# Лекция 1. Информатика

## Подходы к определению информации

Хотя понятие «информация» является одним из фундаментальных в современной науке (наряду с веществом, энергией и временем), единое определение информации на данный момент отсутствует. Существуют различные ***подходы*** к определению.

Варианты определения информации:

* в **быту**: сведения, которые представляют некоторую ценность (интересны);
* в **философии**: отражение реального мира, сведения об одном объекте, имеющиеся у другого или того же самого объекта;
* в **технике**: последовательности сигналов, которые хранятся, передаются или обрабатываются с помощью технических средств (без учета смысла этих сигналов);
* в **журналистике**: любые сведения, обладающие новизной;
* в **теории информации**: сведения, которые полностью снимают или уменьшаю существующую неопределенность;
* в **кибернетике**: знания, которые используются для управления в различных системах;
* в **документалистике**: любой текст, зафиксированный в знаковой форме в виде документа;
* и др.

Само понятие появилось в 30-х гг XXв. в журналистике. Происходит от лат. informatio – «разъяснение», «осведомление», «изложение».

Две крайних точки зрения:

1. Информация объективна. Она присуща всем объектам окружающего мира и cуществует независимо от человека. Человек только раскрывает смысл информации.
2. Информация субъективна. Она не существует в материальном мире, а возникает в процессе восприятия у человека или другого существа. Информация – одна из функций жизни, отличие живого от неживого. Человек создает информацию.

В информации принято выделять три **аспекта**:

* синтаксический (форма) – связан со способом представления информации, независимо от ее смысловых и потребительских качеств;
* семантический (смысл) – формирование понятий и представлений, выявления содержания и обобщения знаний;
* прагматический (полезность) – определяет возможность достижения цели с помощью информации.

## Свойства информации

Качественные признаки, оценка информации. Иногда называют **требованиями** к информации.

1. *Объективность* – субъективность: зависит ли от чьего-либо мнения.
2. *Достоверность* – недостоверность: отражает ли реальное положение дел.

Причины недостоверности:

* преднамеренное искажение (дезинформация);
* непреднамеренное искажение (опечатки, помехи при передаче);
* преуменьшение или преувеличение реальных фактов (слухи, реклама).

1. *Полнота* – неполнота (недостаточность), а также избыточность: хватает ли для достижения цели (принятия решения).
2. *Актуальность* (своевременность) – неактуальность: важность в текущий момент времени. Варианты неактуальности: устаревшая, преждевременная, в принципе незначимая.
3. *Полезность* – бесполезность: ценность для решения поставленной задачи.
4. *Понятность* – непонятность: доступность формы представления для получателя.
5. Прочее: *адекватность*, логичность, удобство, компактность, воспроизводимость, стираемость и др.

## Информационные процессы

Информация не существует сама по себе, она проявляется в информационных процессах.

**Информационный процесс** – совокупность последовательных действий, производимых над информацией для получения какого-либо результата.

Основные:

1. **Передача** (и получение) информации.
2. **Хранение** – способ распространения информации в пространстве и времени.
3. **Обработка** – преобразование информации в семантическом и/или синтаксическом смыслах.

Другие (частные случаи):

* поиск,
* отбор,
* накопление,
* создание,
* использование (принятие решения),
* структурирование,
* кодирование,
* защита
* и др.

**Атрибутами** информационных процессов выступают:

* источник информации;
* получатель информации;
* канал передачи данных;
* носитель информации.

**Носитель** – это среда для записи, хранения и передачи информации. Носителем может быть любой предмет, электромагнитные волны и поля, вещество в различных состояниях и др. Формой передачи информации через носитель являются **знак** (буква, цифра) и **сигнал** (электрический, световой).

Виды сигналов:

* **аналоговый** – непрерывно изменяющийся по амплитуде и во времени;
* **дискретный** (цифровой) – представленный в виде последовательного набора отдельных значений. Сам набор может быть как конечным, так и бесконечным.

*Дискретизация* – преобразование аналогового сигнала в цифровой.

## Информационные технологии (ИТ)

**ИТ** – приемы, способы и методы применения средств вычислительной техники при выполнении функций сбора, хранения, обработки, передачи и использования данных.

**Цель:** производство информации, удовлетворяющей человеческие потребности. В результате получается информация нового качества (**информационный продукт**).

**Особенность технологии:** применение одинаковой технологии к одинаковому сырью всегда дает одинаковый результат.

**Принципы** компьютерных ИТ:

* интерактивность (ориентирование на пользователя);
* интегрированность (с другими ИТ);
* гибкость (как по отношению к данным, так и к постановке задачи).

Общие **требования** к технологии:

* разбиение технологии на отдельные этапы, действия, операции;
* должна включать весь набор средств для достижения цели;
* иметь регулярный характер (повышение стандартизации и унификации).

Примеры ИТ:

* электронный документооборот;
* беспроводная связь;
* 3D-принтер;
* машинный перевод.

## Информационное общество

Общество:

1. *Аграрное* – основная масса трудится в сельскохозяйственном секторе. Низкий уровень стандартизации – ручной труд, индивидуальное производство.
2. *Индустриальное* – основная масса трудится в промышленном секторе. Повышается уровень стандартизации – массовое, конвейерное производство.
3. *Постиндустриальное* (***информационное***) – основная масса трудится в сфере услуг. Очень высокий уровень стандартизации – автоматизированное производство.

**Информационное общество** – общество, в котором большинство работающих занято производством, хранением, переработкой и реализацией информации.

Особенности:

* приоритет информации над другими ресурсами;
* возрастание числа людей, занятых в сфере услуг, а не материального производства;
* высокая степенно проникновения ИТ во все сферы деятельности (информатизация)
* создание глобального информационного пространства для доступа к мировым информационным ресурсам
* развитие информационной экономики, электронного правительства, цифровых рынков, электронных сетей, информационного права, информационной политики.

**Информационные революции:**

1. Изобретение письменности (тысячи лет назад)
2. Изобретение книгопечатания (средние века)
3. Открытие электричества, появление телеграфа и телефона (XIX – начало XXв.)
4. Изобретение компьютеров и Интернета (1940-1980гг)

## Информационная экономика

Термин «информационная экономика» употребляется в нескольких значениях.

1. Экономический уклад информационного общества, характеризуется преобладающей ролью интеллектуального, творческого труда и информационных продуктов.
2. Отрасль экономики, относящаяся к работе с информацией, а также компьютерная индустрия.
3. Экономическая теория информационного общества.

Основные понятия: *информационные* *ресурсы*, *продукты* и *услуги*. К ним относятся документы, патенты, программы, базы данных, информационные системы, консультации.

Для информационной экономики характерна высокая степень стандартизации производства, формирование глобальных рынков.

Возникает **информационный рынок**. Его компоненты:

* рынок техники;
* рынок ИТ;
* рынок информационных продуктов;
* рынок услуг.

Субъекты информационного рынка:

* производители;
* владельцы;
* пользователи.

**Информационный работник** – работник умственного труда, чья деятельность связана с обработкой имеющейся информации и получением новой информации (программисты, аналитики, специалисты по планированию и др.). Иногда в эту группу включают всех работников, обладающих высоким уровнем образования или связанных с образованием (в том числе ученых, преподавателей и студентов).

## Измерение информации. Представление информации в компьютере

Минимальной единицей информации является **бит**. Бит может быть равен 0 или 1, но можно обозначить эти два варианта и по-другому: истина/ложь, да/нет, вкл./выкл., true/false. Физически это выражается в двух различных состояниях устройств: есть ток или нет тока, намагничено – не намагничено, отражает свет – не отражает свет.

Это означает, что в компьютере используется *двоичная система счисления*. Для человека наиболее привычна десятичная система. Любое десятичное число можно перевести в двоичную систему:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 010 | 02 |  | 910 | 10012 |
| 110 | 12 |  | 1010 | 10102 |
| 210 | 102 |  | 1110 | 10112 |
| 310 | 112 |  | 1210 | 11002 |
| 410 | 1002 |  | 1310 | 11012 |
| 510 | 1012 |  | 1410 | 11102 |
| 610 | 1102 |  | 1510 | 11112 |
| 710 | 1112 |  | 1610 | 100002 |
| 810 | 10002 |  | … |  |

Можно заметить, что числа, равные степеням двойки (4=22, 8=23, 16=24) записываются как 1 и столько же нулей, что и степень 2. Например,

25610 = 28 = 1 0000 00002

А числа перед ними состоят из одних единиц:

310 = 22 – 1 = 112

710 = 23 – 1 = 1112

1510 = 24 – 1 = 11112

Исторически сложилось, что биты объединяются в группы по 8. Группа из 8 бит называется **байт**. Байт является минимальной адресуемой ячейкой памяти компьютера. Т.е. вы не можете записать в память один бит, только целый байт.

В один байт можно записать числа от 0 до 255, самое большое:

1111 11112 = 28 – 1 = 25510

Байт – это очень маленькая единица информации. Даже для записи числа жителей России (142 млн. чел.) потребуется три байта.

Для измерения информации используются более крупные единицы – килобайты, мегабайты, гигабайты и др.

Однако с этими приставками существует небольшая путаница. Дело в том, что

210 = 1024 ≈ 1000 =103.

В физике «кило-» означает 1000, но в информатике под килобайтом или килобитом подразумевается 1024 байта или бита соответственно. Из-за этого, например, флешка, на которой написано «4ГБ» в компьютере будет отображаться как 3,8ГБ. В России во избежание в соответствующем ГОСТе указаны различные сокращения: ГБ = 1000 байт, а Гбайт = 1024 байт.

**Двоичные приставки**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Приставка | По степеням 10 (СИ/ГОСТ) | | По степеням 2 (ГОСТ) | |
| кило- | 103Б | кБ | 210Б = 1024 Б | Кбайт |
| мега- | 106Б | МБ | 220Б ≈ 1,05∙106Б | Мбайт |
| гига- | 109Б | ГБ | 230Б ≈ 1,07∙109Б | Гбайт |
| тера- | 1012Б | ТБ | 240Б ≈ 1,10∙1012Б | Тбайт |
| пета- | 1015Б | ПТ | 250Б ≈ 1,13∙1015Б | Пбайт |
| экса- | 1018Б | ЭБ | 260Б ≈ 1,15∙1018Б | Эбайт |
| зетта- | 1021Б | ЗБ | 270Б ≈ 1,18∙1021Б | Збайт |
| йотта- | 1024Б | ЙБ | 280Б ≈ 1,21∙1024Б | Йбайт |

На сегодня йоттабайт – самая крупная единица измерения информации. По оценкам экспертов, в 2011г. в мире накоплено в общей сложности 1,8 ЗБ информации, т.е. около 1 800 000 000 000 000 000 000 байт.

В общем случае могут использоваться любые удобные единицы измерения – страницы, книги, DVD-диски, экраны.

Но компьютер не умеет работать ни с чем, кроме двоичных чисел. Поэтому вся остальная информация – текст, даты, рисунки, звук, видео – *кодируется* в числовом виде.

**Кодирование дат**

Любая дата вместе со временем представляется в виде числа дней, прошедших с 1 января 1900г. Поэтому даты можно вычитать одну из другой, получится число дней между ними. Но если ошибиться с форматом представления, то вместо даты отобразится какое-то большое число. Например, 1 января 2000г. = 36526.

**Кодирование текста**

Каждый символ текста представляется в виде определенного числа (кода). Это соответствие записано в виде таблицы.

Первой такой таблицей была **ASCII** (читается «аски»). В ней каждому символу ставится в соответствие 1 байт, т.е. всего она содержит 256 символов (от 0 до 255).

Вот ее фрагмент:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Код** | **32** | **33** | **34** | **35** | **36** | **37** | **38** | **39** | **40** | **41** | **42** | **43** | **44** | **45** | **46** | **47** |
| **Символ** | пробел | ! | " | # | $ | % | & | ' | ( | ) | \* | + | , | - | . | / |
| **Код** | **48** | **49** | **50** | **51** | **52** | **53** | **54** | **55** | **56** | **57** | **58** | **59** | **60** | **61** | **62** | **63** |
| **Символ** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| **Код** | **64** | **65** | **66** | **67** | **68** | **69** | **70** | **71** | **72** | **73** | **74** | **75** | **76** | **77** | **78** | **79** |
| **Символ** | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| **Код** | **80** | **81** | **82** | **83** | **84** | **85** | **86** | **87** | **88** | **89** | **90** | **91** | **92** | **93** | **94** | **95** |
| **Символ** | P | Q | R | S | T | U | V | w | X | Y | Z | [ | \ | ] | ^ | \_ |

Любой ASCII-символ можно ввести с клавиатуры через код. Нужно зажать клавишу Alt и, не опуская ее, на дополнительной цифровой клавиатуре (справа) набрать код (четырьмя цифрами). Например, Alt+0167 = §, Alt+0169 = ©. Так можно вводить символы, которых нет на клавиатуре.

Однако в мире очень много языков с различными символами, все они в эту таблицу не уместятся. Проблему решили следующим образом.

Первая половина таблицы (0-127) стала универсальной, международной. В ней содержатся английские буквы, цифры, знаки препинания, арифметические действия, пробел, а также *непечатаемые символы* – «удаление», «переход на новую строку», «выход» и др.

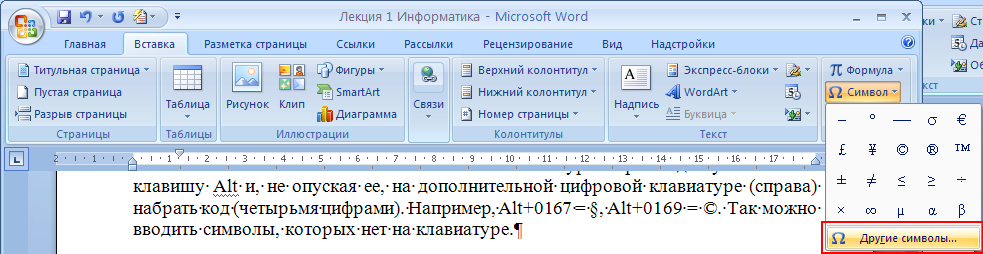
Вторая половина таблицы – национальная. Каждый сам решает, что и как в нее записывать. Например, существуют варианты кодировки кириллицы – Windows-1251 (американская), КОИ8 (российская, сейчас выходит из употребления). Для кодирования расширенной латиницы (с точками, ударениями над буквами) используется Windows-1252.

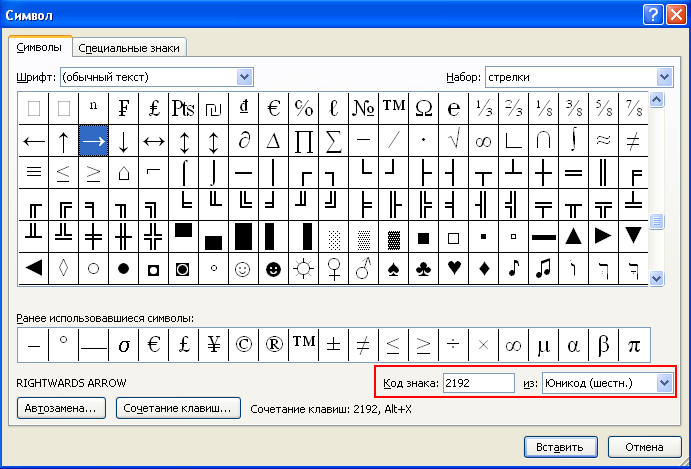
К сожалению, это приводит к тому, что в тексте нельзя использовать одновременно несколько разных языков, кроме английского. Из-за неправильной кодировки вместо текста на экране будут отображаться «зюки» и «кракозябры».

Поэтому позже была создана таблица **Unicode** (читается «юникод»).

В ней каждый символ кодируется двумя байтами, всего в нее можно записать 216 = 65 536 символов. Этого оказалось достаточно для всех современных языков Земли (по крайней мере, использующихся в компьютерах). В таблице даже осталось немного места «про запас», на будущее.

Все символы Unicode можно найти в MS Word.





В Word символ можно набрать, сначала введя его код, а потом нажав Alt+X. Так можно поставить ударение: 0301, Alt+X = ́, 0300, Alt+X = ̀. Ударение смещается на предыдущую букву (например, замо́к).

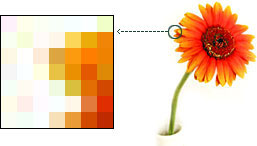
**Кодирование графики**

Существует два основных способа представления рисунков в компьютере:

* растровый
* векторный.



В *растровой* графике каждый рисунок представляется в виде точек (пикселей). Каждая точка имеет свой цвет. Так лучше всего хранить фотографии, художественные картины и другие реалистичные изображения с плавными переходами и большим количеством цветов.



В *векторной* графике используются линии и кривые, которые компьютер обрабатывает как математические формулы.

Например, формула окружности:

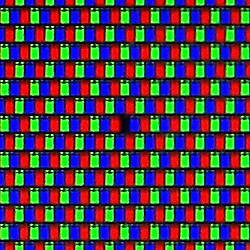
(x – x0)2 + (y – y0)2= R2

(x0, y0) – координаты центра, R – радиус.

Поэтому, когда в векторной графике рисуется круг, компьютер просто запоминает эти числа и цвет линии и заливки.

Получается, что в любом случае нужно закодировать **цвет**.

Основной способ кодировки цвета – **RGB** (red, green, blue – красный, синий зеленый). Все остальные цвета смешиваются из них. Так представляются точки на экране монитора. Под лупой он выглядит так:



Каждый цвет записывается одним байтом, т.е. всего каждый пиксель занимает три байта. Каждый цвет кодируется числом от 0 до 255.

Примеры:

|  |  |  |
| --- | --- | --- |
| RGB(0,0,0) | черный | ни один цвет не горит на экране |
| RGB(255,255,255) | белый | светятся все цвета |
| RGB(127,127,127) | серый | все цвета поровну, но неярко |
| RGB(255,0,0) | красный | только красный |
| RGB(255,255,0) | желтый | красный и зеленый |
| RGB(78,169,214) |  | сложный оттенок синего |

Всего так можно закодировать очень много цветов – 224 = 16,5млн. Это примерно соответствует числу оттенков, которые способен различить человеческий глаз.

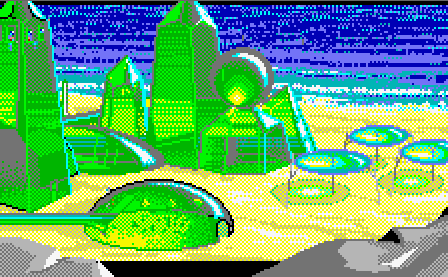
Поскольку в картинке пикселей очень много (тысячи и миллионы), рисунки занимают много места, и для них используется *сжатие*. Без сжатия одна фотография занимала бы столько же места, как книга из 1000 страниц текста.

Замечание. Раньше, когда компьютеры еще не могли работать с такими большими объемами, обходились гораздо меньшим количеством цветов: 16, потом 256. Каждый цвет, как и символ, кодировался определенным числом, записанным в таблице – **палитре**.

Системная палитра:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Код* | *R* | *G* | *B* | *Название* | *Цвет* |
| 0 | 0 | 0 | 0 | Черный |  |
| 1 | 255 | 255 | 255 | Белый |  |
| 2 | 255 | 0 | 0 | Красный |  |
| 3 | 0 | 255 | 0 | Зеленый |  |
| 4 | 0 | 0 | 255 | Синий |  |
| 5 | 255 | 255 | 0 | Желтый |  |
| 6 | 0 | 255 | 255 | Голубой |  |
| 7 | 255 | 0 | 255 | Розовый (фуксин) |  |
| 8 | 128 | 0 | 0 | Бордовый |  |
| 9 | 0 | 128 | 0 | Темно-зеленый |  |
| 10 | 0 | 0 | 128 | Темно-синий |  |
| 11 | 128 | 128 | 0 | Оливковый |  |
| 12 | 0 | 128 | 128 | Морской волны |  |
| 13 | 128 | 0 | 128 | Фиолетовый |  |
| 14 | 128 | 128 | 128 | Серый |  |
| 15 | 255 | 128 | 128 | Серебристый |  |

Пример рисунка из 16 цветов:



Можно создать и свою собственную палитру, с цветами для конкретного рисунка. При этом даже с 16 цветами он будет выглядеть не очень страшно.

|  |  |
| --- | --- |
| http://upload.wikimedia.org/wikipedia/commons/6/69/Truecolor.png | 16,7 млн. цветов |
| http://upload.wikimedia.org/wikipedia/commons/f/ff/8_bit.png | 256 цветов |
| http://upload.wikimedia.org/wikipedia/commons/0/0d/4_bit.png | 16 цветов |

**Аудио** и **видео** кодируется множеством разных способов. Для них существуют **кодеки**. В кодеке как раз записывается, как именно нужно кодировать или раскодировать определенный формат аудио и видео.