情報のコード化とCOMET IIでの文字の取り扱い

山本昌志*

2005年11月11日

1　前回の復習と本日の内容

1.1　前回の復習

前回の講義内容は以下の通りで，教科書のP.9〜P.12に対応する．課題ができていれば，十分である．

- COMET IIのメモリー
- 負の数の表現方法 (2の補数) を示した．

1.2　本日の内容

本日の講義内容は以下の通りで，教科書のP.12〜14に対応する．

- 情報のコード化
- COMET IIの文字の取り扱い方
- これまでのまとめ

1.3　情報のコード化とメモリーへ格納

一般に，情報を記号によって表現することをコード化 (符号化) と言う．表現されたものをコード (符号) と呼ぶ．情報を表す記号は，何を使っても良いが，整数を使うのが最も簡単であるし，コンピューターとの相性も良い．

コンピューターは情報 (データ) を加工する機械である．コンピューターで取り扱う情報は，全てコード化されて，整数で表現される．数値あろうが文字であろうが，音や絵さえも整数で表されるのである．整数で表すことができたら，2進数での取り扱いが可能となり，論理回路で処理できるようになる．データを整数で表すことは，コンピューターを用いての情報処理の第一歩となる．

*国立秋田工業高等専門学校　電気工学科

整数にコード化された情報を処理するためには，その整数を記憶し，計算する必要がある．記憶する役目を担っている装置がメモリーで，計算する役目を担っている装置が CPU である．ここでは，情報を整数にコード化して，それがメモリーに格納される様子を学習する．

整数をコード化して，コンピューター (COMET II) のメモリーに格納する方法は，先週までに学習した．例えば，図 1 のようにである．これを見れば，整数がコード化されて，それがメモリーに格納されることが分かるだろう．

- 符号無し整数の場合，整数そのものがコードになる．
- 符号付き整数の場合，2 の補数にしたものがコードになる．

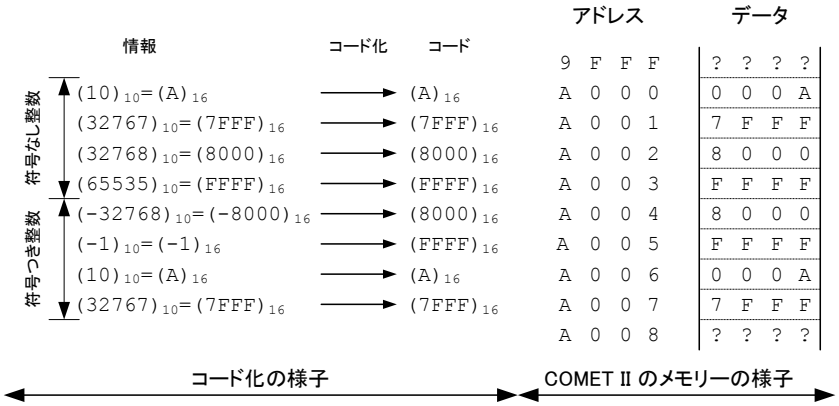

図 1: 整数のコード化の例．コード化された整数が，メモリーに格納されている．データの’????’は不明を示す．

この図を見て分かるように，COMET II のメモリーの中では，符号無し整数の $(65535)_{10}$ と符号付き整数の $(-1)_{10}$ は，メモリーの内容は全く同じである．したがって，COMET II では，それら 2 つの値は全く区別できず，同じ値となってしまうのである．これらの情報にはとても大きな違いがあり，これでは使い物にならない．コンピューターそのものは，アホでこれらの区別はできないが，実際のところプログラムがこれらを区別しているのである．これらの値を処理するプログラムでは，符号の有無に区別して計算するのである．もちろん，この区別はプログラマー (人間) が決めなくてはならない．アホなプログラマーがアホなプログラムを書くと，アホなコンピューターはアホな処理しかしないのである．

2 COMET II の文字の取り扱い

2.1 文字の取り扱い方

COMET II で文字列をコード化するためには，JIS X0201 ラテン文字・片仮名用 8 単位符号を用いる (付録，教科書 p.13 の表 2.3)．表を見て分かるように，この符号は各文字が 8 ビットの整数で表現されているだけである．たとえば，

文字	2 進数	16 進数
!	00100001	21
+	00101011	2B
3	00110011	33
K	01001011	4B
t	01110100	74

のようになっている．次の問いに答えよ．

- このコードのビット数?．それは，何バイト?．
- このビット数で表現できる文字数は，いくつか?
- 漢字やカタカナ，ひらがながある日本語の文字はこのビット数で表現できるか?

このコードでは，文字を 8 ビットで表すが，COMET II は 16 ビット単位でデータを取り扱う．一つのアドレスあたり 2 文字，格納することも考えられるが，そうすると 1 文字の取り扱いに困難をきたす．実際，COMET II では，上位 8 ビットを 0 にして，下位 8 ビットを使うと決められている．それについては，教科書 P.211 の真ん中あたりの文字定数の部分に書いてある．

文字がコード化されて，COMET II のメモリーに格納される様子を図 2 に示す．メモリーのアドレス `#A000` から，`'Akita kosen'` と文字が格納される様子が分かるだろう．

- 図 2 で，上位 8 ビットが 0 になっていることを確認せよ．

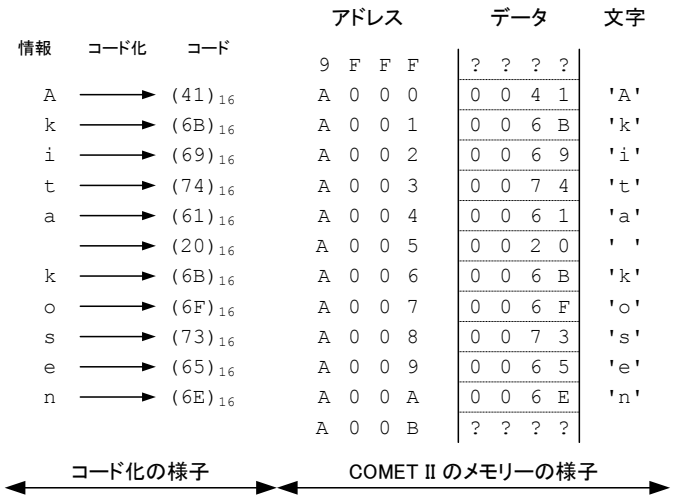

図 2: 文字列のコード化の例．文字列`'Akita kosen'`がコード化され，メモリーに格納されている．データの'????' は不明を示す．

2.2 文字と数値の違い

教科書に書いてある通り (P.14).

3 これまでのまとめ

コンピューターを構成する最も重要な要素は,

- Central Processing Unit (CPU:中央処理装置)
- メインメモリー (main memory:主記憶装置). 単にメモリーと呼ぶことも多い.

である. これまでは, メインメモリーの中でのデータ (数値, 文字) の格納方法を学習した. 次のようなことを理解していないと, 次からわからなくなる.

- COMET II では, 16 ビットを 1 ワード (1 語) と言い, この単位でデータの処理を行う.
- メモリーには, 1 ワード毎にアドレスが割り振られている.
- COMET II では, 整数は 16 ビットで表現する. 符号付整数は次のようにして表す (コード化する).
  - 正の整数はそのまま, 16 桁の 2 進数で (16 ビット) で表す.
  - 負の整数は, 16 ビットの 2 の補数で表す. 2 の補数への変換方法は以下の通りである.
    1. 絶対値を 2 進数で表して, ビット反転する.
    2. ビット反転した値に+1 加算する.
- 符号付き整数が表すことができる範囲は, 以下の通りである.
  - 正の数の絶対値の最大値は, $(0111\ 1111\ 1111\ 1111)_2 = (2^{15}\text{-}1)_{10}=(32767)_{10}$
  - 負の数の絶対値の最大値は, $(1000\ 0000\ 0000\ 0000)_2$ である. これは第 15 ビットが 1 なので負の数で, 2 の補数表示となっている. したがって, その絶対値を求めるためには, ビット反転を行い, 1 を加算すればよい. すると, これは$\text{-}(2^{15})_{10}=\text{-}(32768)_{10}$ を表すことが分かる.
- 符号無し整数の場合は, 以下の通りである.
  - 表現可能な最小値は, $(0000\ 0000\ 0000\ 0000)_2 = (0)_{10}$ である.
  - 表現可能な最大値は, $(1111\ 1111\ 1111\ 1111)_2 = (2^{16}\text{-}1)_{10}=(65535)_{10}$ である.
- 数値と異なり, 文字にはそれぞれ, 番号をつけて区別 (コード化) する. COMET II の文字のコード化は, 規格 JIS X0201 ラテン文字・片仮名用 8 単位符号をつかう.
- このコードは, 8 ビットなので, 最大 256 文字しか使えない. 数字とアルファベットと片仮名と記号を表すのであれば十分である. 漢字は, 使えない.

- COMET II の 1 ワード 16 ビットに対して，文字は 8 ビットしかつかわない．COMET II では 1 ワードで 1 文字を表すため，16 ビットのうち上位 8 ビットは 0 として，下位 8 ビットで 1 文字分を表す．例えば，アルファベットの Yama を表す場合，Y は $(59)_{16}$，a は $(61)_{16}$，m は $(6D)_{16}$，とコード化されるので，COMET のメモリーには，次のように格納される．ただし，アドレスの実際の割り当ては，OS が決める．

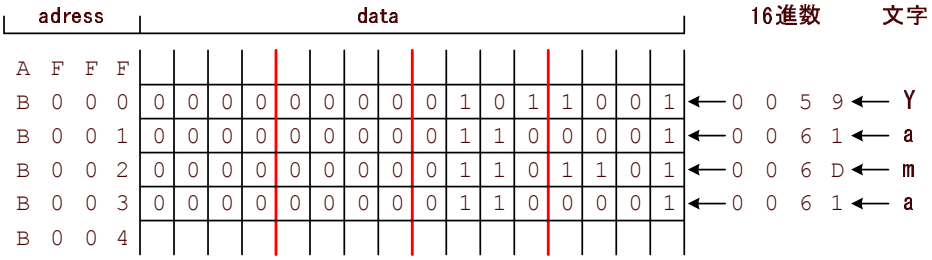

図 3: 文字列”Yama”のメモリーへの格納

- 数値と文字では，メモリーの中身は異なる．例えば，数値の $(9)_{10}$ と文字の”9”は，以下のようになる．文字の”9”は，JIS X0201 では，$(39)_{16}$ である．

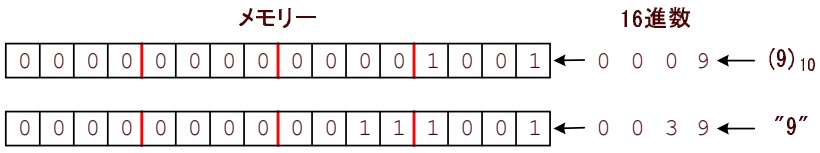

図 4: 数値の $(9)_{10}$ と文字”9”のメモリーへの格納

- メモリーの中身を見ると，それが数値なのか文字なのか，判断できない．命令毎に数値を扱うのか，文字を扱うのか決まっている．これは，以降の学習で分かる．

4 課題 (レポート)

以下のデータを格納せよ．ただし，データは，アドレスの B001 から格納される者とする．

[**問 1**] 文字列の”Keisanki”

[**問 2**] 学籍番号 (整数 2 桁) と自分の名字 (ローマ字)．ただし，名字のアルファベットは大文字とする．

[**問 3**] 文字列の”15AB”と 16 進整数の $(15AB)_{16}$

アドレス (16 進数)	データ (2 進数)																データ (16 進数)
AFFF	1	0	0	1	1	1	1	1	0	1	0	1	1	0	1	0	9F5A
B000																	
B001																	
B002																	
B003																	
B004																	
B005																	
B006																	
B007																	
B008																	
B009																	
B00A																	

図 5: メモリー

4.1 レポート提出要領

提出方法は，次の通りとする．

期限　11 月 18 日 (金)PM1:00 まで
用紙　A4
提出場所　山本研究室の入口のポスト
表紙　表紙を 1 枚つけて，以下の項目を分かりやすく記述すること．
　授業科目名「電子計算機」
　課題名「課題 4　メモリー中の文字の表現」
　3E　学籍番号　氏名
　提出日
内容　問題の解答．

5 付録

5.1 JIS X0201

表 1: JIS X0201 コード表．A1〜A5 の記号とカタカナは半角文字である．

下位 4 ビット	上位 4 ビット															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DEL	間隔	0	@	P	‘	p				-	タ	ミ		
1	SOH	DC1	!	1	A	Q	a	q			。	ア	チ	ム		
2	STX	DC2	"	2	B	R	b	r			「	イ	ツ	メ		
3	ETX	DC3	#	3	C	S	c	s			」	ウ	テ	モ		
4	EOT	DC4	$	4	D	T	d	t			、	エ	ト	ヤ		
5	ENQ	NAK	%	5	E	U	e	u			・	オ	ナ	ユ		
6	ACK	SYN	&	6	F	V	f	v			ヲ	カ	ニ	ヨ		
7	BEL	ETB	’	7	G	W	g	w			ァ	キ	ヌ	ラ		
8	BS	CAN	(	8	H	X	h	x			ィ	ク	ネ	リ		
9	HT	EM	)	9	I	Y	i	y			ゥ	ケ	ノ	ル		
A	LF	SUM	*	:	J	Z	j	z			ェ	コ	ハ	レ		
B	VT	ESC	+	;	K	[	k	{			ォ	サ	ヒ	ロ		
C	FF	FS	,	<	L	\	l	|			ャ	シ	フ	ワ		
D	CR	GS	-	=	M	]	m	}			ュ	ス	ヘ	ン		
E	SO	RS	.	>	N	^	n	~			ョ	セ	ホ	゛		
F	SI	US	/	?	O	_	o	DEL			ッ	ソ	マ	゜		