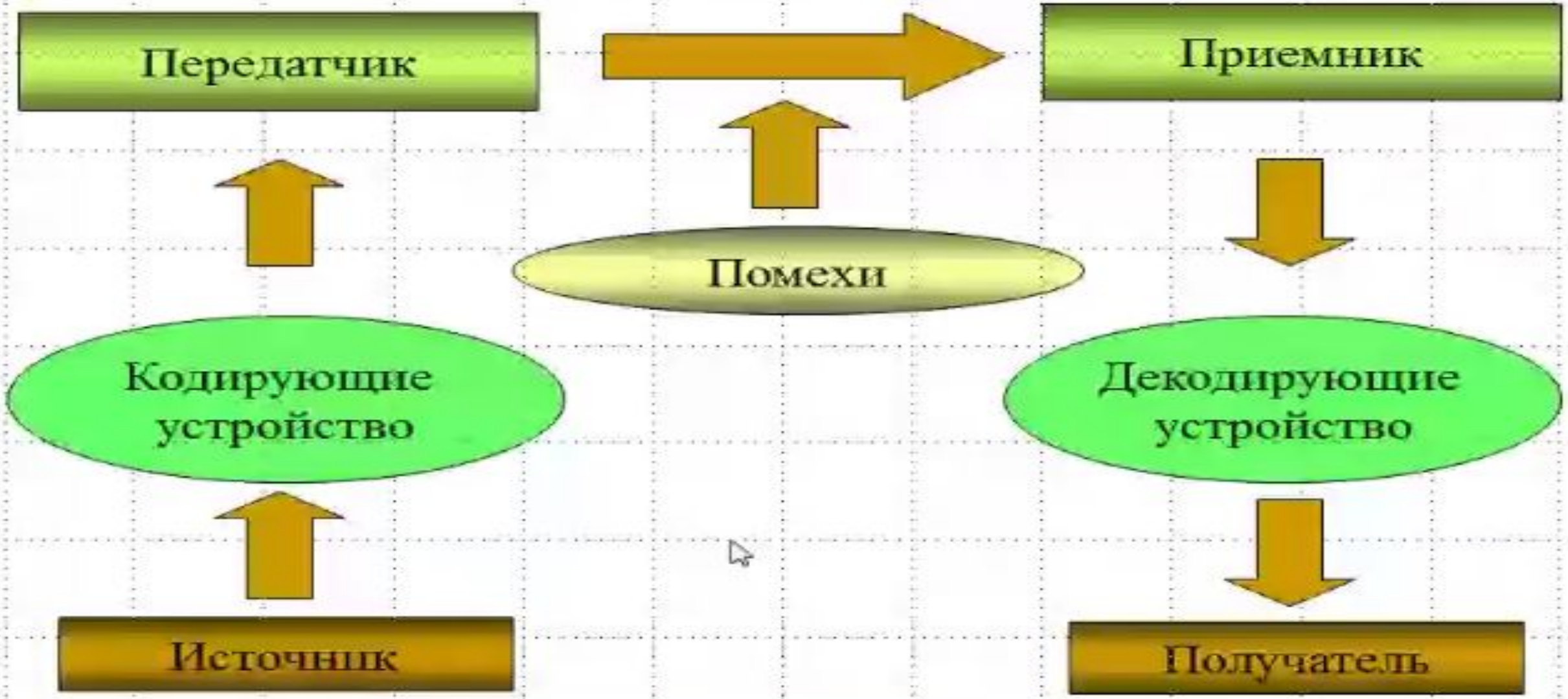
Схема передачи информации



Формальное определение алфавита

Алфави́т — конечное не пустое множество различных символов. Что такое символ — не определяется, как не определяется, например, точка в геометрии.

Символом (буквой или знаком) называется любой элемент x *алфавита* X

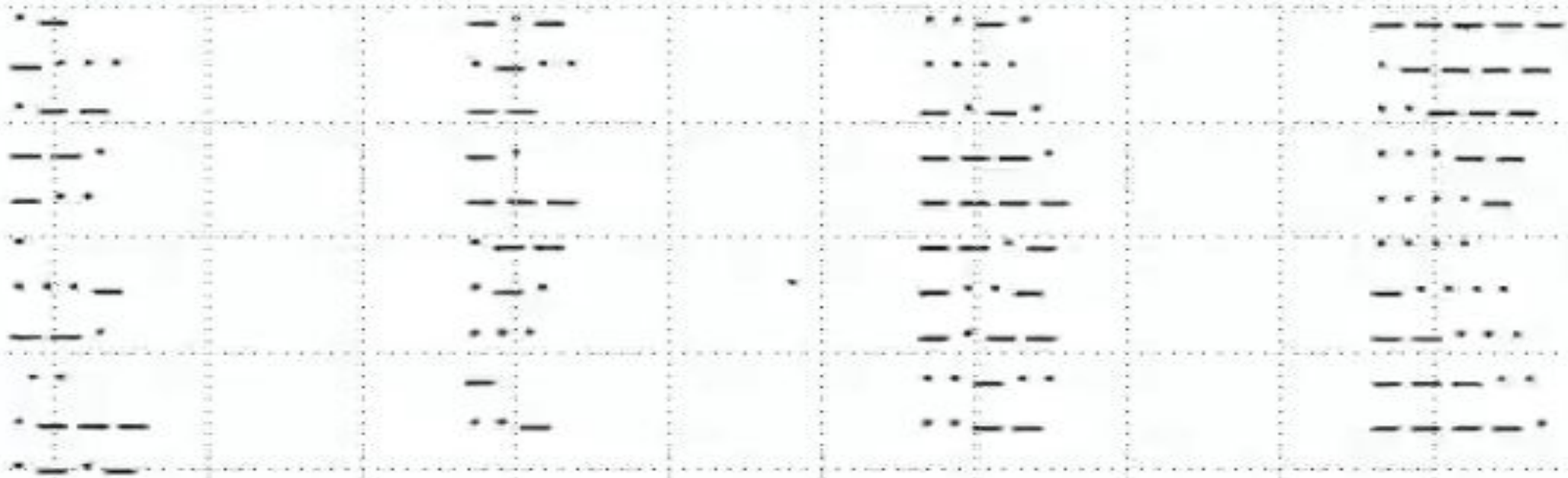
Для символов определена операция **конкатенации** (приписывания, присоединения символа к символу или цепочке символов), с помощью которой по определенным правилам соединения символов и слов можно получать *слова* (т.е. цепочки символов) и *словосочетания* (цепочки слов) в этом алфавите (над этим алфавитом).

Примеры алфавитов

- **алфавит прописных русских букв**

А Б В Г Д Е Ё Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я

- **алфавит Морзе**



- **алфавит арабских цифр**

0 1 2 3 4 5 6 7 8 9

Понятие слова

- Конечная последовательность символов алфавита называется ***словом*** над алфавитом.
- ***Длиной |p| некоторого слова*** *p* над алфавитом *X* называется число составляющих его символов.
- *Слово* (обозначаемое символом Ø) имеющее нулевую *длину*, называется *пустым* словом: $|Ø| = 0$.
- Множество различных *слов* над *алфавитом X* обозначим через *S(X)* и назовем **языком (словарным запасом)** над алфавитом *X*.
- В отличие от конечного алфавита, язык может быть и бесконечным.

Слова над некоторым заданным алфавитом определяют сообщения

- В *алфавите* должен быть определен порядок следования *букв* (порядок типа «предыдущий элемент – последующий элемент»), то есть любой *алфавит* имеет упорядоченный вид

$$X = \{x_1, x_2, \ldots, x_n\}.$$

- Алфавит должен позволять решать задачу лексикографического (алфавитного) упорядочивания, или задачу расположения *слов* над этим *алфавитом*, в соответствии с порядком, определенным в *алфавите* (то есть по символам *алфавита*).

Кодирование информации

Источник представляет сообщение в алфавите, который называется ***первичным***, далее это сообщение попадает в кодирующее устройство, преобразующее и представляющее его во ***вторичном алфавите***.

Код – правило, описывающее соответствие набора символов первичного алфавита символам вторичного алфавита.
Кодирование – процедура преобразования информации, представленной сообщением над первичным алфавитом, в сообщение над вторичным алфавитом.
Декодирование – операция обратная кодированию.
Кодер – устройство, обеспечивающее выполнение операции кодирования.
Декодер – устройство, производящее декодирование.

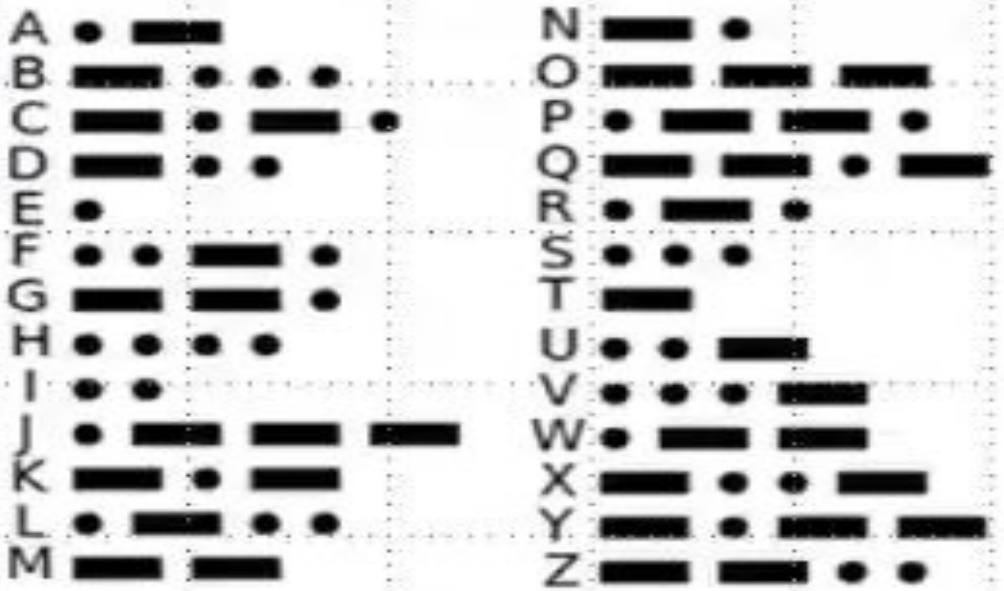
Операции кодирования и декодирования называются ***обратимыми,*** если их последовательное применение обеспечит возврат к исходной информации без каких-либо ее потерь.

Примеры кодов

- азбука Морзе

  … — — — …

  неравномерный код



- Код Трисиме (символам латинского алфавита ставятся в соответствие комбинации из трех символов: 1,2,3)

  31122331

  равномерный код, в котором кодовые комбинации содержат одинаковое число символов

A	111	J	211	S	311
B	112	K	212	T	321
C	113	L	213	U	313
D	121	M	221	V	321
E	122	N	222	W	322
F	123	O	223	X	323
G	131	P	231	Y	331
H	132	Q	232	Z	332
I	133	R	233		333

Избыточность информации

- Известно, что естественный язык обладает большой избыточностью. Этим объясняется помехоустойчивость сообщений, составленных из символов алфавитов таких языков.
- **Примеры избыточности**:

«в словох всо глосноо зомононо боквой о»

«По рзелульаттам илссеовадний одонго анлигйсокго унвиертисета, не иеемт занчнеия, в кокам пряокде рсапожолены бкувы в солве»

- Избыточность может использоваться при передачи кодированных сообщений в технических системах. Например, каждый фрагмент сообщения может передаваться трижды. Верным считаются полностью совпавшие фрагменты. Однако большая избыточность:
  - приводит к большим временным затратам при передаче информации,
  - требует большого объема памяти при ее хранении.

Математическая постановка задачи кодирования

- Пусть A – первичный алфавит, состоящий из N символов со средней информацией на символ I_A.
- B – вторичный алфавит из M символов со средней информацией на символ I_B.
- Пусть сообщение в первичном алфавите содержит n символов, а закодированное – m символов.
- Пусть $I_{исх}$ – информация в исходном сообщении, $I_{код}$ – информация в закодированном сообщении.

- Тогда условие обратимости кодирования (т.е. неисчезновения информации) имеет вид

$$I_{исх} \le I_{код}$$

операция обратимого кодирования может увеличить количество информации, но не может его уменьшить

$$n * I_A \le m * I_B$$

- Отношение m/n – характеризует среднее число символов вторичного алфавита, которое используется для кодирования одного символа первичного. Оно называется *длиной кода* и обозначается через $K(A,B)$. Следовательно,

$$K(A,B) \ge I_A / I_B > 1.$$

- Поскольку способов построения кодов при фиксированных алфавитах A и B существует множество, возникает проблема построения наилучшего варианта – *оптимального кода*. Минимальная средняя длинна кода

$$K_{min}(A,B) = I_A / I_B$$

- Когда возможно $K(A,B) \approx K_{min}(A,B)$?

Первая теорема Шеннона

(основная теорема о кодировании при отсутствии помех)

При отсутствии помех существует возможность кодирования сообщения, при которой среднее число символов кода, приходящихся на один символ первичного алфавита, будет сколь угодно близко к отношению средних информаций на символ первичного и вторичного алфавитов (т.е. избыточность кода будет сколь угодно близкой к нулю).

Пусть $M=2$. Существует возможность создания системы эффективного кодирования сообщений, у которой среднее число двоичных символов на один символ сообщения асимптотически стремится к энтропии источника сообщений.

$$A = \{a_i\} \rightarrow \text{кодирующее устройство} \rightarrow B = \{0,1\}$$

Требуется оценить минимальную среднюю длину кодовой комбинации.

Шенноном была рассмотрена ситуация, когда при кодировании сообщения в первичном алфавите учитывается *различная* вероятность появления символов, а также *равная* вероятность появления символов вторичного алфавита (по формуле Хартли $I_B = \log_2 M = 1$). **Тогда**

$$K_{\min}(A,B) = \frac{I_A}{I_B} = I_A$$

Вторая теорема Шеннона

При наличии помех в канале всегда можно найти такую систему кодирования, при которой сообщения будут переданы с заданной достоверностью (существует код, обеспечивающий передачу со сколь угодно малой частотой ошибок). **При этом пропускная способность канала должна превышать производительность источника сообщений (скорость создания сообщений).**

Таким образом, вторая теорема Шеннона устанавливает принципы помехоустойчивого кодирования.

Эта теорема не дает конкретного метода построения кода, но указывает на пределы достижимого в создании помехоустойчивых кодов, стимулирует поиск новых путей решения этой проблемы.

1.Способ кодирования только устанавливает факт искажения сообщения, что позволяет потребовать повторную передачу.
2.Используемый код находит и автоматически исправляет ошибку передачи.

Таблицы кодировки

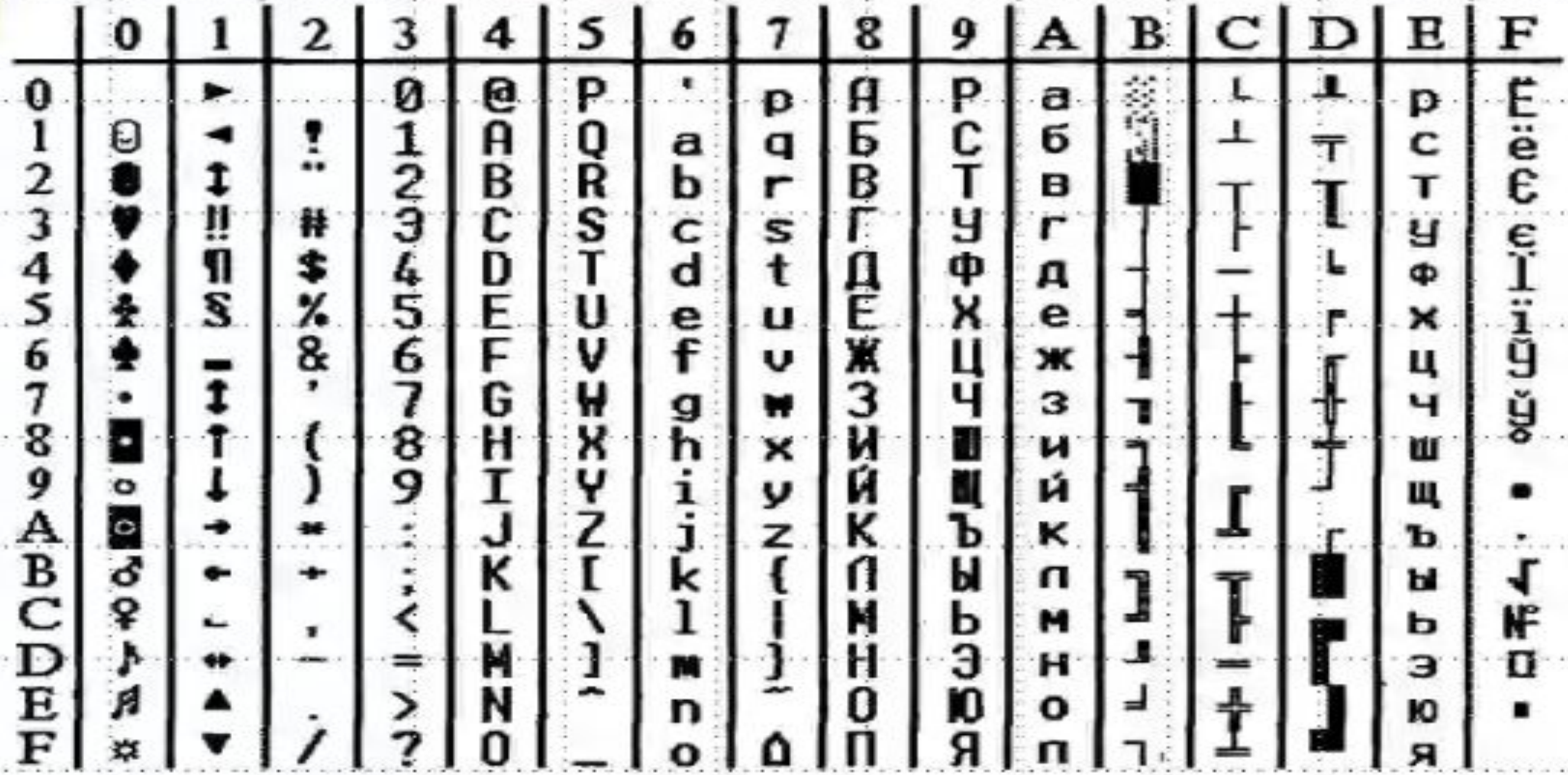
Таблица, в которой устанавливается однозначное соответствие между символами и их порядковыми номерами, называется **таблицей кодировки**.

Для разных типов ЭВМ используют различные таблицы кодировки:

ANSI - (American National Standards Institute)

ASCII - (American Standard Cod for Information Interchange)

Таблица кодировки ASCII



Знакам алфавита ЭВМ ставятся в соответствие шестнадцатиричные числа по правилу: первая - номер столбца, вторая - номер строки. Например, английская 'A' - код 41, русская 'и' - код A8.

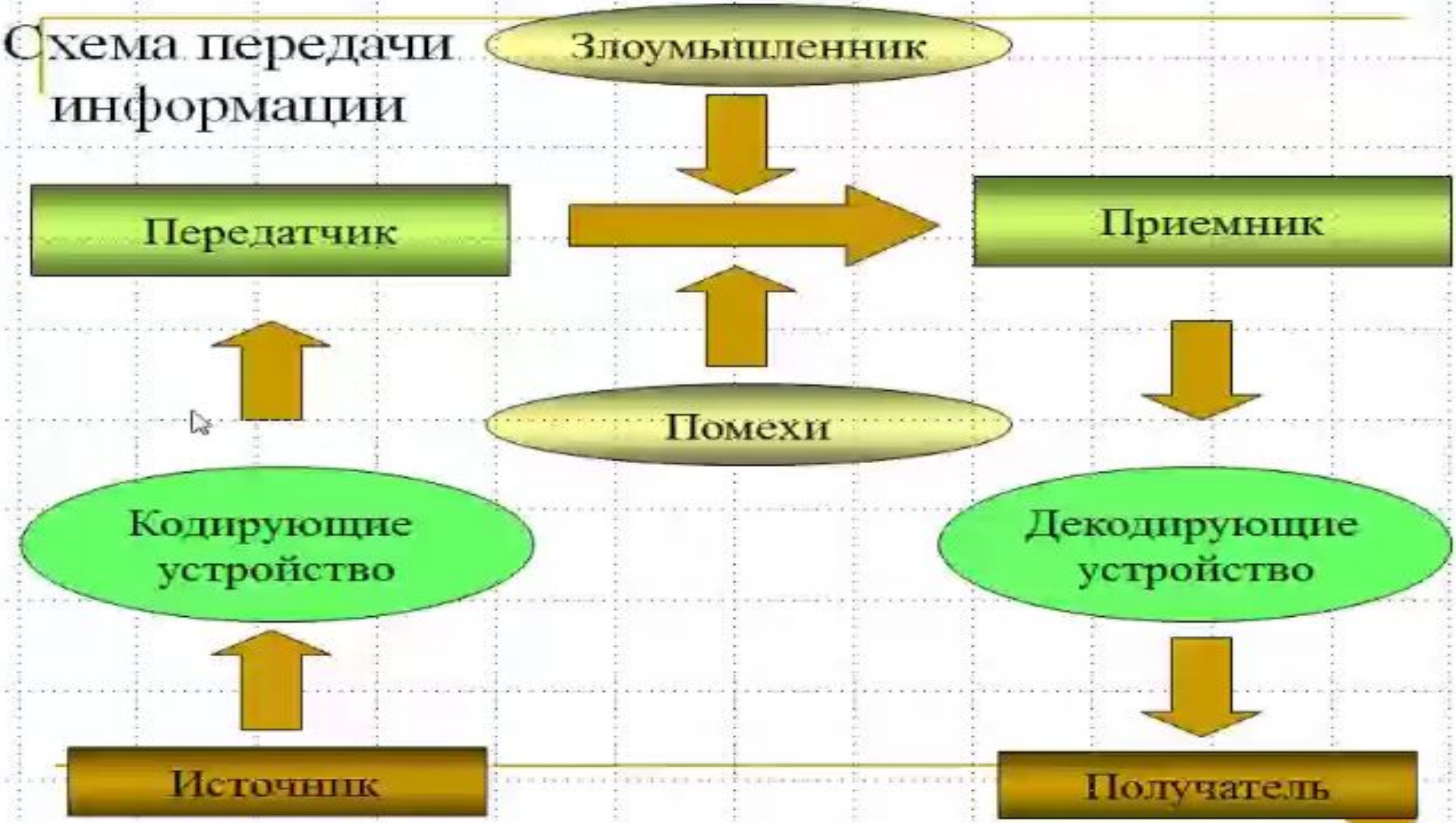
Системы кодирования

- КОИ-7
- Windows-1251
- КОИ-8
- ISO
- Unicode
- UTF-8

Неверная кодировка

Название	РЎРєР°С‡Р°С‚СЊ РїСЂРѕРіСЂР°РјРјС‹ Р±РµСЃРїР»Р°С‚РЅРѕ. РЎРєР°С‡Р°С‚СЊ РїРѕР»РµР·РЅС‹Рµ Рё РЅРѕРІС‹Рµ РєРѕРјРїСЊСЋС‚РµСЂРЅС‹Рµ РїСЂРѕРіСЂР°РјРјС‹.
Описание	Freesoftyk.ru - СѓРЅРёРєР°Р»СЊРЅС‹Р№ РєР°С‚Р°Р»РѕРі РЅРѕРІС‹С… Рё РїРѕР»РµР·РЅС‹С… РєРѕРјРїСЊСЋС‚РµСЂРЅС‹С… РїСЂРѕРіСЂР°РјРј СЃ РїРѕРґСЂРѕР±РЅС‹Рј РѕРїРёСЃР°РЅРёРµРј РєР°Р¶РґРѕР№, СЃСЃС‹Р»РєРё РґР»СЏ СЃРєР°С‡РёРІР°РЅРёСЏ РїСЂРѕРіСЂР°РјРј Рё СЃР°РјС‹Р№ СѓРґРѕР±РЅС‹Р№ РёРЅС‚РµСЂС„РµР№СЃ.....
Ключевые слова	СЃРєР°С‡Р°С‚СЊ РїСЂРѕРіСЂР°РјРјС‹ Р±РµСЃРїР»Р°С‚РЅРѕ, СЃРєР°С‡Р°С‚СЊ РєРѕРјРїСЊСЋС‚РµСЂРЅС‹Рµ РїСЂРѕРіСЂР°РјРјС‹, СЃРєР°С‡Р°С‚СЊ РїСЂРѕРіСЂР°РјРјС‹ РґР»СЏ С‚РµР»РµС„РѕРЅР°, QIP Infium 9024, ICQ 6.0, РїСЂРѕРіСЂР°РјРјС‹ РґР»СЏ web РјР°СЃС‚РµСЂР°, СЃРѕС„С‚,Р°РЅС‚РёРІРёСЂСѓСЃС‹,РёРЅС‚РµСЂРЅРµС‚, РјСѓР»СЊС‚РёРјРµРґРёСЏ,Р·Р°СЃС‚Р°РІРєРё, СЂР°Р±РѕС‡РёР№ СЃС‚РѕР»,Р±РµР·РѕРїР°СЃРЅРѕСЃС‚СЊ,СЃРёСЃС‚РµРјР°, РїРѕР»РµР·РЅС‹Рµ, СЃРєР°С‡Р°С‚СЊ СЃРѕС„С‚, РїСЂРѕРіРё ...

Схема передачи информации



Актуальность защиты информации

В современном обществе успех любого вида деятельности зависит от обладания определенными сведениями (информацией) и от отсутствия их (ее) у конкурентов.

Возникновение индустрии обработки информации привело к возникновению индустрии средств ее защиты и к актуализации самой проблемы защиты информации, проблемы *информационной безопасности*.

Одна из наиболее важных задач – задача *кодирования* сообщений и *шифрования* информации.

Вопросами защиты и скрытия информации занимается наука **криптология** (*криптос – тайный, логос – наука*).

1) **Криптография** (*криптос* и *графейн – писать*). **Цель:** построение и исследование математических методов преобразования информации. Криптография – это тайнопись, система перекодировки сообщения с целью сделать его непонятным для непосвященных лиц и дисциплина, изучающая общие свойства и принципы систем тайнописи.
2) **Криптоанализ**. **Цель:** исследованием возможности расшифровки информации без *ключа*.

Основные понятия *шифрования*

- Сообщение, которое мы хотим передать адресату, назовем **открытым** сообщением. Оно, естественно, определено над некоторым алфавитом.
- Зашифрованное сообщение может быть построено над другим алфавитом. Назовем его **закрытым** сообщением.
- Процесс преобразования открытого сообщения в закрытое сообщение и есть ***шифрование***.
- Если A – открытое сообщение, B – закрытое сообщение (***шифр***) , f – правило *шифрования*, то

$$\mathbf{f(A) = B.}$$

Шифрование

- Правила *шифрования* должны быть выбраны так, чтобы зашифрованное сообщение можно было расшифровать.
- Однотипные правила (например, все *шифры* типа шифра Цезаря, по которому каждый символ алфавита кодируется отстоящим от него на k позиций символом) объединяются в классы, и внутри класса определяется некоторый параметр (числовой, символьный табличный и т.д.), позволяющий перебирать (варьировать) все правила.
- Такой параметр называется шифровальным *ключом* . Он, как правило, секретный и сообщается лишь тому, кто должен прочесть зашифрованное сообщение (обладателю *ключа*).

Кодирование

- При кодировании секретного ключа нет, так как кодирование ставит **целью** лишь более **сжатое**, **компактное** представление сообщения.

Шифр

- Если k – *ключ*, то можно записать

$$\mathbf{f(k(A)) = B.}$$

- Для каждого *ключа* k, преобразование f(k) должно быть обратимым, то есть

$$\mathbf{f(k(B)) = A.}$$

- Совокупность преобразования f(k) и соответствия множества k называется ***шифром*** .

Две группы шифров: шифры перестановки и шифры замены

- *Шифр перестановки* изменяет только порядок следования символов исходного сообщения. Это такие *шифры*, преобразования которых приводят к изменению только следования символов открытого, исходного сообщения.
- *Шифр замены* заменяет каждый символ кодируемого сообщения на другой(ие) символ(ы), не изменяя порядок их следования. Это такие *шифры*, преобразования которых приводят к замене каждого символа открытого сообщения на другие символы, причем порядок следования символов закрытого сообщения совпадает с порядком следования соответствующих символов открытого сообщения.

Надежность

- Под ***надежностью*** понимается способность противостоять взлому *шифра*.
- При дешифровке сообщения может быть известно все, кроме *ключа*, то есть *надежность шифра* определяется секретностью *ключа*, а также числом его *ключей*. Применяется даже открытая криптография, которая использует различные *ключи* для *шифрования*, а сам *ключ* может быть общедоступным, опубликованным. Число *ключей* при этом может достигать сотни триллионов.

Пример

- Один из лучших примеров алгоритма шифрования – принятый в 1977 году Национальным бюро стандартов США алгоритм стандарта шифрования данных DES (Data Encrypted Standard).
- Исследования алгоритма специалистами показали, что пока нет уязвимых мест, на основе которых можно было бы предложить метод криптоанализа, существенно лучший, чем полный перебор ключей. В июле 1991 года введен в действие аналогичный отечественный криптоалгоритм (стандарта ГОСТ 28147-89), который превосходит DES по надежности.

Криптографическая система

- **Криптографическая система** – семейство X преобразований открытых текстов. Члены этого семейства индексируются, обозначаются символом k; параметр k является *ключом*.
- Множество *ключей* K – это набор возможных значений *ключа* k. Обычно *ключ* представляет собой последовательный ряд букв алфавита.
- Открытый текст обычно имеет произвольную длину. Если текст большой и не может быть обработан шифратором (компьютером) целиком, то он разбивается на блоки фиксированной длины, а каждый блок шифруется отдельно, независимо от его положения во входной последовательности. Такие криптосистемы называются системами *блочного шифрования*.

Криптосистемы:

- Симметричные с открытым *ключом* и системы электронной подписи.
- В **симметричных криптосистемах**, как для шифрования, так и для дешифрования, используется один и тот же *ключ*.
- В системах с открытым *ключом* используются два *ключа* – открытый и закрытый, которые математически (алгоритмически) связаны друг с другом. Информация шифруется с помощью открытого *ключа*, который доступен всем желающим, а расшифровывается лишь с помощью закрытого *ключа*, который известен только получателю сообщения.

Криптосистемы:

- **Электронной (цифровой) подписью (ЭЦП)** называется присоединяемое к тексту его криптографическое преобразование, которое позволяет при получении текста другим пользователем проверить авторство и подлинность сообщения.
- К ЭЦП предъявляются два основных требования:
  - *легкость проверки* подлинности подписи;
  - *высокая сложность* подделки подписи.

Криптосистемы:

- Системы управления *ключами* – это информационные системы, целью которых является составление и распределение *ключей* между пользователями информационной системы.

Криптосистемы:

- Разработка *ключевой*, парольной информации является типовой задачей администратора безопасности системы. *Ключ* может быть сгенерирован как массив нужного размера статистически независимых и равновероятно распределенных на двоичном множестве {0, 1} элементов.

Пример. Разработка ключей

- Для таких целей можно использовать программу, которая вырабатывает *ключ* по принципу "электронной рулетки". Когда число пользователей, то есть объем необходимой ключевой информации, очень большой, используют чаще аппаратные датчики случайных (псевдослучайных) чисел.
- Пароли также необходимо менять. Например, известный *вирус* Морриса пытается войти в систему, последовательно пробуя пароли из своего внутреннего эвристически составленного списка в несколько сотен процедур, имитирующих "сочинение" паролей человеком.

Принцип Кирхгоффса

- Все современные криптосистемы построены по **принципу Кирхгоффса**: секретность зашифрованных сообщений определяется секретностью *ключа*.
- Если даже алгоритм *шифрования* будет известен криптоаналитику, тот тем не менее не в состоянии будет расшифровать закрытое сообщение, если не располагает соответствующим *ключом*.
- Все классические *шифры* соответствуют этому принципу и спроектированы таким образом, чтобы не было пути вскрыть их более эффективным способом, чем полный перебор по всему ключевому пространству, то есть перебор всех возможных значений *ключа*. Ясно, что стойкость таких *шифров* определяется размером используемого в них *ключа*.

Пример.

- В российских *шифрах* часто используется 256-битовый *ключ*, а объем ключевого пространства составляет 2^{256}. Ни на одном реально существующем или возможном в недалеком будущем компьютере нельзя подобрать *ключ* (полным перебором) за время, меньшее многих сотен лет. Российский криптоалгоритм проектировался с большим запасом *надежности*, стойкости.