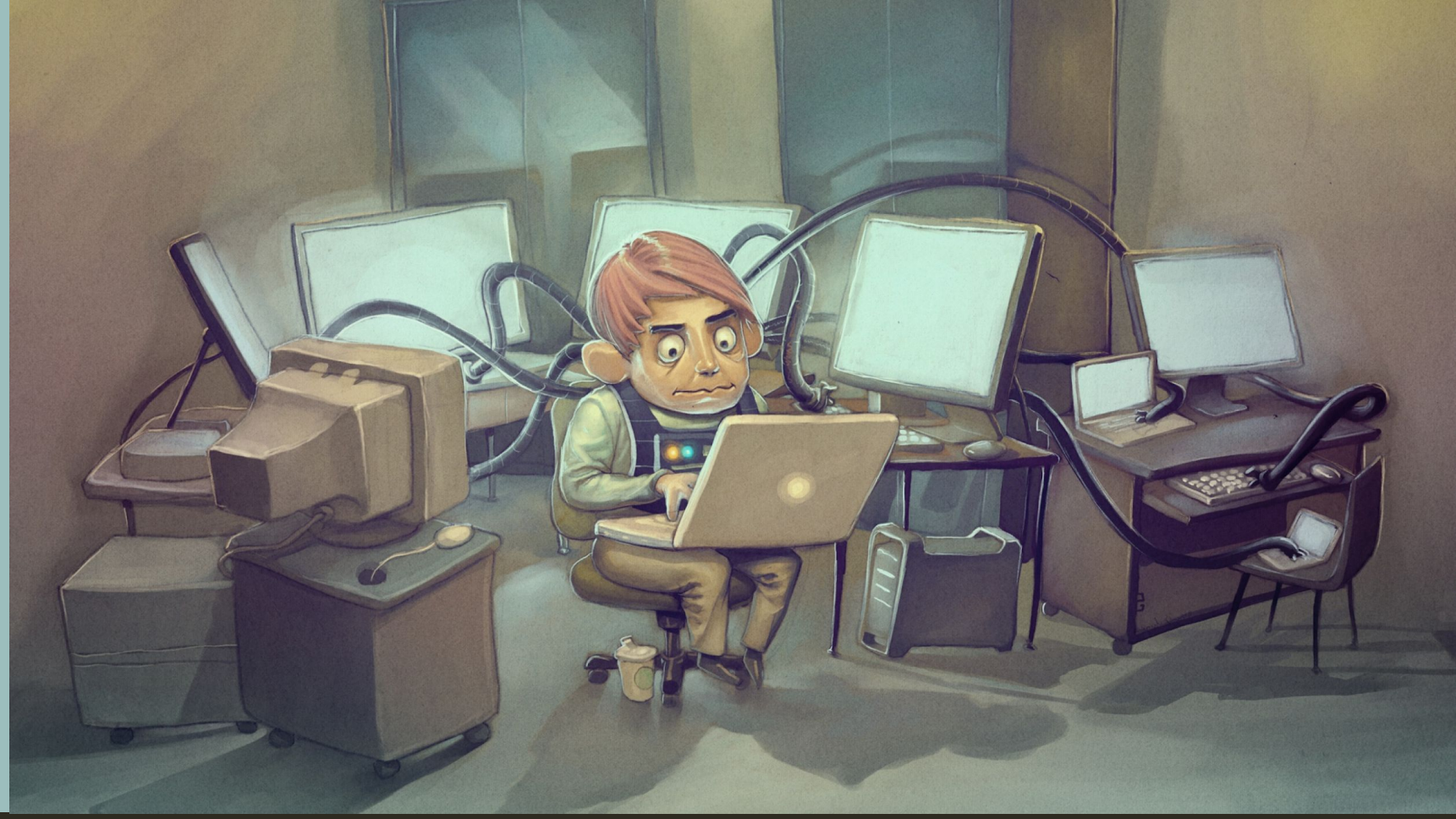


КОДИРОВАНИЕ ТЕКСТОВОЙ ИНФОРМАЦИИ

информатика

Текстовую информацию кодируют двоичным кодом через обозначение каждого символа алфавита определенным целым числом. С помощью восьми двоичных разрядов возможно закодировать 256 различных символов. Данного количества символов достаточно для выражения всех символов английского и русского алфавитов.
В первые годы развития компьютерной техники трудности кодирования текстовой информации были вызваны отсутствием необходимых стандартов кодирования. В настоящее время, напротив, существующие трудности связаны с множеством одновременно действующих и зачастую противоречивых стандартов.
Для английского языка, который является неофициальным международным средством общения, эти трудности были решены. Институт стандартизации США выработал и ввел в обращение систему кодирования ASCII (American Standard Code for Information Interchange – стандартный код информационного обмена США).

Для кодировки русского алфавита были разработаны несколько вариантов кодировок:

1) Windows-1251 – введена компанией Microsoft; с учетом широкого распространения операционных систем (ОС) и других программных продуктов этой компании в Российской Федерации она нашла широкое распространение;

2) КОИ-8 (Код Обмена Информацией, восьмизначный) – другая популярная кодировка российского алфавита, распространенная в компьютерных сетях на территории Российской Федерации и в российском секторе Интернет;

3) ISO (International Standard Organization – Международный институт стандартизации) – международный стандарт кодирования символов русского языка. На практике эта кодировка используется редко.

Код — правило (алгоритм) сопоставления каждому конкретному сообщению строго определённой комбинации символов (знаков) (или сигналов) Представляет собой систему условных знаков для представления информации.

Кодирование – перевод информации в удобную для передачи, обработки, хранения формы с помощью некоторого кода.

Декодирование – процесс восстановления содержания закодированной информации.

Алфавит – множество символов, с помощь которых записывается текст.

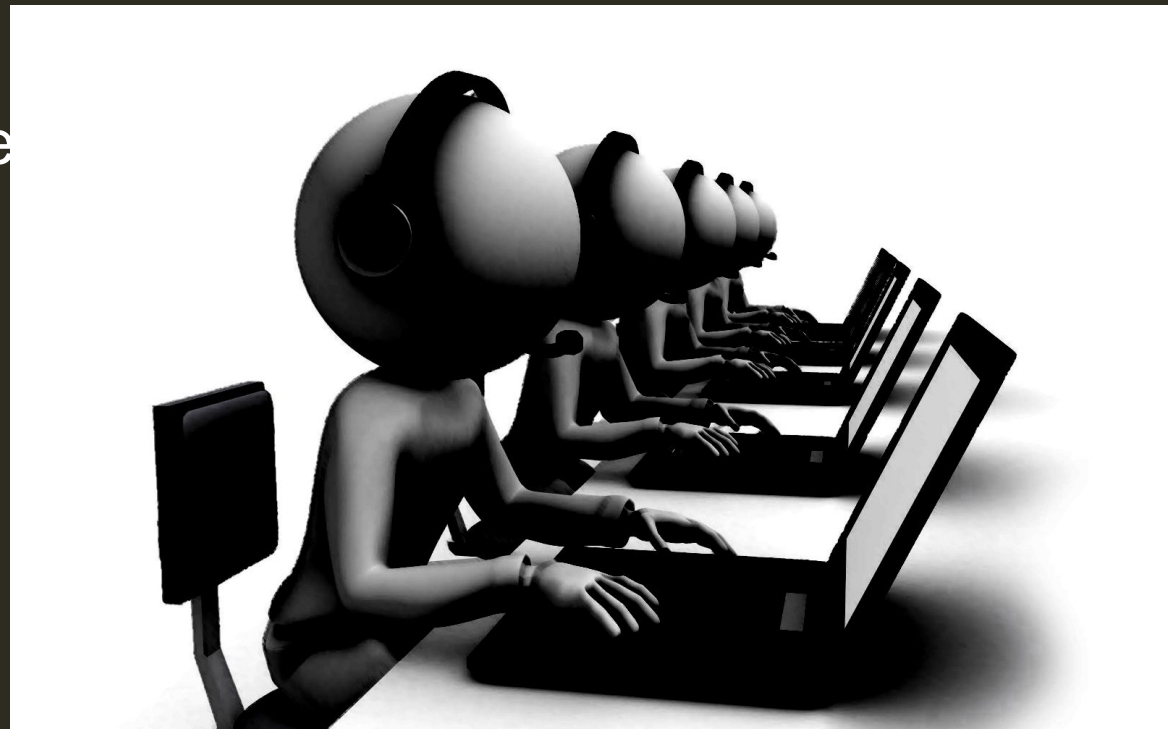
Мощность – число символов алфавите.

Проблемы кодирования:

1) Отсутствие информации о кодировке

2) Программа не знает кодировки

3) Избыток информации о кодировках

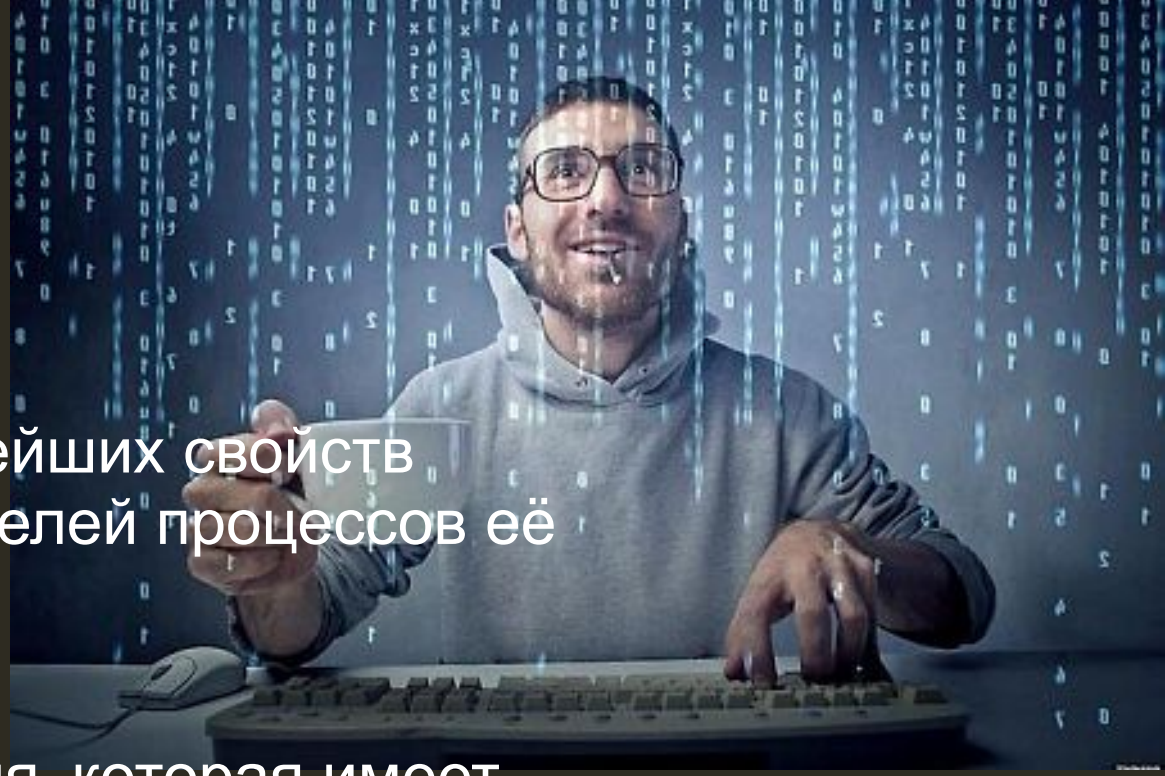


Ключевые свойства текстовых материалов:

- ценность
- новизна
- полезность
- адекватность
- истинность

Текстовая информация - последовательность символов, печатных знаков, которые принадлежат тому или иному набору символов. Может храниться в формативом или не нормативном виде.

ПОДРОБНЕЕ



Ценность информации – одно из важнейших свойств информации, оценка которого зависит от целей процессов её генерации и обработки.

Полезность информации – информация, которая имеет значение.

Адекватность информации – уровень соответствия образа, создаваемого с помощью информации, реальному объекту.

КЛАССИФИКАЦИЯ ИНФОРМАЦИИ:

- синтаксическая – отображает формально-структурные характеристики информации;

- прагматическая – отображает соответствие информации цели управления, реализуемой на её основе;

- семантическая – определяет степень соответствия образа объекта самому объекту.

На сегодняшний день большое колличество пользователей при помощи компьютера обрабатывает текстовую информацию, которая состоит из: букв, цифр, знаков препинания и других элементов.

Обычно для кодирования одного символа, используеться 1 байт памяти то есть 8 бит. По теории вероятностей с помощью простой формулы, которая связывает количество возможных событий (К) и количество информации (I), можно вычислить сколько не одинаковых символов можно закодировать: К = 2^I = 28 = 256.

Принцип данного кодирования заключается в том, что каждому символу (букве, знаку) соответствуе свой двоичный код от 00000000 до 11111111, так-же текстовая информация может быть представлена в десятичном коде от 0 до 255.

Нужно запомнить, что на сегодняшний день для кодирования букв российского алфавита используют пять разных кодировачных таблиц (КОИ - 8, CP1251, CP866, Mac, ISO), запомните, что тексты закодированные с помощью одной таблицы не будут корректно отображаться в другой кодировке. Это можно увидить в обьединенной таблице кодировки символов.