

Содержание:

Введение

Важнейшей частью такой науки как информатики является теория информации, что занимается изучением принципов информации как таковой, появлением ее, развитием и уничтожением. Для этой науке очень близко примыкает и теория кодирования, в задачи которой входит изучение разных форм представления информации в ее передаче по самым различным каналам связи.

Тема, рассматриваемая в данном курсовом исследовании – актуальна, поскольку кодирование информации – это одна из самых первых обязанностей программистов для обеспечения надежной работы проектируемой информационной системы.

На первый взгляд понятие кодирования информации – это очень сложное дело, с которым далеко не всем справиться.

Но все это может рассматриваться как увлекательный и захватывающий процесс.

Разные учащиеся с удовольствием рассматривают, изучают методы расшифровки текстов, создают собственные шифры, изучают и рассматривают шифрующие устройства по кодированию и декодированию информации, получившие свое последующее развитие еще в эпоху Возрождения.

Целью работы является описание принципов и методов кодирования данных.

Исходя из цели в работе необходимо выполнить задачи:

- выполнить обработку литературы по теории кодирования;
- рассмотреть основные определения, а также классификацию методов кодирования;
- дать характеристику процессу кодирования информации разных видов информации
 - рассмотреть основные системы для выполнения кодирования;
 - описать особенности развития систем кодирования.

Объектом исследования является теория цифрового кодирования информации.

Предмет исследования – анализ методов кодирования информации и сообщений.

По мере развития современной вычислительной техники возникали также разные методики кодирования, которые исследовались разными учеными: Динман М.И. [6], Р. Лафоре, [13], Харви Дейтел [10].

Глава 1. Основы теории кодирования данных

Основные определения

Информация – сведения об объектах или явлениях окружающей среды, а также их параметрах, свойствах, состояниях, которые уменьшают уже имеющуюся степень неопределенности или неполноты знаний.[5]

В процессе обработки данных информация может менять структуру или форму. Признаком такой структуры являются элементы специальной информации, а также их взаимосвязь.

Все формы представления информации также могут быть различны. Главными из них являются следующие:[5]

- символная;
- жестов и мимики;
- текстовая;
- световых и звуковых сигналов;
- графическая;
- электрических или нервных импульсов;
- радиоволн;
- магнитных записей;
- запахов или вкусовых ощущений и т.п.

В нынешней повседневной практике такие основные понятия, как информация или данные, часто рассматриваются в качестве синонимов. На самом деле имеются между ними существенные различия.

Данными является информация, представленная в самом удобном для их обработки виде. Все данные могут быть также представлены в виде текстов, графики, аудиовизуального ряда. [10]

Любое представление данных называется специальным языком информатики, что представляются совокупностью символов, соглашений и определенных правил, используемых для тесного общения, отображения, а также передачи информации в цифровом виде.

Стоит отметить, что информация имеет несколько основных свойств (рисунок 1):



Рисунок 1 - Свойства информации

Информация может классифицироваться по следующим признакам (рисунок 2):

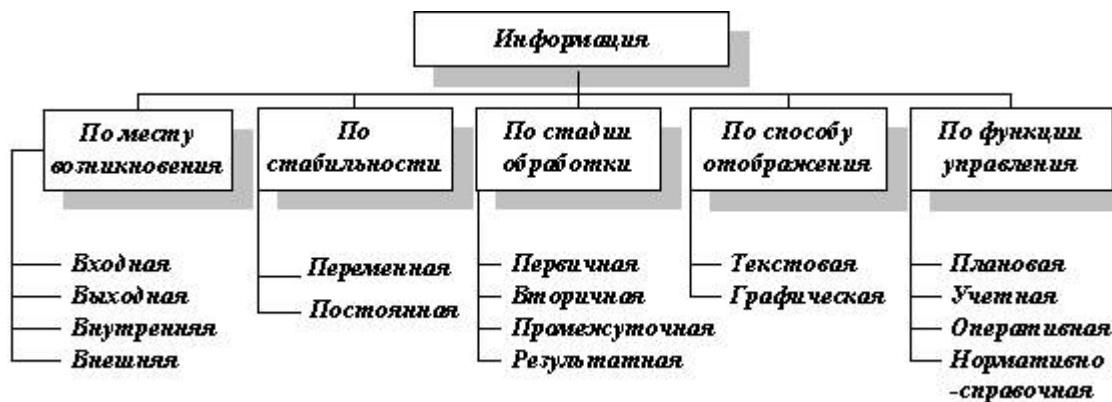


Рисунок 2 – Классификация информации

Все люди имеют дело с многими видами информации. Когда услышав прогноз погоды, можно также записать его в персональный компьютер (ПК), чтобы затем им воспользоваться.

В ПК можно поместить и фотографию своего друга и видеосъемку о проведенных каникулах. Но ввести в ПК вкус мороженого или же мягкость покрывала – нельзя.

Компьютером является электронная машина, которая выполняет работу с сигналами. Компьютер также может работать лишь с такой информацией, что можно превратить в некоторые сигналы. Если бы все люди умели превращать в специальные сигналы вкус и запах, то ПК мог бы работать с такой информацией. [4]

Для компьютера очень хорошо может получаться работать с числами и символами. Он может делать все, что угодно с ними. Все числа в ПК закодированы "двоичным кодом", представлены с помощью двух символов 1 или 0, которые легко представлены сигналами.

Вся информация для которой работает КП кодируется числами. Независимо от этого, графическая, текстовая, звуковая ли эта информация, что ее мог бы обрабатывать центральный процессор ПК она должна тем и иным образом может быть представлена числами.

Для преобразования числовой, графической, текстовой, звуковой информации в бинарную необходимо применить принцип кодирования

Кодирование – преобразование данных с одного типа через информацию другого типа.

В ПК применяется система для двоичного кодирования, основанная на принципах представления данных последовательностью 2-х знаков: 1 или 0, которые называются еще двоичными цифрами.

На рисунке 3 показан пример кодирования элементарной информации:



Рисунок 3 – Пример кодирования информации

1.2. Кодирование основных типов информации

Множество символов, что используются при записи текста, называют алфавитом. Количество таких символов в алфавите называют его мощностью.

Для такого представления информации на ПК в виде текстов чаще всего используется специальный алфавит мощностью в 256 символа.

Один символ с такого алфавита несет сразу 8 бит информации, поскольку $2^8 = 256$.

Но также 8 бит составляют и один байт, следовательно, бинарный код каждого символа будет занимать только 1 байт памяти ПК. Все символы этого алфавита пронумерованы с 0 до 255, каждому номеру соответствует свой 8-разрядный двоичный код с 00000000 до 11111111. [8]

Такой код является порядковым значением символа в бинарной системе счисления.

Для самых разных типов ПК и операционных систем (ОС) используются различные таблицы для кодировки, отличающиеся порядком размещения всех символов алфавита в таблице. Международным стандартом в ПК является таблица ASCII (рисунок 4).

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	.A	.B	.C	.D	.E	.F
0.		☺	☹	♥	♦	♣	♠	•	◼	○	◉	♂	♀	♪	♫	⚙
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.	▶	◀	↕	!!	¶	§	—	↕	↑	↓	→	←	↵	↔	▲	▼
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2.		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
3.	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
4.	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
5.	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
6.	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
7.	p	q	r	s	t	u	v	w	x	y	z	{		}	~	␣
	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127

Рисунок 4 – Пример таблицы ASCII

Принцип последовательного процесса кодирования алфавита заключается также в том, что в приведенной кодовой таблице ASCII все латинские буквы (и прописные, и строчные) располагаются только в алфавитном порядке.

Непосредственное расположение цифр также может быть упорядочено в направлении возрастания значений.[12]

Стандартными в такой таблице являются лишь первые 128 символов, поскольку символы с номерами с нуля (бинарный код 00000000) и до 127 (01111111).

Входят сюда буквы латинского алфавита, знаки препинания, цифры, скобки и некоторые иные символы. Остальные 128 символов, начиная с 128 (код 10000000) и кончая значением 255 (11111111), используются только для кодировки букв для национальных алфавитов, символов из псевдографики и разных научных символов.

Сейчас есть несколько различных таблиц для кодировки русских букв (CP-1251, CP-866, KOI-8, Mac, ISO), причем все тексты, созданные также в одной из кодировок,

могут неправильно отображаться также в другой. Решается и такая проблема с использованием специальных программ перевода текстов из одной кодировки сразу в другую.

В ОС Windows пришлось передвинуть также русские буквы в таблицах на место псевдографики, получили кодировку Win-1251.

В течение многого времени термины «байт» и «символ» стали почти синонимами. Но, в конце концов, было ясно, что сразу 256 различных символов – не очень много.

Математикам надо использовать в формулах также специальные математические знаки, а переводчикам необходимо создавать иногда тексты, где могут встречаться символы из самых различных алфавитов, экономистам надобны символы валют (£, \$, ¥). [3]

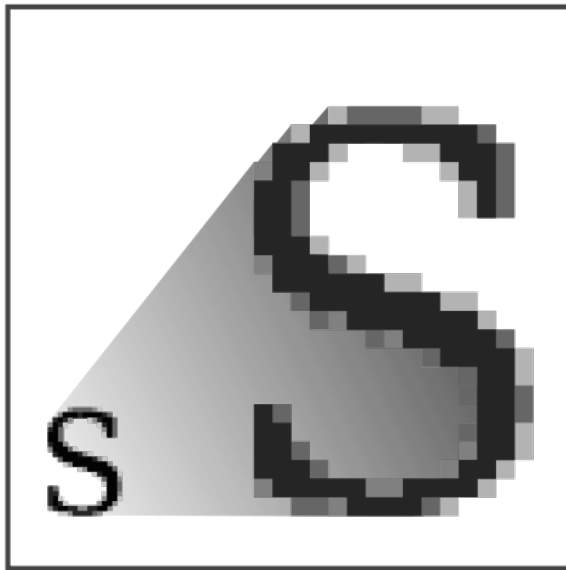
Для решения такой проблемы была разработана универсальная система для кодирования текстовой информации под названием Unicode. В данной кодировке для каждого символа будет отводиться не один, сразу два байта, то есть шестнадцать бит.

Графический формат – способ записи графических данных. Графические форматы файлов также предназначены для хранения разных изображений, таких как рисунки и фотографии.

Почти все создаваемые, а также обрабатываемые или просматриваемые при использовании компьютера изображения можно также разделить на 2 большие части, а именно векторную и растровую графику.

Для представления графических данных растровым способом используется точечный подход. На самом первом этапе горизонтальными и вертикальными линиями делят изображение. Чем больше получилось элементов (пикселей) при этом, тем точнее будет передаваться информация об изображении.

На рисунке 5 показаны различия в принципах кодирования графической информации:



РАСТР
.jpeg .gif .png



ВЕКТОР
.svg

Рисунок 5 – Примеры векторной и растровой графики

Любой цвет может представляться в виде суммы самой различной яркости красного, синего и зеленого цветов. Поэтому надо кодировать информацию о яркостях каждого из 3-х цветов для отображения пикселя. В видеопамати будет находиться двоичная информация на изображении, выводимом на дисплей.[6]

Таким образом, все растровые изображения представляются однослойной сеткой точек, называемых пикселями, а код пикселя может содержать информацию о цвете.[3]

Для черно-белого изображения пиксель может принимать лишь два значения: черный и белый (не светится и светится), а для его закодирования достаточно только одного бита.

Пиксель на цветном дисплее имеет различную окраску, поэтому только одного бита на пиксель будет недостаточно. Для кодирования, к примеру, 4-цветного изображения нужны два бита на один пиксель, поскольку 2 бита могут принимать 4 различных состояния.

На RGB-мониторах разнообразие всех цветов получается сочетанием основных цветов: красного, зеленого, синего, из которых можно получать 8 основных комбинаций.

Из курса физики уже известно, что звук – колебание частиц воздуха, а также непрерывный сигнал с изменчивой амплитудой.

При кодировании звуковой информации этот сигнал надо представлять в виде последовательности единиц и нулей.

Через равные промежутки по времени, очень часто измеряется амплитуда колебаний. Все измерения производится с точностью и записывается с помощью двоичного вида. Частоту, с которой и записывается амплитуда, называются частотой дискретизации. Полученный сигнал сначала сглаживается с помощью аналогового фильтра, а потом преобразуется в звук усилителями и динамиками.[9]

На качество воспроизведения кодированного звука в основном влияют 2 параметра: частота дискретизации – это количество измерений амплитуды в секунду в герцах, а также глубина кодирования звука – это размер в битах, что отводится под запись значений амплитуды.

К примеру, при записи на CD-диски используются 16-разрядные их значения, а частота для дискретизации равна 44030 Гц. Такие параметры обеспечивают отличное качество звучания музыки и речи. Для стереозвука отдельно также записывают данные с левого и правого канала.

Если преобразовывать звук в электрические сигналы (к примеру, с помощью микрофона), будет выполнено плавно изменяющееся с течением некоторого времени напряжение. Для обработки такой сигнал нужно преобразовать каким-то образом в последовательность бинарных чисел.

Для того чтобы воспроизводить закодированный звук, нужно выполнять обратное преобразование, а затем сглаживая получившийся ступенчатый сигнал.

Чем частота дискретизации выше (количество отсчетов в секунду) и чем больше таких разрядов отводится для отсчета, тем точнее представлен звук. Также при этом увеличивается размер звукового файла. Поэтому для хранения характера звука, требований, что предъявляются к его качеству или объему занимаемой памяти, могут выбирать некоторые компромиссные значения параметров.

Описанный способ для кодирования звуковой информации универсален, он позволяет представлять любой звук и его преобразовывать самыми разными методами (рисунок 6).[2]

Кодирование звуковой информации

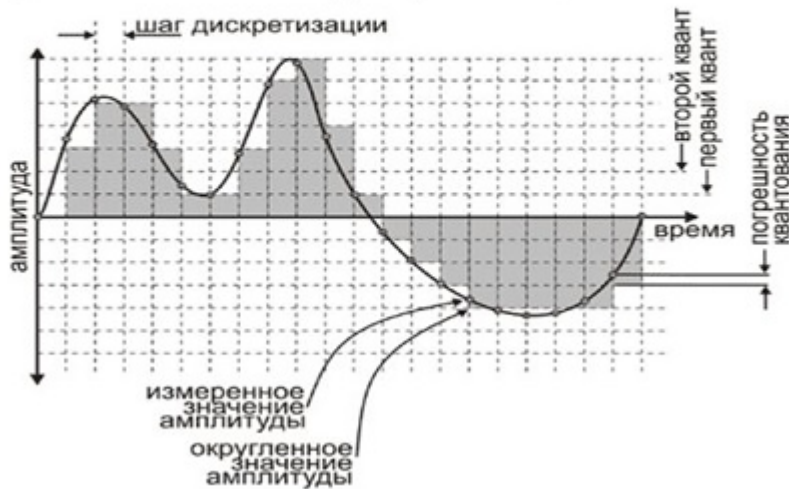


Рисунок 6 – Принцип кодирования звуковой информации

В результате написания первого раздела курсовой работы подробно рассмотрены все основные определения теории кодирования, представлены методы классификации для самых популярных видов информации.

Глава 2. Анализ методов кодирования данных

2.1. Системы кодирования

Под кодированием понимают процесс присвоения условных кодов признакам объектов. Процесс кодирования используется для представления данных в ПК при хранении, обработке и передаче информации. Вопросам кодирования, связанным хранением и передачей данных (представление в ПК, шифрование, сжатие и помехозащищённость), посвящены соответствующие главы теории информации.

Код строится на основании алфавита, что может состоять с букв, цифр и иных символов. Наибольшее распространение для обработки информации с

использованием ЭВМ получили специальные цифровые коды. Код может характеризоваться длиной, структурой или степенью информативности.

Длина – это количество позиций в коде.

Структура – порядок расположения символов в кодах, используемых в обозначении классификационного признака.

Степени информативности – отношение числа закодированных признаков к длине кода.

К системам для кодирования также предъявляются и следующие такие требования:

- отображение необходимой информации по объектах;
- идентификация каждого из объектов кодируемого множества;
- достаточная их гибкость и резерв для кодовых обозначений при самой минимальной длине;
- ориентация на обработку информации в автоматическом режиме и возможность обнаружения ошибок.[6]

Все системы кодирования информации распределяют на 2 основных вида (рисунок 7):

- регистрационные системы кодирования, что не требуют предварительной классификации;
- классификационные системы кодирования, что ориентированные на выполнение предварительной классификации для объектов.

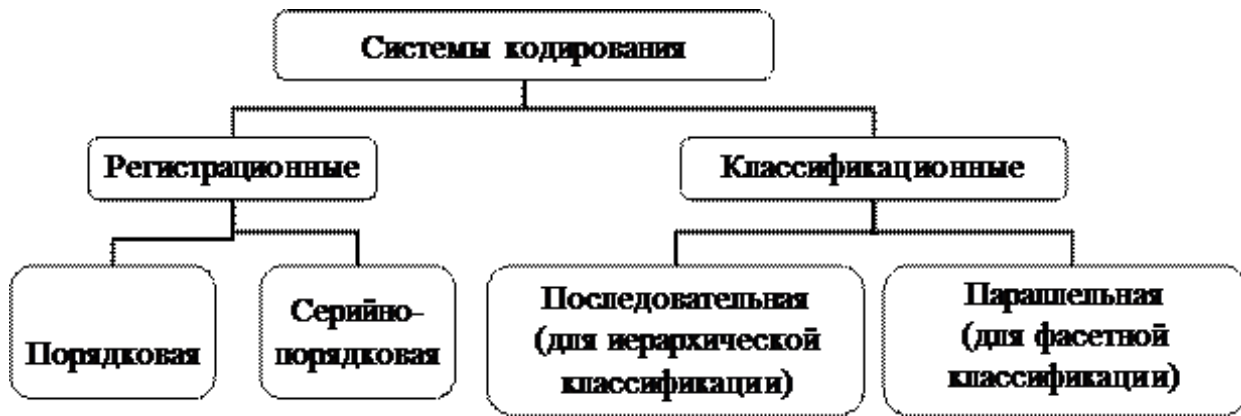


Рисунок 7 –

Классификация систем кодирования

Особо надо выделить комбинированные системы для кодирования, представляющие собой совокупность нескольких различных систем для кодирования. Они используются также для кодирования больших объектов, которые можно сгруппировать по нескольким из подчинённых или независимых признаков. Комбинированные системы дают возможность получать более информативный код.

2.2. Порядковые методы кодирования

Порядковая система для кодирования была применена на самом начале зарождения кодирования информации и предварительной классификации объектов.

Она основана на самом простом методе. Суть его заключается в постепенной нумерации объектов с помощью натуральных чисел. Таким образом, каждый из объектов идентифицируется регистрационным номером. Объект при регистрации может располагаться случайно и в определённом порядке, к примеру, алфавитном (список студентов) или хронологическом (журнал для регистрации событий).[1]

Принцип порядкового кодирования показан на рисунке 8:

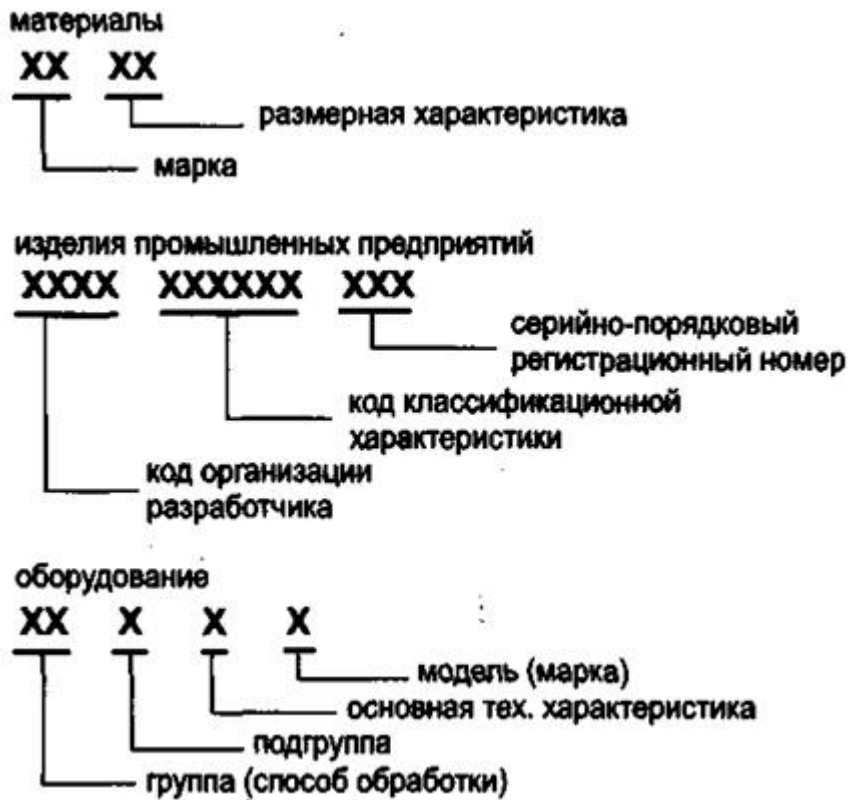


Рисунок 8 – Пример порядкового кодирования

Количество всех объектов номенклатуры N , что может быть закодировано при помощи порядковой системы, зависит также от количества символов, составляющих алфавит кода, длины кодового обозначения (количества знаков), что выражается такой формулой:

$$N = A^l \quad (1)$$

Например, с применением алфавита из 10 цифр и 3-значного кода можно закодировать 10^3 объектов.[7]

Достоинствами порядковой системы считаются простота кодирования, минимальная длина кода. Такая долговечность порядкового кода всегда достигается практически за счёт незначительной ее избыточности, поскольку добавление только 1 разряда увеличивает ёмкость в A раз.

Недостатком для порядковой системы кодирования считается отсутствие в коде данных о свойствах объектов.

2.3. Серийно-порядковое кодирование

Следующий этап развития, что используется кодировании отобразился в серийно-порядковой системе.

Данный метод отличается от обычного порядкового кодирования только тем, что объекты разбиваются предварительно на подмножества, а другие порядковые номера – на некоторые серии. Элементы для каждого подмножества в пределах отведённой серии регистрируются. В каждой серии предусматриваются обычно и резервные номера.

По своей сути такая система является смешанной, а именно классифицирующей и идентифицирующей. Часто применяется и в тех случаях, при которых деление на подмножества может осуществляться только по одному из классификационных признаков, а их количество будет малым (рисунок 9).

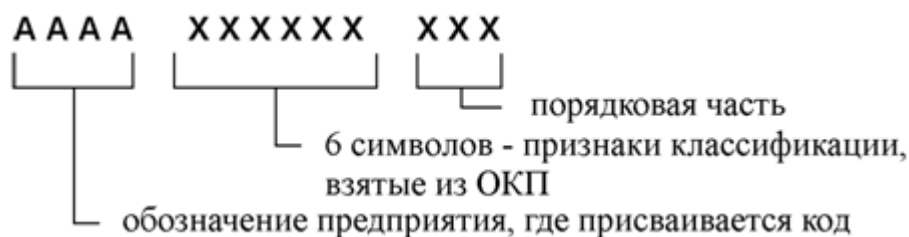


Рисунок 9 – Пример метода кодирования

Например, для нумерации аудиторий ВУЗа, размещённого в 4-х небольших близкорасположенных зданиях, есть по 20 используемых помещений для занятий, можно использовать 4 серии по 25 номеров (25–49; 00–24; 50–74; 75–99).

Также в каждой из серий остаётся резерв из 5 номеров (на случай ее перепланировки), а длина такого кода аудитории остаётся все же прежней – двузначной.[11]

Иной пример: использование «нечётных» серии для нумерации адресов домов, расположенных с левой, а «чётной» – с правой стороны улицы.

Серийно-порядковой системе также присущи в основном все те же достоинства или недостатки, что и обычному порядковому кодированию. Но тут коды, помимо идентификации объектов, несут некоторую дополнительную хранимую информацию, хотя восприятие её человеком может являться затруднено.

Резервирование является достоинством, что обеспечивает некоторую долговечность кода, а серийно-порядковая система будет непригодной из-за переполнения одной серии, при этом все резервы других серий не будут ещё

исчерпаны.

2.4. Последовательное кодирование

Последовательное кодирование, аналогично параллельному, требует проведения предварительного процесса классификации кодируемых объектов.

Такие классификационные системы кодирования часто называются также позиционными, ведь для выражения каждого из классификационных признаков в структуре кода выделяются позиции (группа символов) из нескольких разрядов.

Последовательное глубокое кодирование используется для реализации иерархической классификационной структуры.

Сама суть метода заключается в том, что сначала записывается код для старшей группировки из 1-го уровня, затем код для группировки 2-го уровня, а далее код группировки 3-го уровня и т.п.

В результате получим кодовую комбинацию, каждый разряд которой будет содержать информацию о специфике всей выделенной группы на каждом из уровней иерархической структуры (рисунок 10).

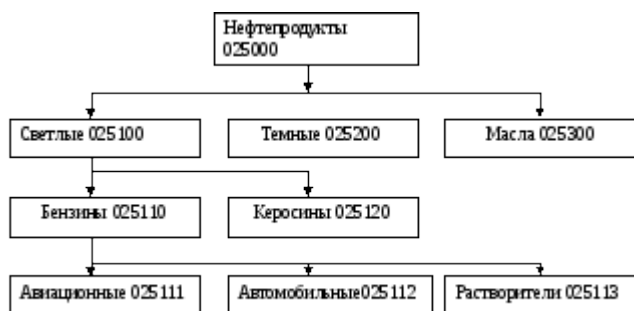


Рисунок 10 – Пример последовательного кодирования

Также при этом значение характеристик объекта, выраженного каким-либо количеством на определённой позиции, будет зависеть от конкретного значения в предыдущих разрядах кодовой комбинации (только за исключением разрядов для старшей группировки).

Последовательная система кодирования будет обладать теми же достоинствами и недостатками, что и иерархическая система. Если же сравнивать ее с порядковой системой, представление дополнительной информации будет достигаться за счёт

увеличения избыточности и длины кода.

2.5. Параллельное кодирование

Параллельное кодирование используют для фасетной системы в классификации. Суть метода также заключается в следующем: фасеты кодируются друг от друга независимо (параллельно, одновременно); в каждом фасете в структуре кода уже определяется позиция с одного или нескольких таких разрядов.

Параллельная система кодирования будет обладать основными достоинствами или недостатками, присущими для фасетной системы классификации. Но параллельный код оказывается более избыточным, чем рассмотренный последовательный, поскольку на практике многие сочетания таких признаков могут не существовать вообще и, следовательно, ёмкость не будет использоваться полностью (рисунок 11).

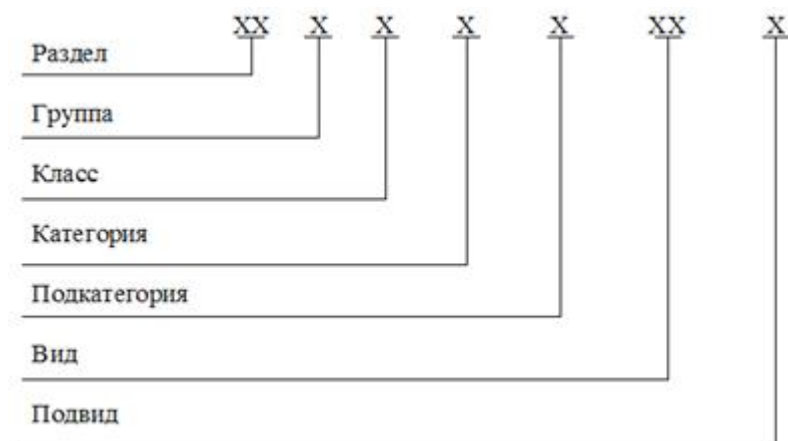


Рисунок 11 – Пример параллельного кодирования

Для использования при автоматизированной обработке и обмена информацией во всех установленных государством уровнях в Российской Федерации предназначен Общероссийский классификатор для специальностей по образованию. Но в целях примера (удобство восприятия и понимание) можно закодировать также каждое направление только первыми тремя буквами наименования. Для кодирования фасета с названием «Форма обучения» будем применять первую букву значения, для номера курса – десятичную цифру.

Структуру кода можно выражать записью «ННН Ф К», ткт ННН – присвоенный код направления; Ф – форма обучения, К – курс.

2.6. Штриховое кодирование

Последним в развитии систем кодирования считается штриховое кодирование.

В настоящее время при автоматической идентификации признаков широко используется оно, основанное на оптическом описании информации различными устройствами – сканером штрих-кода. Для примера можно привести широко известные коды, как:

- UPC – универсальный товарный код;
- EAN – европейский товарный код;

Опуская особенности преобразования для штрихового кода в цифровой, рассмотрим структуру широко распространённого в РФ товарного кода под названием EAN-13. Он состоит из тринадцати цифр, что можно разделить на 4 части:

- первые 2 цифры обозначаются код страны-производителя);
- следующие 5 цифр (в России они начинаются с 2000 г.) – код предприятия-производителя;
- следующие 5 цифр – код продукта, что устанавливается предприятием-производителем;
- последняя цифра – контрольная.

Таким образом, в основании кода EAN-13 лежит также последовательная система для кодирования, поскольку определение значения 2-й части кода возможно после определения первой, третьей – после второй (рисунок 12).



Рисунок 12 - Пример штрихового кодирования

Во втором разделе курсовой работы рассмотрены основные принципы кодирования с помощью специальных систем, приведены определения каждой системы кодирования.

Заключение

Рассмотрение данной темы курсовой работы многими учеными дало возможность возникнуть вычислительной технике, поскольку первоначально она возникла как средство для автоматизации вычислений.

Самым первым видом обрабатываемой информации для кодирования стала текстовая. Сначала разные тексты просто поясняли некоторые труднообозримые столбики цифр, а затем машины все более последовательным образом стали преобразовывать имеющуюся текстовую информацию.

Оформление разных текстов достаточно быстро вызывало у людей стремление как-то дополнить их графиками или рисунками. Делались попытки решить эти проблемы частично в рамках специального символического подхода: вводились также специальные символы для выполнения рисования таблиц и схем.

Но практические потребности разных людей в графике сделали ее появление среди основных видов компьютерной информации практически неизбежной. Числа, а также тексты и графика образовывали некоторый относительно набор, которого было также достаточно для решаемых многих на компьютере задач.

Постоянный рост быстродействия ПК создал широкие технические возможности по обработке звуковой информации, а также и для быстро сменяющихся графических изображений. Все это далее обусловило и развитие методов представления и кодирования самых различных видов информации по обработке ее на компьютере.

В процессе написания курсовой работы были реализованы следующие задачи:

- выполнена обработка литературы по теории кодирования;
- рассмотрены основные определения, а также классификацию методов кодирования;
- дана характеристика процессу кодирования информации
 - рассмотрены основные методы для выполнения кодирования;
 - описаны особенности развития систем кодирования.

В процессе анализа работы выявлены следующие достоинства для кодирования информации, а именно:

- удобность создания кода;
 - гибкость в использовании языков программирования;
- интегрированность в программные средства модулей для кодирования

Список использованной литературы

1. Баззел, Р.Д. Информация и риск в маркетинге / Р.Д. Баззел, Д.Ф. Кокс, Р.В. Браун. - М.: Финстатинформ, 2015. - 758 с.
2. Белоногов, Г.Г. Автоматизация процессов накопления, поиска и обобщения информации / Г.Г. Белоногов, А.П. Новоселов. - М.: Наука, 2017. - 256 с.
3. Берлекэмп, Э. Алгебраическая теория кодирования / Э. Берлекэмп. - М.: [не указано], 2017. - 575 с.
4. Берновский, Ю.Н. Классификация и кодирование промышленной и сельскохозяйственной продукции / Ю.Н. Берновский, В.А. Захаров, Р.А.

- Сергиевский, и др.. - М.: Стандартов, 2015. - 183 с.
5. Верещагин, Н. К. Информация, кодирование и предсказание / Н.К. Верещагин, Е.В. Щепин. - М.: ФМОП, МЦНМО, 2016. - 240 с.
 6. Воскобойников, Я.С. Журналист и информация. Профессиональный опыт западной прессы / Я.С. Воскобойников, В.К. Юрьев. - М.: РИА-Новости, 2015. - 208 с.
 7. Галицкий, А.В. Защита информации в сети - анализ технологий и синтез решений / А.В. Галицкий, С.Д. Рябко, В.Ф. Шаньгин. - М.: ДМК Пресс, 2013. - 615 с.
 8. Гоппа, В.Д. Введение в алгебраическую теорию информации / В.Д. Гоппа. - М.: [не указано], 2013. - 434 с.
 9. Гоппа, В.Д. Введение в алгебраическую теорию информации / В.Д. Гоппа. - М.: ФИЗМАТЛИТ, 2013. - 112 с.
 10. Информация президиума правления всероссийского театрального общества (январь-апрель 1980 г.). - М.: Всероссийское театральное общество, 2017. - 269 с.
 11. Кадомцев, Б.Б. Динамика и информация / Б.Б. Кадомцев. - М.: [не указано], 2016. - 856 с.
 12. Кельберт, М. Я. Вероятность и статистика в примерах и задачах. Том 3. Теория информации и кодирования / М.Я. Кельберт, Ю.М. Сухов. - М.: МЦНМО, 2014. - 568 с.
 13. Кельберт, М. Я. Вероятность и статистика в примерах и задачах. Том 3. Теория информации и кодирования / М.Я. Кельберт. - М.: МЦНМО, 2016. - 366 с.