Представление информации, языки, кодирование

Из базового курса известно:

- Историческое развитие человека, формирование человеческого общества связано с развитием речи, с появлением и распространением языков. **Язык** — это знаковая система для представления и передачи информации.

- Люди сохраняют свои знания в записях на различных носителях. Благодаря этому знания передаются не только в пространстве, но и во времени — от поколения к поколению.

- Языки бывают **естественными**, например, русский, китайский, английский, и **формальные**, например, математическая символика, нотная грамота, языки программирования. Естественные языки развивались веками и служат для общения людей между собой. Формальные языки разрабатываются для специальных применений.

- Каждый язык имеет свой **алфавит**. Под алфавитом языка понимают набор используемых символов. Под **мощностью алфавита** понимают количество составляющих алфавит символов.

Под словом «**кодирование**» понимают процесс представления информации, удобный для её хранения и/или передачи. Следовательно, запись текста на естественном языке можно рассматривать как способ кодирования речи с помощью графических элементов (букв, иероглифов). Записанный текст является **кодом**, заключающим в себе содержание речи, т. е. информацию.

Код — система условных знаков (символов), предназначенных для представления информации в соответствии с определенными правилами.

Кодирование — переход от одной формы представления информации к другой, наиболее удобной для её хранения, передачи или обработки.

Декодирование — процесс по восстановлению первоначальной формы представления информации, т. е. операция, обратная кодированию.

При кодировании ставятся разные цели и, соответственно, применяются различные способы кодирования.

Наиболее распространенные цели кодирования:

1) **экономность** (сократить запись);

2) **надежность** (засекретить информацию);

3) **удобство обработки или восприятия**.

Чаще всего кодированию подвергаются тексты на естественных языках (русском, английском и пр.).

Существуют три основных способа кодирования текста:

1) **графический** — с помощью специальных рисунков или значков;

2) **числовой** — с помощью чисел;

3) **символьный** — с помощью символов того же алфавита, что и исходный текст.

Процесс чтения текста — это обратный по отношению к письму процесс, при котором письменный текст преобразуется в устную речь. Чтение можно назвать декодированием письменного текста.

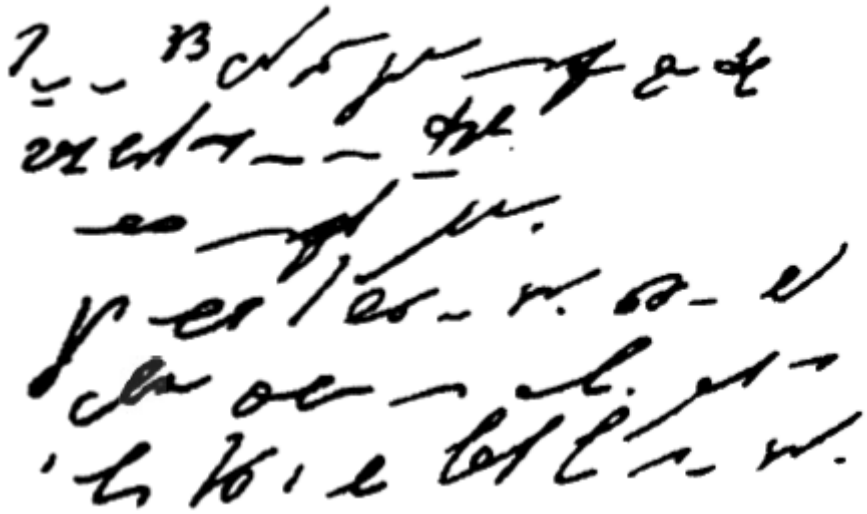
А теперь обратим внимание на то, что может существовать много способов кодирования одного и того же текста на одном и том же языке.

Пример:

Русский текст мы привыкли записывать с помощью русского алфавита. Но то же самое можно сделать, используя латинский алфавит. Иногда так приходится поступать, отправляя SMS по мобильному телефону, на котором нет русских букв, или электронное письмо на русском языке за границу, если у адресата нет русифицированного программного обеспечения. Например, фразу «Здравствуй, дорогой Саша!» приходится писать так: «Zdravstvui, dorogoi Sasha!».

Существует множество способов кодирования. Например, **стенография** — быстрый способ записи устной речи. Ею владеют лишь немногие специально обученные люди — стенографисты. Они успевают записывать текст синхронно с речью выступающего человека. В стенограмме один значок обозначает целое слово или сочетание букв. Скорость стенографического письма превосходит скорость обычного в 4-7 раз. Расшифровать (декодировать) стенограмму может только сам стенографист.

Пример стенографии, в которой написано следущее: «*Говорить умеют все люди на свете. Даже у самых примитивных племен есть речь. Язык — это нечто всеобщее и самое человеческое, что есть на свете*»:



Стенография — скоростное письмо, основанное на применении специальных систем знаков и сокращений слов и словосочетаний, позволяющее вести синхронную запись устной речи и рационализировать технику письма.

Приведённые примеры иллюстрируют следующее важное правило:

Для кодирования одной и той же информации могут быть использованы разные способы; их выбор зависит от ряда обстоятельств: цели кодирования, условий, имеющихся средств.

Если надо записать текст в темпе речи, делаем это с помощью стенографии; если надо передать текст за границу, пользуемся латинским алфавитом; если надо представить текст

в виде, понятном для грамотного русского человека, записываем его по правилам грамматики русского языка.

Еще одно важное обстоятельство:

Выбор способа кодирования информации может быть связан с предполагаемым способом её обработки.

Обсудим это на примере представления чисел — количественной информации. Используя русский алфавит, можно записать число «тридцать пять». Используя же алфавит арабской десятичной системы счисления, пишем: 35. Пусть нам надо произвести вычисления. Скажи, какая запись удобнее для выполнения расчётов: «тридцать пять умножить на сто двадцать семь» или «$35x127$»? Очевидно, что для перемножения многозначных чисел вы будете пользоваться второй записью.
Заметим, что две эти записи, эквивалентные по смыслу, используют разные языки: первая — естественный русский язык, вторая — формальный язык математики, не имеющий национальной принадлежности. Переход от представления на естественном языке к представлению на формальном языке можно также рассматривать как кодирование. Человеку удобно использовать для кодирования чисел десятичную систему счисления, а компьютеру — двоичную систему.

В некоторых случаях возникает потребность засекречивания текста сообщения или документа, для того чтобы его не смогли прочитать те, кому не положено. Это называется защитой от несанкционированного доступа. В таком случае секретный текст шифруется. В давние времена шифрование называлось тайнописью.

Шифрование представляет собой процесс превращения открытого текста в зашифрованный, а дешифрование — процесс обратного преобразования, при котором восстанавливается исходный текст. Шифрование — это тоже кодирование, но с засекреченным методом, известным только источнику и адресату. Методами шифрования занимается наука криптография.

Для осуществления шифрования используются специальные математические алгоритмы (криптоалгоритмы). Шифрование гарантирует защиту секретной информации от несанкционированного доступа со стороны третьих лиц. Для восстановления зашифрованной информации осуществляется обратное преобразование — **расшифровка**. Для расшифровки информации необходимо наличие соответствующего секретного **ключа**.
Подробнее об основных алгоритмах шифрования можно узнать здесь.

Шифрование — медод защиты любой информации от несанкционированного доступа, просмотра, а также её использования, основанный на преобразовании данных в зашифрованный формат.

Криптография — это наука о методах и принципах передачи и приема зашифрованной с помощью специальных ключей информации.

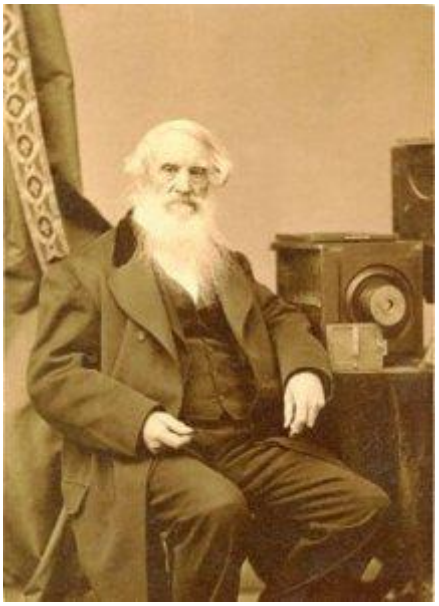
Ключ — секретная информация, используемая криптографическим алгоритмом при шифровании/расшифровке сообщений.

История технических способов кодирования информации

С появлением технических средств хранения и передачи информации возникли новые идеи и приемы кодирования.

Первым техническим средством передачи информации на расстояние стал телеграф, изобретенный в 1837году американцем Сэмюэлем Морзе.
Телеграфное сообщение — это последовательность электрических сигналов, передаваемая от одного телеграфного аппарата по проводам к другому телеграфному аппарату.

Эти технические обстоятельства привели Морзе к идее использования всего двух видов сигналов — короткого и длинного — для кодирования сообщения, передаваемого по линиям телеграфной связи.



Сэмюэль Финли Бриз Морзе
(1791-1872), США

Такой способ кодирования получил название **азбуки Морзе**. В ней каждая буква алфавита кодируется последовательностью коротких сигналов (точек) и длинных сигналов (тире). Буквы отделяются друг от друга паузами — отсутствием сигналов. В кодовой таблице ниже показана азбука Морзе применительно к русскому алфавиту. Специальных знаков препинания в ней нет. Их обычно записывают словами: «тчк» — точка, «зпт» — запятая и т. п.

Азбука Морзе				
А • —	К — • —	Ф •• — •	1 •— — — —	. • • • • • •
Б — • • •	Л •— ••	Х • • • •	2 •• — — —	, • — • — • —
В • — —	М — —	Ц — • — •	3 • • • — —	; — • — • — •
Г — — •	Н — •	Ч — — — •	4 • • • • —	: — — — • • •
Д — ••	О — — —	Ш — — — —	5 • • • • •	? •• — — ••
Е •	П • — — •	Щ — — • —	6 — • • • •	! — — •• — —
Ж • • • —	Р • — •	Ь,Ъ — •• —	7 — — • • •	- — • • • • —
З — — ••	С • • •	Ы — • — —	8 — — — ••	« • — •• — •
И ••	Т —	Э •• — ••	9 — — — — •	(— • — — • —
Й • — — —	У •• —	Ю •• — —	0 — — — — —	/ — •• — •
		Я • — • —		

Кодовая таблица — это соответствие между набором знаков (символов) и их кодами.

Самым знаменитым телеграфным сообщением является сигнал бедствия «**SOS**» (**S**ave **O**ur **S**ouls - *спасите наши души*).

Вот как он выглядит в коде азбуки Морзе: ● ● ● ▬ ▬ ▬ ● ● ●
Три точки обозначают букву S, три тире — букву О. Две паузы отделяют буквы друг от друга.

Характерной особенностью азбуки Морзе является переменная длина кода разных букв, поэтому код Морзе называют **неравномерным кодом**. Буквы, которые встречаются в тексте чаще, имеют более короткий код, чем редкие буквы. Например, код буквы «Е» — одна точка, а код буквы «Ъ» состоит из шести знаков. Зачем так сделано? Чтобы сократить длину всего сообщения. Но из-за переменной длины кода букв возникает проблема отделения букв друг от друга в тексте. Поэтому приходится для разделения использовать паузу (пропуск). Следовательно, телеграфный алфавит Морзе является троичным, так как в нём используется три знака: точка, тире, пропуск.

Азбука Морзе — неравномерный телеграфный код, где каждая буква и знак представлены при помощи длинных и коротких сигналов, так называемых «тире» и «точек».

Равномерный телеграфный код был изобретен французом Жаном Морисом Бодо в конце XIX века. В нём использовалось всего два вида сигналов. Неважно, как их назвать: точка и тире, плюс и минус, ноль и единица. Это два отличающихся друг от друга электрических сигнала.

Жан Морис Эмиль Бодо
(1845-1903), Франция

В коде Бодо длина кодов всех символов алфавита одинакова и равна пяти. В таком случае не возникает проблемы отделения букв друг от друга: каждая пятерка сигналов — это знак текста.

Код Бодо — это первый в истории техники способ двоичного кодирования информации. Благодаря идее Бодо удалось автоматизировать процесс передачи и печати букв. Был создан клавишный телеграфный аппарат. Нажатие клавиши с определенной буквой вырабатывает соответствующий пятиимпульсный сигнал, который передается по линии связи. Принимающий аппарат под воздействием этого сигнала печатает ту же букву на бумажной ленте.

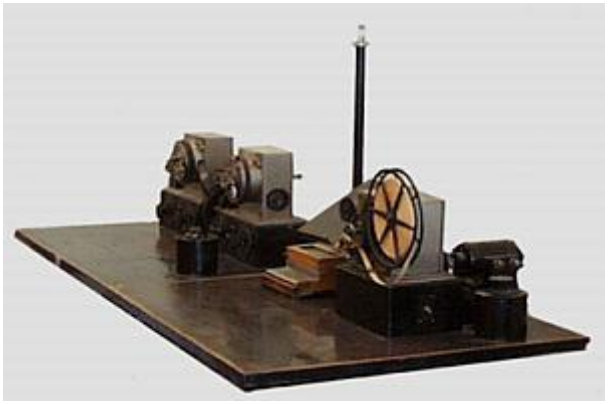
Код Бодо — равномерный телеграфный 5-битный код, использующий два отличающихся друг от друга электрических сигнала.

Оригинальный код Бодо

Управляющие символы	
o. ...	пробел, перейти к таблице букв
.o ...	пробел, перейти к таблице цифр
oo ...	удалить последний знак

таблица букв						таблица цифр					
.. o..	A	oo o..	K			.. o..	1	o. o..	.		
.. oo.	É	oo oo.	L			.. .o.	2	o. .o.	9/		
.. .o.	E	oo .o.	M			.. ..o	3	o. ..o	7/		
.. .oo	I	oo .oo	N			.. o.o	4	o. o.o	2/		
.. ooo	O	oo ooo	P			.. ooo	5	o. ooo	'		
.. o.o	U	oo o.o	Q			.. oo.	1/	o. oo.	:		
.. ..o	Y	oo ..o	R			.. .oo	3/	o. .oo	?		
.o ..o	B	o. ..o	S			.o o..	6	oo o..	(		
.o o.o	C	o. o.o	T			.o .o.	7	oo .o.	)		
.o ooo	D	o. ooo	V			.o ..o	8	oo ..o	-		
.o .oo	F	o. .oo	W			.o o.o	9	oo o.o	/		
.o .o.	G	o. .o.	X			.o ooo	0	oo ooo	+		
.o oo.	H	o. oo.	Z			.o oo.	4/	oo oo.	=		
.o o..	J	o. o..	—			.o .oo	5/	oo .oo	£		

В 1872 г. Жан Бодо конструирует телеграфный аппарат многократного действия, при помощи которого можно передавать по одной линии два и более сообщения.
Кроме своего аппарата, Бодо создает весьма удачный телеграфный код (Код Бодо), который впоследствии принимается повсеместно и получает звание Международного телеграфного кода № 1. Через два года изобретатель модернизирует свое изобретение и создает двукратный аппарат, который передаёт информацию со скоростью 360 знаков в минуту. Ещё через два года он создает уже пятикратный аппарат, скорость передачи в котором увеличивается по сравнению с первым уже в пять раз. Первые подобные аппараты вводятся в эксплуатацию в 1877 году на линии Париж – Бордо.



Аппарат Бодо становится самым значимым достижением техники передачи информации на расстоянии в XIX веке. В честь Бодо называют единицу скорости передачи информации — **Бод**.