

入学年度	学部	学科	組	番号	検	フリガナ
						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 日本人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。そこで、1-Q型を選ぶことにする。(漢字4文字ならさらに訂正能力の高い1-H型でも可能。)
- 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4bit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8bitで表す(英数字モードの場合は9bit)。例えば「草野みどり」のように5文字なら0000101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、1byte=8bitで表現できるが、漢字などの非欧文文字は2byteを用いないと表せない。逆に2byteで表せる文字数は $2^{16} = 65535$ であるが、常用漢字は約2000字程度であり、これより使用頻度の低いJIS第2水準漢字を含めても約6000字程度である。QRコードで用いられるShift_JISコードはJIS第2水準漢字までの漢字が扱え、全体の文字数は $2^{13} = 8192$ より少ない。そこで、QRコードでは、漢字1文字を2byte=16bit用いて表すのではなく、13bitに圧縮して格納する。

圧縮の原理は以下の通り。まず、漢字を対応するShift_JIS漢字コードに直す。Shift_JIS漢字コードは通常2byteを16進法で4桁の数にして表す。例えば、「草」、「野」、「み」、「ど」、「り」はそれぞれ、「9190₁₆」、「96EC₁₆」、「82DD₁₆」、「82C7₁₆」、「82E8₁₆」となる。

Shift_JISコードは第1byteを縦軸に、第2byte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1byteが81₁₆から9F₁₆、第2byteが40₁₆からFC₁₆、第2の長方形は第1byteがE0₁₆からEB₁₆、第2byteが40₁₆からFC₁₆の長方形である。JIS第1水準の漢字はすべて第1の長方形に入るので、大抵の人名はこの中に収まるはずである。そこで、Shift_JIS漢字コードの原点を「8140₁₆」にずらすために、各漢字コードから「8140₁₆」を引き算する。万一、漢字コードがE040₁₆を超える場合には8140₁₆の代わりにC140₁₆を引き算する。なお、4桁の引き算ではあるが、理論上桁の繰り下がりはないので、桁ごとに引き算をすればよい。

さらに、得られた結果を「192進法化」する。つまり、最初の1byteにC0₁₆ = 192を掛け、つぎのbyteを足す。例えば、1050₁₆なら、10₁₆ × C0₁₆ = 0C00₁₆となる。最後にこれを13bitの2進法で表す。

「草」 → 9190₁₆ → 9190₁₆ - 8140₁₆ = 1050₁₆ → 10₁₆ × C0₁₆ + 50₁₆ = 0C50₁₆ → 0110001010000₂

氏名	草	野	み	ど	り
Shift_JIS漢字コード	9190 ₁₆	96EC ₁₆	82DD ₁₆	82C7 ₁₆	82E8 ₁₆
—	8140 ₁₆				
差	1050 ₁₆	15AC ₁₆	019D ₁₆	0187 ₁₆	01A8 ₁₆
「192進法化」	0C50 ₁₆	106C ₁₆	015D ₁₆	0147 ₁₆	0168 ₁₆
13bit化	0110001010000	1000001101100	0000101011101	0000101000111	0000101101000

なお、Mathematicaで計算する場合、「草」のShift_JISコードを求めるにはToCharacterCode["草",16]とするが、結果は{91₁₆, 90₁₆}と1byteごとに表示される。そこで、Shift_JISコードを16進法で表すのはやめ、81₁₆ = 129, 40₁₆ = 64であることを用い、

$$\begin{aligned} & \text{ToCharacterCode["草"]} - \{129, 64\} \text{ (Shift_Enter) } \{16, 80\}, \\ & 16 \times 192 + 80 \text{ (Shift_Enter) } 3152 \\ & \text{BaseForm}[3152, 2] \text{ (Shift_Enter) } 110001010000_2 \end{aligned}$$

と計算すればよい。

氏名				
Shift_JIS漢字コード				
—				
差				
「192進法化」				
13bit化				

すべてをbit化したら、最後に終端パターンとして0000を付加する。

$$\underbrace{1000}_{\text{漢字モード}} \underbrace{0000100}_{\text{文字数4}} \underbrace{0110001010000}_{\text{「草」}} \underbrace{1000001101100}_{\text{「野」}} \dots \underbrace{0000}_{\text{終端パターン}}$$

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」を交互に付加する。漢字4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)
1000	0000	0100	
		(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
0000	0000	11101100	00010001
			埋め草パターン1
			11101100
		埋め草パターン2	
		00010001	

漢字3文字の場合には裏の表のようになる。

