с одного конца которой добавляются шарики, а с другого – вынимаются. Трамваи, стоящие на перекрестке, – это тоже пример очереди, они не могут обогнать друг друга.

Возможен и другой вариант, когда элементы добавляются и удаляются с одного и того же конца списка. Это значит, что тот, кто пришел последним, уйдет первым. Такая структура называется **стек**. Например, стопка с книгами или автоматный магазин:



Во всех этих примерах для того, чтобы добраться до нужного объекта, нам нужно сначала удалить все те, что расположены выше него.

¹³ Англ. FIFO = *First In – First Out*.

Стек – это линейный список, в котором элементы добавляются и удаляются только с одного конца («последний пришел – первым ушел»¹⁴).

Более общий случай – это список, в котором добавление и удаление элементов разрешается с обоих концов. Такая структура называется **дек**.

Еще одна знакомая вам структура – **таблица**. С помощью таблиц устанавливается связь между несколькими элементами. Например, в следующей таблице элементы в каждой строке связаны между собой – это свойства некоторого объекта (человека):

Фамилия	Имя	Рост, см	Вес, кг	Год рождения
Иванов	Иван	175	67	1996
Петров	Петр	164	70	1998
Сидоров	Сидор	168	63	2000

Именно так хранится информация в базах данных: строка таблицы, содержащая информацию об одном объекте, называется *записью*, а столбец (название свойства) – *полем*.

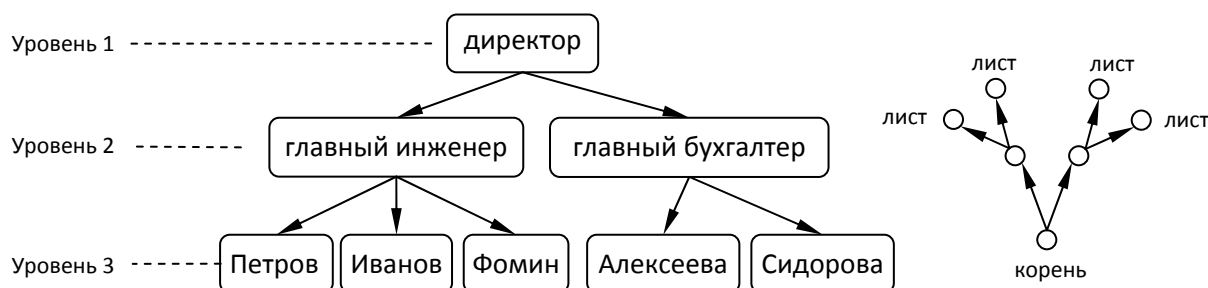
Возможен и другой вариант таблицы, когда роли строк и столбцов меняются. В первом столбце записываются названия свойств, а данные в каждом из следующих столбцов описывают свойства какого-то объекта. Например, вот таблица с характеристиками разных марок автомашин:

Марка	Лада Приора	Лада Калина	ВАЗ 2110	ВАЗ 21099
Мощность, л.с.	98	89	89	81
Максимальная скорость, км/ч	183	165	187	167,5
Время разгона до 100 км/ч, с	12	12,5	12	13,5

В математике и программировании таблицы обычно называют *матрицами*.

1.4.3. Иерархия (дерево)

Линейных списков и таблиц иногда недостаточно для того, чтобы представить все связи между элементами. Например, в некоторой фирме есть директор, ему подчиняются главный инженер и главный бухгалтер, у каждого из них есть свои подчиненные. Если мы захотим нарисовать схему управления этой фирмы, она получится *многоуровневой*.



Такая структура, в которой одни элементы «подчиняются» другим, называется *иерархия* (от древнегреческого *ἱεραρχία* – «священное правление»). В информатике иерархию называют **деревом**. Дело в том, что если перевернуть эту схему вверх ногами, она становится похожа на дерево (точнее, на куст, см. рисунок справа). Несколько деревьев образуют **лес**.

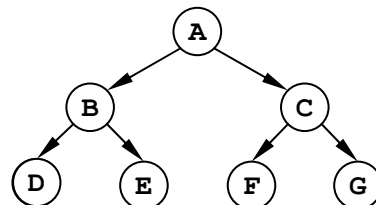
Дерево состоит из узлов и связей между ними (они называются дугами). Самый первый узел, расположенный на верхнем уровне (в него не входит ни одна стрелка-дуга) – это *корень дерева*. Конечные узлы, из которых не выходит ни одна дуга, называются *листьями*. Все остальные узлы, кроме корня и листьев – это промежуточные узлы.

¹⁴ Англ. LIFO = *Last In – First Out*.

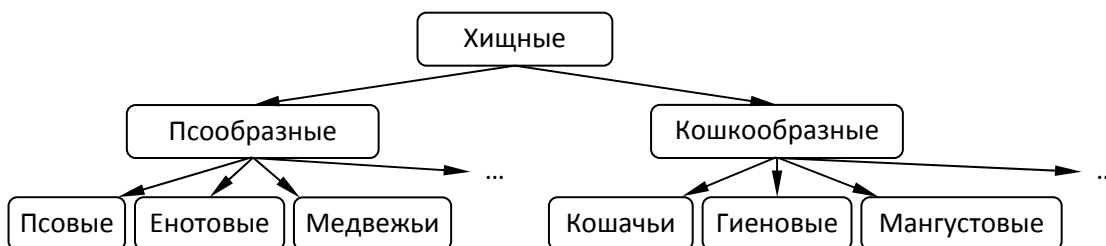
Из двух связанных узлов тот, который находится на более высоком уровне, называется «родителем», а другой – «сыном». Корень – это единственный узел, у которого нет «родителя»; у листьев нет «сыновей».

Используются также понятия «предок» и «потомок». «Потомок» какого-то узла – это узел, в который можно перейти по стрелкам от узла-предка. Соответственно, «предок» какого-то узла – это узел, из которого можно перейти по стрелкам в данный узел.

В дереве на рисунке справа родитель узла E – это узел B, а предки узла E – это узлы A и B, для которых узел E – потомок. Потомками узла A (корня) являются все остальные узлы.



Типичный пример иерархии – различные *классификации* (животных, растений, минералов, химических соединений). Например, отряд *Хищные* делится на два подотряда: *Псообразные* и *Кошкообразные*. В каждом из них выделяют несколько семейств:



Конечно, на этой схеме показаны не все семейства, остальные обозначены многоточием.

В текстах иерархию часто оформляют в виде многоуровневого списка. Например, оглавление книги о хищниках может выглядеть так:

Глава 1. Псообразные

1.1 Псовые

1.2 Енотовые

1.3 Медвежьи

...

Глава 2. Кошкообразные

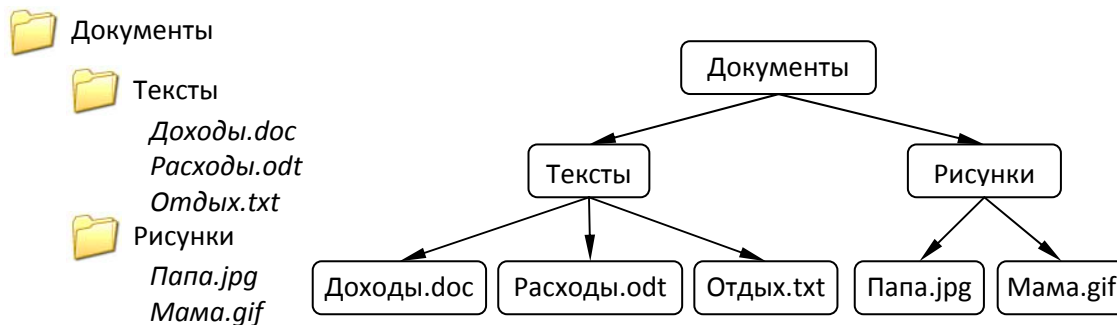
2.1 Кошачьи

2.2 Гиеновые

2.3 Мангустовые

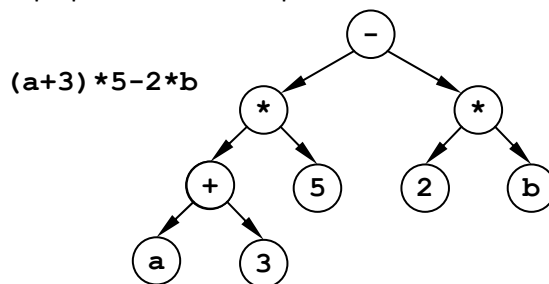
...

С иерархией мы встречаемся, работая с файлами и папками: классическая файловая система имеет древовидную структуру¹⁵. Вход в папку – это переход на следующий (более низкий) уровень иерархии:



¹⁵ В современных файловых системах файл может «принадлежать» нескольким каталогам одновременно. При этом древовидная структура, строго говоря, нарушается.

Алгоритм вычисления арифметического выражения тоже может быть представлен в виде дерева:



Здесь листья — это числа и переменные, тогда как корень и промежуточные узлы — знаки операций. Вычисления идут «снизу вверх», от листьев — к корню. Показанное дерево можно записать так:

$$(- (* (+ (a, 3), 5), * (2, b)))$$

Самое интересное, что скобки здесь не обязательны; если их убрать, то выражение все равно может быть однозначно вычислено:

$$- * + a 3 5 * 2 b$$

Такая запись, которая называется *префиксной* (операция записывается *перед* данными), просматривается с конца. Как только встретится знак операции, эта операция выполняется с двумя значениями, записанными справа. В рассмотренном выражении сначала выполняется умножение:

$$- * + a 3 5 (2*b)$$

затем — сложение:

$$- * (a+3) 5 (2*b)$$

и еще одно умножение

$$- (a+3)*5 (2*b)$$

и, наконец, вычитание: $(a+3)*5 - (2*b)$.

Для того, чтобы вычислять выражение не с конца, а с начала, префиксную форму «разворачивают» задом наперед, тогда получается *постфиксная* форма (операция *после* данных). Например, рассмотренное выше выражение может быть записано в виде

$$b 2 * 5 3 a + * -$$

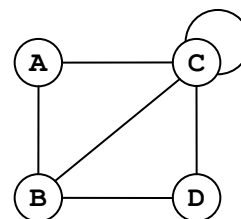
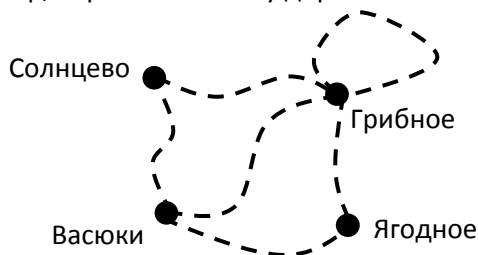
Для вычисления такого выражения скобки также не нужны, и это очень удобно для автоматических расчетов. Когда программа на языке программирования высокого уровня переводится в машинные коды, все выражения записываются в бесскобочной постфиксной форме и именно так и вычисляются.

1.4.4. Графы

Подумайте, как можно структурировать такую информацию:

«От пос. Васюки три дороги идут в Солнцево, Грибное и Ягодное. Между Солнцевым и Грибным и между Грибным и Ягодным также есть дороги. Кроме того, есть дорога, которая идет из Грибного в лес и возвращается обратно в Грибное».

Можно, например, нарисовать схему дорог:



В информатике такие схемы называются *графами*.

Граф – это набор узлов (вершин) и связей между ними (рёбер).

Для хранения информации об узлах и связях показанного выше графа (см. правый рисунок) можно использовать таблицу (матрицу) такого вида,

	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	1	1
D	0	1	1	0

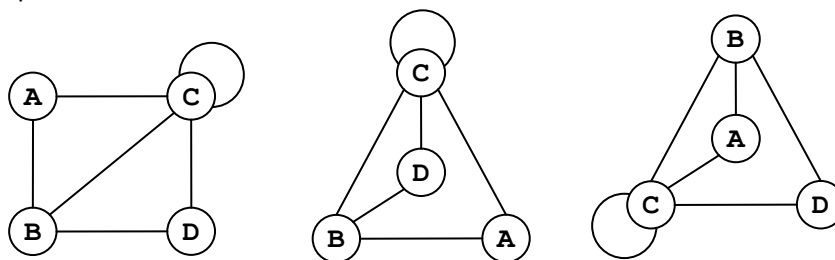
Единица на пересечении строки A и столбца B означает, что между узлами A и B есть связь. Ноль указывает на то, что связи нет. Такая таблица называется **матрицей смежности**. Она симметрична относительно главной диагонали (серые клетки в таблице).

На пересечении строки C и столбца C стоит единица, которая говорит о том, что в графе есть **петля** – ребро, которое начинается и заканчивается в одной и той же вершине.

Можно поступить иначе: для каждого узла перечислить все узлы, с которыми связан данный узел. В этом случае мы получим **список смежности**. Для рассмотренного графа список смежности выглядит так:

(A (B, C), B (A, C, D), C (A, B, C, D), D (B, C))

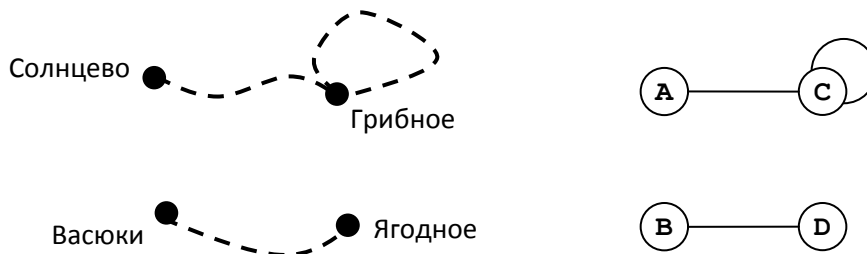
Матрица смежности (и список смежности) не дают никакой информации о том, как именно расположены узлы друг относительно друга. Для таблицы, приведенной выше, возможны, например, такие варианты:



В рассмотренном примере все узлы связаны, то есть, между любой парой узлов существует **путь** – последовательность ребер, по которым можно перейти из одного узла в другой. Такой граф называется **связным**.

Связный граф – это граф, в котором все узлы связаны.

Теперь представьте себе, что дороги Васюки – Солнцево, Васюки – Грибное и Грибное – Ягодное завалило снегом (ли размыло дождем) так, что по ним не пройти и не проехать.



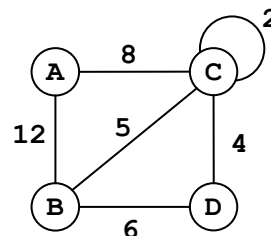
Эту схему тоже можно считать графом (она подходит под определение), но в таком графе есть две несвязанные части, каждая из которых – связный граф. Такие части называют **компонентами связности**.

Вспоминая материал предыдущего пункта, можно сделать вывод, что дерево – это частный случай связного графа. Но у него есть одно важное свойство – в дереве нет замкнутых путей (**цик-**

лов). Граф в последнем примере не является деревом, потому что в нем есть циклы: ABCA, BCDB, ABCDA.

Дерево – это связный граф, в котором нет циклов.

Если в первом примере с дорогами нас интересуют еще и расстояния между поселками, каждой связи нужно сопоставить число (вес).



Такой граф называется **взвешенным**, поскольку каждое ребро имеет свой **вес**. В реальных задачах это может быть не только расстояние, но и, например, стоимость проезда или другая величина.

Как хранить информацию о таком графе? Ответ напрашивается сам собой – нужно в таблицу записывать не 1 или 0, а вес ребра. Если связи между двумя узлами нет, на бумаге можно оставить ячейку таблицы пустой, а при хранении в памяти компьютера записывать в нее условный код, например, –1. Такая таблица называется **весовой матрицей**, потому что содержит веса ребер. В данном случае она выглядит так:

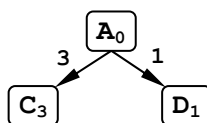
	A	B	C	D
A		12	8	
B	12		5	6
C	8	5		4
D		6	4	

Также как и матрица смежности, весовая матрица симметрична относительно диагонали. Две пустые ячейки говорят о том, что между узлами A и D нет связи.

Если в графе немного узлов, весовая матрица позволяет легко определить наилучший маршрут из одного узла в другой простым перебором вариантов. Рассмотрим граф, заданный весовой матрицей (числа определяют стоимость поездки между соседними пунктами):

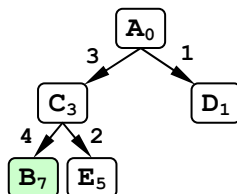
	A	B	C	D	E
A			3	1	
B			4	5	1
C	3	4			2
D	1	5			1
E		1	2	1	

Найдем наилучший путь из A в B – такой, при котором общая стоимость поездки минимальная. Сначала видим, что из пункта A напрямую в B ехать нельзя, а можно ехать только в C и D. Изобразим это на схеме:

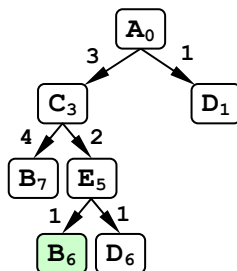


Числа около каждого ребра показывают стоимость поездки по этому участку, а индексы у названий узлов показывают общую стоимость проезда в данный узел из узла A.

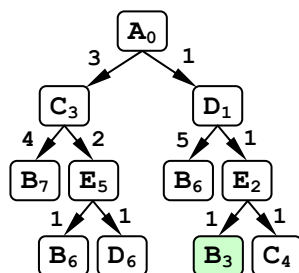
Теперь разберем варианты дальнейшего движения из узла С (узел А уже не нужно рассматривать, так как мы из него пришли).



Видим, что из С сразу можно попасть в В, стоимость проезда в этом случае равна 7. Но, возможно, это не самый лучший вариант, и нужно проверить еще путь через узел Е. Действительно, оказывается, что можно сократить стоимость до 6:



Исследовать дальше маршрут, содержащий цепочку ACED, нет смысла, потому что его стоимость явно будет больше 6. Аналогично строим вторую часть схемы (вы догадались, что она представляет собой дерево!).

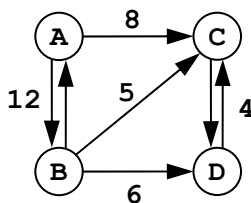
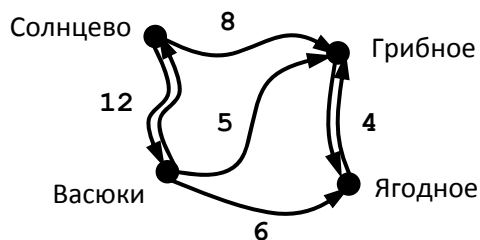


Таким образом, **оптимальный (наилучший) маршрут – ADEB**, его стоимость – 3. Маршруты ACED и ADEC, не дошедшие до узла В, далее проверять не нужно, они не улучшат результат.

Конечно, для более сложных графов метод перебора работает очень долго, поэтому используются более совершенные (но более сложные) методы, о которых мы поговорим в 11-м классе.

Наверное, вы заметили, что при изображении деревьев, которые описывают иерархию (подчинение), мы ставили стрелки от верхних уровней к нижним. Это означает, что для каждого ребра указывается направление, и двигаться можно только по стрелкам, но не наоборот. Такой граф называется *ориентированным* (или коротко *орграфом*). Он может служить, например, моделью системы дорог с односторонним движением. Матрица смежности и весовая матрица для орграфа уже не обязательно будут симметричными.

На этой схеме всего две дороги с двусторонним движением, по остальным можно ехать только в одну сторону:



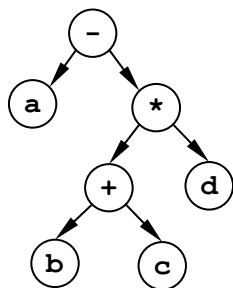
	A	B	C	D
A			8	
B	12		5	6
C				4
D			4	

Ребра в орграфе называют *дугами*. Дуга, в отличие от ребра, имеет начало и конец.

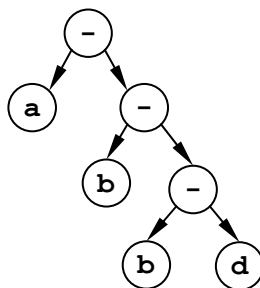
? Вопросы и задачи

1. Что такое структура?
2. Что такое структурирование информации? Зачем оно нужно?
3. Что такое алфавитный порядок? Как поступают, если начальные символы слов совпали?
4. Расскажите, как используются оглавление, словарь и индекс для быстрого поиска нужной информации. Чем эти средства отличаются друг от друга?
5. Какими способами можно задать множество? Что такое пустое множество?
6. Приведите примеры множеств.
7. Чем отличаются множество и линейный список?
8. Что такое очередь? Приведите примеры.
9. Что такое стек? Чем он отличается от очереди? Приведите примеры.
10. Расшифруйте английские сокращения LIFO и FIFO.
11. Какая структура позволяет объединить возможности стека и очереди?
12. Что такое матрица?
13. Как можно записать табличные данные в виде списка?
14. Что такое иерархия? Приведите примеры.
15. Вспомните известные вам классификации, которые вы изучали на других предметах.
16. Как называется соответствующая структура в информатике?
17. Что такое «корень», «лист», «родитель», «сын», «предок», «потомок»?
18. В дереве 4 потомка и все они являются листьями. Нарисуйте это дерево. Сколько в нем узлов?
19. Что такое «лес»?
20. Определите выражения, соответствующие каждому из деревьев, в «нормальном» виде со скобками (эту форму называют *инфиксной* – операция записывается *между* данными). Постройте для каждого из них постфиксную форму.

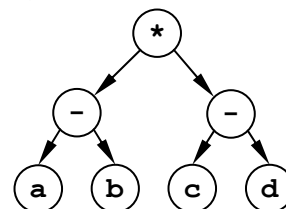
а)



б)



в)



21. Постройте деревья, соответствующие следующим арифметическим выражениям. Запишите эти выражения в префиксной и постфиксной формах.

а) $(a+b) * (c+2*d)$

в) $(a+b+2*c) * d$

б) $(2*a-3*d) * c+2*b$

г) $3*a-(2*b+c) * d$

22. Вычислите выражение, записанное в постфиксной форме

а) $12\ 6\ +\ 7\ 3\ -\ 1\ -\ * \ 12\ +$

б) $12\ 10\ -\ 5\ 7\ +\ * \ 7\ -\ 2\ *$

в) $5\ 6\ 7\ 8\ 9\ +\ -\ +\ -$

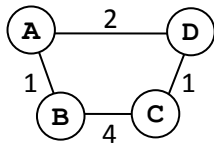
г) $5\ 4\ 3\ 2\ 1\ -\ -\ -\ -$

Запишите его в инфиксной и в префиксной формах. Постройте дерево, соответствующее этому выражению. Единственно ли такое дерево? В этом дереве назовите корень, листья и промежуточные узлы.

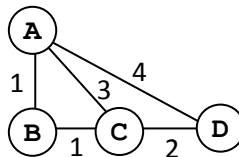
(Ответ: 66; 34; 9; 3)

23. В чем отличие понятий «дерево», «лес», «граф»?
24. Какой граф называется связным?
25. Что такое компонента связности?
26. Нарисуйте граф, в котором 5 вершин и три компонента связности. Постройте его матрицу смежности.
27. Структурируйте эту информацию разными способами: «Между поселками Верхние Васюки и Нижние Васюки есть проселочная дорога длиной 10 км. Село Сергеево соединяется двумя асфальтовыми шоссе с Нижними Васюками (22 км) и Верхними Васюками (16 км). В Солнечное можно доехать только из Сергеева по грунтовой дороге (5 км)». Можно ли сказать, как точно расположены эти пункты?
28. Для графа, полученного в предыдущей задаче, постройте матрицу смежности, список смежности, весовую матрицу. Является ли этот граф деревом?
29. Что такое петля? Как по матрице смежности определит, есть ли петли в графе?
30. Постройте матрицы смежности и весовые матрицы для каждого графа:

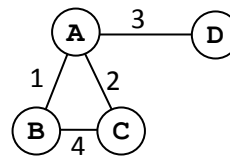
а)



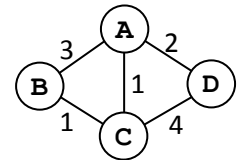
б)



в)



г)



31. Постройте графы, соответствующие каждой из матриц смежности:

а)

	A	B	C	D	E
A		0	1	1	0
B	0		1	0	1
C	1	1		0	1
D	1	0	0		0
E	0	1	1	0	

б)

	A	B	C	D	E
A		0	1	1	1
B	0		1	0	0
C	1	1		0	1
D	1	0	0		0
E	1	0	1	0	

в)

	A	B	C	D	E
A		0	1	1	1
B	0		1	0	1
C	1	1		0	1
D	1	0	0		0
E	1	1	1	0	

г)

	A	B	C	D	E
A		0	0	1	0
B	0		1	0	1
C	0	1		1	1
D	1	0	1		0
E	0	1	1	0	

32. Постройте графы, соответствующие каждой из весовых матриц:

а)

	A	B	C	D	E
A		4	3		7
B	4			2	
C	3			6	
D		2	6		1
E	7			1	

б)

	A	B	C	D	E
A		2	5		6
B	2			3	
C	5				
D		3			1
E	6			1	

в)

	A	B	C	D	E
A			2	2	6
B				2	
C	2			2	
D	2	2	2		
E	6				

г)

	A	B	C	D	E
A		5	2		6
B	5			5	
C	2			2	
D		5	2		3
E	6			3	

33. Стоимость перевозок между пунктами, которые для краткости обозначены буквами А, В, С, D и Е, задается таблицей (весовой матрицей графа). Нужно перевезти груз из пункта А в пункт В. Для каждого из четырех вариантов определите оптимальный маршрут и полную стоимость перевозки.

а)

	A	B	C	D	E
A			3	1	
B			4		2
C	3	4			2
D	1				
E		2	2		

б)

	A	B	C	D	E
A			3	1	1
B			4		
C	3	4			2
D	1				
E	1		2		

в)

	A	B	C	D	E
A			3	1	4
B			4		2
C	3	4			2
D	1				
E	4	2	2		

г)

	A	B	C	D	E
A				1	
B			4		1
C		4		4	2
D	1		4		
E		1	2		

34. Что такое орграф? Когда для представления данных используются орграфы? Приведите примеры.

35. Постройте орграф, соответствующий каждой из этих таблиц.

а)

	A	B	C	D	E
A			3	1	
B	2		4		2
C	3				
D	1				
E			2		

б)

	A	B	C	D	E
A			5	1	1
B			6	4	
C	3	4			2
D		2			
E			3		

в)

	A	B	C	D	E
A			3	1	4
B			4		2
C		4			2
D					
E	4		2		

г)

	A	B	C	D	E
A				1	
B			4		1
C	3	4		4	2
D	1	2	4		
E	1	1	2		

36. В чем разница между понятиями «ребро» и «дуга»?

37. Выберите наиболее подходящий способ структурирования информации для хранения

- а) данных по крупнейшим озерам мира;
- б) рецепта приготовления шашлыка;
- в) схемы железных дорог;
- г) схемы размещения файлов на флэш-диске.