

Лабораторная работа № 5

Кодирование Quoted-Printable

5.1. Цель работы:

Изучение алгоритма кодирования/декодирования Quoted-Printable.

5.2. Задание на лабораторную работу:

Разработать приложение, реализующее алгоритмы кодирования и декодирования Quoted-Printable.

5.3. Методические указания:

5.3.1. Механизм конвертации Quoted-Printable

Quoted-Printable – механизм конвертации, относящийся к спецификации MIME (Multipurpose Internet Mail Extensions). Текущая спецификация MIME описана в RFC 1521.

Данный механизм конвертации применим только к текстовым документам. Для конвертации исполняемых файлов, графических и видеофайлов лучше использовать механизм конвертации BASE64.

Quoted-Printable предполагает конвертацию всех символов в символы использующиеся в наборе символов ASCII. Данный набор представлен в таблице 5.1.

5.3.2. Правила механизма конвертации Quoted-Printable

Для данного механизма конвертации существует набор правил:

- 1) Все символы, кроме символов конца строки должны быть закодированы наборами, состоящими из знака “=” и двух шестнадцатеричных цифр. Данное правило является общим, поэтому следующие правила могут предложить альтернативу по кодированию различных символов.
- 2) Символы, номера которых в таблице ASCII находятся в диапазоне с 33 по 60 (!..<) и с 62 по 126 (>..~), не должны подвергаться кодированию.

- 3) Символы под номерами 9(ТАВ) и 32(Пробел) также не должны подвергаться кодированию, за исключением тех случаев, где данные символы находятся в конце сформированной с помощью механизма конвертации Quoted-Printable строки.
- 4) Для передачи конца строки используются символы 13(CR) 10(LF), так же, как и в обычном тексте. Данные символы соответствуют символам “\n\r”.
- 5) Максимальная длина одной строки не должна превышать 76 символов. Если строка имеет больший объём применяется так называемый “мягкий” перевод строки.

Все перечисленные правила позволяют найти некий компромисс между читаемостью закодированного текста и его объёмом.

Область применения механизма конвертации Quoted-Printable – текстовые сообщения, в которых не более 30% символов, не представленных в наборе символов ASCII. Такие текстовые сообщения чаще всего представляют собой web-страницу, и в последнее время довольно часто отправляются в виде рассылок, подтверждений и новостей.

Если количество символов, не представленных в наборе символов ASCII, больше 30%, то желательно использовать механизм конвертации BASE64, так как при этом происходит неоправданный рост объёма сообщения.

5.3.3. Алгоритм кодирования/декодирования Quoted-Printable

Для примера возьмём строку:

“name = Сергей “

5.3.3.1. Алгоритм кодирования

Начинаем побайтный анализ строки. Если байт соответствует символу принадлежащему набору символов ASCII, то копируем данный символ в нашу новую строку, в противном случае начинаем кодирование символа.

Кодирование каждого байта осуществляется путём конвертации данного байта в вид символа “=” и представлением данного байта в шестнадцатеричной форме.

Таблица 5.2

Пример кодирования строки

Символ	Код символа(bin)	Преобразование	Результат
n	01101110	Принадлежит ASCII, переносим	n
a	01100001	Принадлежит ASCII, переносим	a
m	01101101	Принадлежит ASCII, переносим	m
e	01100101	Принадлежит ASCII, переносим	e
пробел	00100000	Принадлежит ASCII, переносим	пробел
=	00111101	$00111101_2 = 3D_{16}$	=3D
пробел	00100000	Принадлежит ASCII, переносим	пробел
C	11010000	$11010000_2 = D0_{16}$	=D0
	10100001	$10100001_2 = A1_{16}$	=A1
e	11010000	$11010000_2 = D0_{16}$	=D0
	10110101	$10110101_2 = B5_{16}$	=B5
p	11010001	$11010001_2 = D1_{16}$	=D1
	10000000	$10000000_2 = 80_{16}$	=80
г	11010000	$11010000_2 = D0_{16}$	=D0
	10110011	$10110011_2 = B3_{16}$	=B3
e	11010000	$11010000_2 = D0_{16}$	=D0
	10110101	$10110101_2 = B5_{16}$	=B5
й	11010000	$11010000_2 = D0_{16}$	=D0
	10111001	$10111001_2 = B9_{16}$	=B9
пробел	00100000	$00100000_2 = 20_{16}$	=20
/n	00001101	Принадлежит ASCII, переносим	/n
/r	00001010	Принадлежит ASCII, переносим	/r

После кодирования наша строка изменилась:

“name =3D =D0=A1=D0=B5=D1=80=D0=B3=D0=B5=D0=B9=20”

Обратите внимание, что последний пробел тоже был закодирован, согласно правилу № 3.

Так же необходимо следить за количеством символов в строке, если строка будет превышать размер 76 символов, необходимо произвести “мягкий” перевод строки. Для этого нужно, при достижении строкой размера 73 символа, дописать символ “=” и “/n/r”.

5.3.3.2. Алгоритм декодирования

Начинаем побайтный анализ строки. Если байт не является символом “=”, то копируем данный символ в нашу новую строку, в противном случае необходимо посмотреть на следующие символы.

Если следующие два символа соответствуют шестнадцатеричным символам, необходимо произвести обратное преобразование. Для этого отбрасываем символ “=” и составляем байт из шестнадцатеричных символов. Записываем полученный байт в результирующую строку и переходим к анализу следующего байта, за рассматриваемыми символами.

В случае если следующие два символа представляют собой связку конца строки “/n/r”, то просто переходим к анализу следующего байта, за рассматриваемыми символами.

Таблица 5.3

Пример декодирования строки

Символ или группа символов	Преобразование	Код символа(bin)	Результат
n	Переносим	01101110	n
a	Переносим	01100001	a
m	Переносим	01101101	m
e	Переносим	01100101	e
пробел	Переносим	00100000	пробел
=3D	$3D_{16} = 00111101_2$	00111101	=
пробел	Переносим	00100000	пробел
=D0	$D0_{16} = 11010000_2$	11010000	
=A1	$A1_{16} = 10100001_2$	10100001	С
=D0	$D0_{16} = 11010000_2$	11010000	
=B5	$B5_{16} = 10110101_2$	10110101	e
=D1	$D1_{16} = 11010001_2$	11010001	
=80	$80_{16} = 10000000_2$	10000000	p
=D0	$D0_{16} = 11010000_2$	11010000	
=B3	$B3_{16} = 10110011_2$	10110011	г
=D0	$D0_{16} = 11010000_2$	11010000	
=B5	$B5_{16} = 10110101_2$	10110101	e
=D0	$D0_{16} = 11010000_2$	11010000	
=B9	$B9_{16} = 10111001_2$	10111001	й
=20	$20_{16} = 00100000_2$	00100000	пробел
/n	Переносим	00001101	/n
/r	Переносим	00001010	/r

В результате декодирования получен исходный текст.

5.4. Приложение: набор символов ASCII

Таблица 5.1

Набор символов ASCII

Decimal	Hex	ASCII	Decimal	Hex	ASCII	Decimal	Hex	ASCII	Decimal	Hex	ASCII
0	00	NUL	32	20	(blank)	64	40	@	96	60	`
1	01	SOH	33	21	!	65	41	A	97	61	a
2	02	STX	34	22	"	66	42	B	98	62	b
3	03	ETX	35	23	#	67	43	C	99	63	c
4	04	EOT	36	24	\$	68	44	D	100	64	d
5	05	ENQ	37	25	%	69	45	E	101	65	e
6	06	ACK	38	26	&	70	46	F	102	66	f
7	07	BEL	39	27	'	71	47	G	103	67	g
8	08	BS	40	28	(72	48	H	104	68	h
9	09	HT	41	29)	73	49	I	105	69	i
10	0A	LF	42	2A	*	74	4A	J	106	6A	j
11	0B	VT	43	2B	+	75	4B	K	107	6B	k
12	0C	FF	44	2C	,	76	4C	L	108	6C	l
13	0D	CR	45	2D	-	77	4D	M	109	6D	m
14	0E	SO	46	2E	.	78	4E	N	110	6E	n
15	0F	SI	47	2F	/	79	4F	O	111	6F	o
16	10	DLE	48	30	0	80	50	P	112	70	p
17	11	DC1	49	31	1	81	51	Q	113	71	q
18	12	DC2	50	32	2	82	52	R	114	72	r
19	13	DC3	51	33	3	83	53	S	115	73	s
20	14	DC4	52	34	4	84	54	T	116	74	t
21	15	NAK	53	35	5	85	55	U	117	75	u
22	16	SYN	54	36	6	86	56	V	118	76	v
23	17	ETB	55	37	7	87	57	W	119	77	w
24	18	CAN	56	38	8	88	58	X	120	78	x
25	19	EM	57	39	9	89	59	Y	121	79	y
26	1A	SUB	58	3A	:	90	5A	Z	122	7A	z
27	1B	ESC	59	3B	;	91	5B	[123	7B	{
28	1C	FS	60	3C	<	92	5C	\	124	7C	
29	1D	GS	61	3D	=	93	5D]	125	7D	}
30	1E	RS	62	3E	>	94	5E	^	126	7E	~
31	1F	US	63	3F	?	95	5F	_	127	7F	(delete)

5.5. Контрольные вопросы:

- 1) Что такое Quoted-Printable?
- 2) Назовите область применения механизма конвертации Quoted-Printable.
- 3) Назовите правила, по которым функционирует данный механизм.
- 4) Опишите алгоритм кодирования.
- 5) Опишите алгоритм декодирования.