

2주차 3차시 문자데이터

【학습목표】

1. 표준 ASCII 코드와 확장 ASCII 코드를 구분할 수 있다.
2. 한글코드의 문자를 표현하는 방법 두 가지를 설명할 수 있다.

학습내용1 : 표준 BCD코드

* BCD코드(Binary Coded Decimal Code)

- BCD코드는 10진수 각 자리의 숫자를 4비트(네 자리)의 2진수로 표현한 것
- 네 자리의 2진수를 큰 쪽부터 8의 자리, 4의 자리, 2의 자리, 1의 자리라고 하기 때문에 8421코드라고도 함
- 2진화 10진 코드, 8421 코드인 BCD코드는 10진수 0부터 9까지를 2진화한 코드로서 실제 표기는 2진수로 하지만 10진수처럼 사용함
- 1010부터 1111까지의 6개는 사용되지 않음

10진 수	BCD 코드	10진 수	BCD 코드	10진 수	BCD 코드
0	0000	10	0001 0000	20	0010 0000
1	0001	11	0001 0001	31	0011 0001
2	0010	12	0001 0010	42	0100 0010
3	0011	13	0001 0011	53	0101 0011
4	0100	14	0001 0100	64	0110 0100
5	0101	15	0001 0101	75	0111 0101
6	0110	16	0001 0110	86	1000 0110
7	0111	17	0001 0111	97	1001 0111
8	1000	18	0001 1000	196	0001 1001 0110
9	1001	19	0001 1001	237	0010 0011 0111

1. BCD코드의 연산

10진 덧셈 (6+3=9)

$$\begin{array}{r}
 0110 \\
 + 0011 \\
 \hline
 1001
 \end{array}$$

(42+27=69)

$$\begin{array}{r}
 0100\ 0010 \\
 + 0010\ 0111 \\
 \hline
 0110\ 1001
 \end{array}$$

- 1) 계산 결과가 BCD코드를 벗어나는 즉, 9를 초과하는 경우에는 계산 결과에 6(0110)을 더해줌
 - (8+7=15)

$$\begin{array}{r}
 1000 \\
 + 0111 \\
 \hline
 1111 \\
 + 0110 \\
 \hline
 0001\ 0101
 \end{array}$$

2. 3초과 코드

<BCD코드(8421코드)로 표현된 값에 3을 더해 준 값으로 나타내는 코드>

- 자기 보수의 성질이 있음

10진수	BCD 코드	3-초과 코드
0	0000	0011
1	0001	0100
2	0010	0101
3	0011	0110
4	0100	0111
5	0101	1000
6	0110	1001
7	0111	1010
8	1000	1011
9	1001	1100

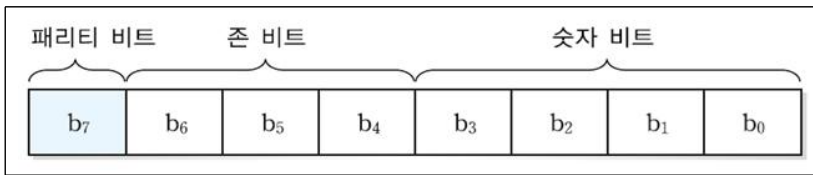
보수
관계

학습내용2 : ASCII 코드

<미국 국립 표준 연구소(ANSI)가 제정한 정보 교환용 미국 표준 코드(American Standard Code for Information Interchange)>

- 1) 코드의 길이는 7비트와 패리티 비트가 추가된 두 종류의 8비트 코드가 있으며, 128(= 2⁷)가지의 정보를 표현 할 수 있음
- ASCII 코드는 128개의 가능한 문자조합을 제공하는 7비트(bit) 부호로, 처음 32개의 부호는 인쇄와 전송 제어용으로 사용함
 - 보통 기억장치는 8비트(1바이트, 256조합)를 기본으로 구성되고, ASCII 코드는 단지 7비트의 128개의 문자만 사용하기 때문에 나머지 하나의 비트를 추가하여 패리티 비트로 사용하거나 특정문자를 표현하는데 사용함
 - 이렇게 하나의 비트가 추가되어 8비트의 코드로 특정문자까지도 표현할 수 있도록 만든 것을 확장 ASCII 코드라고 함

2) ASCII코드의 구성



1. 존 비트에 따른 ASCII코드의 분류

패리티 비트	존비트			숫자비트			
7	6	5	4	3	2	1	0
하위 비트에 따라 달라짐	1	0	0	영문자 A~O (0001~1111)			
	1	0	1	영문자 P~Z (0000~1010)			
	0	1	1	숫자 0~9 (0000~1001)			

2. 표준 ASCII 코드표

	0	1	2	3	4	5	6	7	8	9
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	TAB
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM
2		!	"	#	\$	%	&	'	()
3	0	1	2	3	4	5	6	7	8	9
4	@	A	B	C	D	E	F	G	H	I
5	P	Q	R	S	T	U	V	W	X	Y
6	`	a	b	c	d	e	f	g	h	i
7	p	q	r	s	t	u	v	w	x	y

	A	B	C	D	E	F
0	LF	VT	FF	CR	SO	SI
1	SUB	ESC	FS	GS	RS	US
2	*	+	,	-	.	/
3	:	;		=	>	?
4	J	K	L	M	N	O
5	Z	[W]	^	_
6	j	k	l	m	n	o
7	z	{		}	~	

3. 확장 ASCII 코드표

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	Ç	ü	é	â	ã	ä	å	ç	ê	ë	è	í	î	ï	Ä	Å
9	É	æ	Æ	ô	ö	ò	û	ü	ÿ	Ö	Ü	Ø	£	¥	Pl	f
A	á	í	ó	ú	ñ	Ñ	a	o	¿	¡	½	¼	¾	¼	¾	¾
B	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞
D	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞	⌞
E	α	β	Γ	π	Σ	σ	μ	τ	Φ	Θ	Ω	δ	∞	φ	ε	∩
F	≡	±	≥	≤	[]	÷	≈	°	•	•	√	°	²	•	

학습내용3 : 기타 코드들

1. 가중치코드(weighted code)

그 위치에 따라 정해진 값을 갖는 코드				
10진수	8421코드 (BCD)	2421 코드	5421 코드	84-2-1코드
0	0000	0000	0000	0000
1	0001	0001	0001	0111
2	0010	0010	0010	0110
3	0011	0011	0011	0101
4	0100	0100	0100	0100
5	0101	1011	1000	1011
6	0110	1100	1001	1010
7	0111	1101	1010	1001
8	1000	1110	1011	1000
9	1001	1111	1100	1111

10진수	51111 코드	바이쿼너리코드 (Biquinary Code) 5043210	링 카운터 (ring counter) 9876543210
0	00000	0100001	0000000001
1	00001	0100010	0000000010
2	00011	0100100	0000000100
3	00111	0101000	0000001000
4	01111	0110000	0000010000
5	10000	1000001	0000100000
6	11000	1000010	0001000000
7	11100	1000100	0010000000
8	11110	1001000	0100000000
9	11111	1010000	1000000000

2. 비가중치코드(non-weighted code)

- 1) 각각의 위치에 해당하는 값이 없는 코드
- 2) 데이터 변환과 같은 특수한 용도로 사용되기 위한 코드 (2-out-of-5)

10진수	3-초과코드	5중 2코드 (2-out-of-5)	shift counter	그레이코드
0	0011	11000	00000	0000
1	0100	00011	00001	0001
2	0101	00101	00011	0011
3	0110	00110	00111	0010
4	0111	01001	01111	0110
5	1000	01010	11111	0111
6	1001	01100	11110	0101
7	1010	10001	11100	0100
8	1011	10010	11000	1100
9	1100	10100	10000	1101

3. 그레이 코드(Gray Code)

- 1) 가중치가 없는 코드이기 때문에 연산에는 부적당하지만, 아날로그-디지털 변환기나 입출력 장치 코드로 주로 쓰임
- 2) 연속되는 코드들 간에 하나의 비트만 변화하여 새로운 코드가 됨

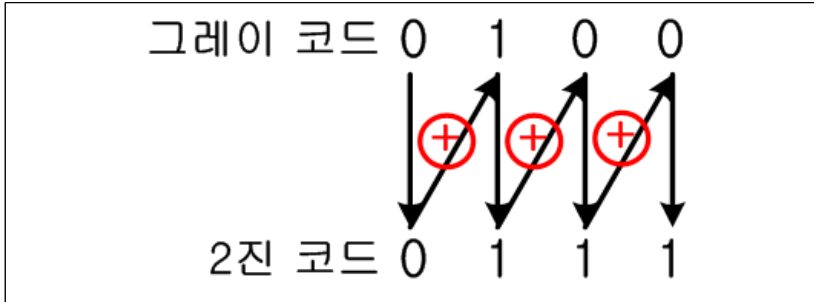
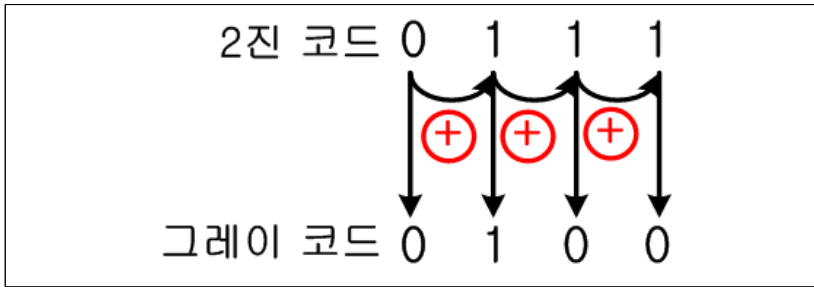
10 진수	2진 코드	Gray 코드	10 진수	2진 코드	Gray 코드
0	0000	0000	8	1000	1100
1	0001	0001	9	1001	1101
2	0010	0011	10	1010	1111
3	0011	0010	11	1011	1110
4	0100	0110	12	1100	1010
5	0101	0111	13	1101	1011
6	0110	0101	14	1110	1001
7	0111	0100	15	1111	1000

↩
이웃하는
코드간에
한 비트만
다름

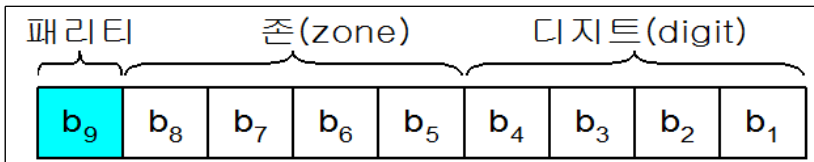
4. 2진 코드를 그레이 코드로 변환하는 방법

5. 그레이 코드를 2진 코드로 변환하는 방법

6. EBCDIC 코드



- Extended Binary Coded Decimal Interchange Code
- 대형 컴퓨터와 IBM 계열 컴퓨터에서 많이 사용되고 있는 8비트 코드(IBM에서 개발)
- 256종류의 문자 코드를 표현할 수 있는 영숫자 코드



7. EBCDIC 코드표

16진	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2진	0	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	1110	1111
0	0	NUL	SOH	STX	ETX	HT		DEL				VT	FF	CR	SO	SI
1	1	DLE					BS		CAN	EM			IFS	IGS	IRS	IUS
2	10					LF	ETB	ESC						ENQ	ACK	BEL
3	11			SYN			EOT							NAK		SUB
4	100	spac														
5	101	e									[.		(+	!
6	110	&									!]	\$	*)		^
7	111	-	/									,	%	_	>	?
8	1000										:	#	@	.	=	"
9	1001		a	b	c	d	e	f	g	h	i					
A	1010		j	k	l	m	n	o	p	q	r					
B	1011		~	s	t	u	v	w	x	y	z					
C	1100	{	A	B	C	D	E	F	G	H	I					
D	1101	}	J	K	L	M	N	O	P	Q	R					
E	1110	W		S	T	U	V	W	X	Y	Z					
F	1111	0	1	2	3	4	5	6	7	8	9					

8. 유니코드(Unicode)

<ASCII 코드의 한계성을 극복하기 위하여 개발된 인터넷 시대의 표준 유니코드 컨소시엄(IBM, Novell, Microsoft, DEC, Apple 등)에 의해서 32(UTF-32), 16(UTF-16), 8bit(UTF-8)의 세 가지 기본 코드로 현재 4.0까지 개발>

- 미국, 유럽, 동아시아, 아프리카, 아시아 태평양 지역 등의 주요 언어들에 적용될 수 있음
- 유니코드는 유럽, 중동, 아시아 등 거의 대부분의 문자를 포함하고 있으며, 96,382개의 문자로 구성되어 있음
- 특히 아시아의 중국, 일본, 한국, 타이완, 베트남, 싱가포르에서 사용하는 표의 문자(한자) 70,207개를 나타낼 수 있음

- 구두표시, 수학기호, 전문기호, 기하학적 모양, 덩벙 기호 등을 포함
- 앞으로도 계속해서 산업계의 요구나 새로운 문자들을 추가하여 나갈 것임

9. 한글코드

<한글은 ASCII코드를 기반으로 16비트를 사용하여 하나의 문자를 표현>



① 조합형

- 조합형으로 표현된 한글은 때에 따라서 다른 응용프로그램에서는 사용할 수 없는 문자들이 많음
- 조합형은 자음과 모음으로 조합 가능한 모든 한글을 사용할 수 있음
- 심지어 우리나라 고어(古語)까지 취급할 수 있는 장점이 있음
- 출력 시 다시 모아 써야 하는 불편이 있다는 것이 단점임

② 완성형

* 완성형 한글코드

- 1987년 정부가 한국표준으로 정한 것으로 가장 많이 사용되는 한글 음절을 2 바이트의 2 진수와 1 대 1로 대응하여 표현하는 방법임

【학습정리】

1. 문자 코드는 대표적인 두가지 BCD와 ASCII 코드가 있다.
2. BCD와 ASCII코드의 명확히 차이가 있다.