入学年度		学部	学 科		組		番　号			検	フリガナ	
		B		1							氏 名	

QR コードで漢字をコード化する仕組みを詳しく知るために，自分の名前を QR コードにしてみよう．

- **型**： 東アジア漢字文化圏の大抵の名前は漢字 7 文字以内なので，誤り訂正レベルが高めの「レベル Q」を選択しても 1 型に収まる．漢字 4 文字以下ならさらに訂正能力の高い 1-H 型を用いることが可能である．今年度の履修者の名前はすべて漢字 4 文字以下なので，1-H 型を用いることにする．
- **モード指示子**： 今回は QR コードで用いられる種々のモードのうち「漢字モード」を用いる．これを指示する最初の 4 bit は 1000 である．
- **文字数指示子**： つぎに文字数指示子は，名前の字数を 2 進法で表示する．漢字モードの場合，これを 8 bit で表す（英数字モードの場合は 9 bit）．例えば「草野美登莉」のように 5 文字なら 00000101 となる．
- **データの bit 列化**： そして，いよいよ実際の漢字データを bit 列になおす．欧文に使われる文字は様々なアクセント記号を含めても $2^8 = 256$ 以内に収まるので，1 byte= 8 bit で表現できるが，漢字などの非欧文文字は 2 byte を用いないと表せない．このように様々な文字を 2 byte の数で表現する仕組みのことを「文字コード体系」と呼ぶ．QR コードでは漢字を表す際には日本独自の体系である “Shift_JIS コード” と呼ばれる文字コードが用いられる．今年度の履修者の名前に使われる漢字も，少々の字体の違いはあるものの Shift_JIS コードで表せるので，その仕組みを理解するために Shift_JIS コードを用いることにする．実用的には “UTF-8” を用いるのが適当であるが，Shift_JIS コードの方がデータがコンパクトになるメリットがある．

Shift_JIS コードの概略

0 と 1 のみを扱うコンピュータ内部では，数は 10 進法ではなく 2 進法で表現した方が都合がよい．しかし，プログラミングの際に数を 2 進法で表すと桁数が大きくなりすぎて人間にとっては理解しづらくなってしまう．そこで，プログラミングなどの際には 16 進法を用いて数を表すことが一般的である．1 byte $= 2^8 = 16^2$ なので，16 進法で 2 桁までの数を 1 byte で表すことができ，2 byte で 16 進法で 4 桁までの数を表現できる．

一方，2 bye で表せる文字数は $2^{16} = 65{,}535$ 字であるが，常用漢字は約 2,000 字程度であり，さらに使用頻度の低い JIS 第 4 水準漢字までをすべて含めても約 11,000 字程度であって，かなりの空き領域ができる．Shift_JIS コードは JIS 第 2 水準漢字までの漢字など 6,879 文字が含まれるが，これは $2^{13} = 8{,}192$ より少ないので，すべて表すには 2 byte = 16 bit も必要はなく，13 bit あれば十分である．そこで，QR コードでは無駄をなくすために，漢字 1 文字を 16 bit ではなく 13 bit の 2 進数で表示することにより，情報を圧縮して格納している．

Mathematica で “草” の Shift_JIS コードを求めるには “`ToCharacterCode`” というコマンドを用いるが，`ToCharacterCode["草","ShiftJIS"]` とすると，結果が 10 進法で表示されるので，16 進法で表示するために次のようにする．

```
BaseForm[ToCharacterCode["草","ShiftJIS"],16]  ⇧+↵
```

$\{91_{16}, 90_{16}\}$

Shift_JIS コードの漢字は第 1 byte を縦軸に、第 2 byte を横軸にとった仮想の正方形の中で，2 つの長方形の領域内に含まれている．第 1 の長方形は第 1 byte が 81_{16} から $9F_{16}$，第 2 byte が 40_{16} から FC_{16}，第 2 の長方形は第 1 byte が $E0_{16}$ から EB_{16}，第 2 byte が 40_{16} から FC_{16} の長方形である．

		0 番	1 番	2 番	…	188 番
	第 1 byte ＼ 第 2 byte	40_{16}	41_{16}	42_{16}	…	FC_{16}
0 番	81_{16}		、	。	…	○
⋮	⋮	⋮				
30 番	$9F_{16}$	槃	槧	檻	…	滌
	⋮					
31 番	$E0_{16}$	漾	漓	滷	…	珱
⋮	⋮	⋮				
42 番	EB_{16}	*	*	*	…	*

圧縮の原理は以下の通り．第 1 byte は 81_{16} を 0 番とし，$9F_{16}$ までを連番で表すと，$9F_{16} - 81_{16} = 1E_{16} = 30$ なので，$9F_{16}$ は 30 番となる．漢字コードが $E040_{16}$ を超える場合，$E0_{16}$ から $C1_{16}$ を引くと，$E0_{16} - C1_{16} = 1F_{16} = 31$ となるので，$E0_{16}$ が 31 番，$E1_{16}$ が 32 番，…，EB_{16} が 42 番となる．一方，第 2 byte は 40_{16} を 0 番とすると，$FC_{16} - 40_{16} = BC_{16} = 188$ より，FC_{16}は188 番となる．そこで，漢字コードを次のようにして「192 進法化」する．すなわち，

（第 1 byte の通し番号（0〜43））× 192 ＋（第 2 byte の通し番号（0〜188））

を漢字コードの番号とする．この計算はすべて 16 進法のまま行う．例えば，“草” なら “9190_{16}” から “8140_{16}” を引くと “1050_{16}” となるが，この最初の 1byte に $C0_{16} = 192$ を掛け，つぎの byte を足す．すなわち，

$$10_{16} \times C0_{16} + 50_{16} = 0C50_{16}$$

こうして “草” は次のように変換される．

$$\begin{aligned} \text{“草”} &\Rightarrow 9190_{16} \\ &\Rightarrow 9190_{16} - 8140_{16} = 1050_{16} \\ &\Rightarrow 10_{16} \times C0_{16} + 50_{16} = 0C50_{16} \\ &\Rightarrow 0110001010000_2 \end{aligned}$$

では，“草”，“野”，“美”，“登”，“莉” を「192 進法化」してみよう．まず，漢字を対応する Shift_JIS 漢字コードに直と，それぞれ，“9190_{16}”，“$96EC_{16}$”，“$94FC_{16}$”，“$936F_{16}$”，“$E4BB_{16}$” となる．

【裏に続く】

氏　名	Shift_JIS コード		差	「192 進法化」	13bit 化
草	9190_{16}	$-$ 8140_{16}	$= 1050_{16}$	$0C50_{16}$	0 1 1 0 0 0 1 0 1 0 0 0 0
野	$96EC_{16}$	$-$ 8140_{16}	$= 15AC_{16}$	$106C_{16}$	1 0 0 0 0 0 1 1 0 1 1 0 0
美	$94FC_{16}$	$-$ 8140_{16}	$= 13BC_{16}$	$0EFC_{16}$	0 1 1 1 0 1 1 1 1 1 1 0 0
登	$936F_{16}$	$-$ 8140_{16}	$= 122F_{16}$	$0DAF_{16}$	0 1 1 0 1 1 0 1 0 1 1 1 1
莉	$E4BB_{16}$	$-$ $C140_{16}$	$= 237B_{16}$	$1ABB_{16}$	1 1 0 1 0 1 0 1 1 1 0 1 1

これを自分の名前でやってみよう．

氏　名	Shift_JIS コード		差	「192 進法化」	13bit 化

すべてを bit 化したら，モード指示子，文字数，漢字データを順に並べ，最後に終端パターンとして 0000 を付加する．

$$\underbrace{1000}_{\text{漢字モード}}\ \underbrace{00000100}_{\text{文字数 4}}\underbrace{0110001010000}_{\text{“草”}}\underbrace{1000001101100}_{\text{“野”}}\ \cdots\ \underbrace{0000}_{\text{終端パターン}}$$

こうして得られたデータを 8bit ごと（1byte ごと）に区切り直す．最後のビット列が 8bit 未満の場合は 0 で埋める．また，1-Q 型では RS(26, 13, 6) 符号を用いるので，得られた byte 数が情報 byte 数である 13 に満たない場合は "11101100" および "00010001" という「埋め草パターン」を交互に付加しする．

漢字 3 文字の場合には下の表のようになる．

漢字モード	文字数	（第一文字）	（第二文字）
1 0 0 0	0 0 0 0 0 0 1 1		

	（第三文字）	終端パターン	0 fill

埋め草パターン 1	埋め草パターン 2	埋め草パターン 1	埋め草パターン 2

埋め草パターン 1

漢字 3 文字の場合には下の表のようになる．

漢字モード	文字数	（第一文字）	（第二文字）
1 0 0 0	0 0 0 0 0 1 0 0		

	（第三文字）	（第四文字）

終端パターン	0 fill	埋め草パターン 1	埋め草パターン 2	埋め草パターン 1
0 0 0 0	0 0 0 0	1 1 1 0 1 1 0 0	0 0 0 1 0 0 0 1	1 1 1 0 1 1 0 0

埋め草パターン 2
0 0 0 1 0 0 0 1

このようにして，学籍番号の場合と同様，13byte からなる情報語を得る．

	8bit データ							
1.	1	0	0	0	0	0	0	0
2.	0							
3.								
4.								
5.								
6.								
7.								
8.								
9.								
10.								
11.								
12.								
13.								

担当：鍬田 政人 (kuwata.04f@g.chuo-u.ac.jp)