

基礎プログラミングI

第9回 計算機の内部表現

今回の内容

- 2進数と16進数
- 文字コード
- 文字列のエンコーディング

今回の内容

- 2進数と16進数
- 文字コード
- 文字列のエンコーディング

2進数

計算機の内部では、データは全て1と0の**2進数**の状態で作保存されている。

なぜ2進数なのか？

データを一時的に記憶するメモリに書き込むために電氣的にONとOFF (または磁氣的にNとS)の形にするため。

データ格納領域の例

| | | | | | | | |
|-----|----|-----|-----|-----|-----|-----|----|
| OFF | ON | OFF | OFF | OFF | OFF | OFF | ON |
|-----|----|-----|-----|-----|-----|-----|----|

2進数

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
|---|---|---|---|---|---|---|---|

10進数

| 128倍 | 64倍 | 32倍 | 16倍 | 8倍 | 4倍 | 2倍 | 1倍 |
|------|-----|-----|-----|----|----|----|----|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

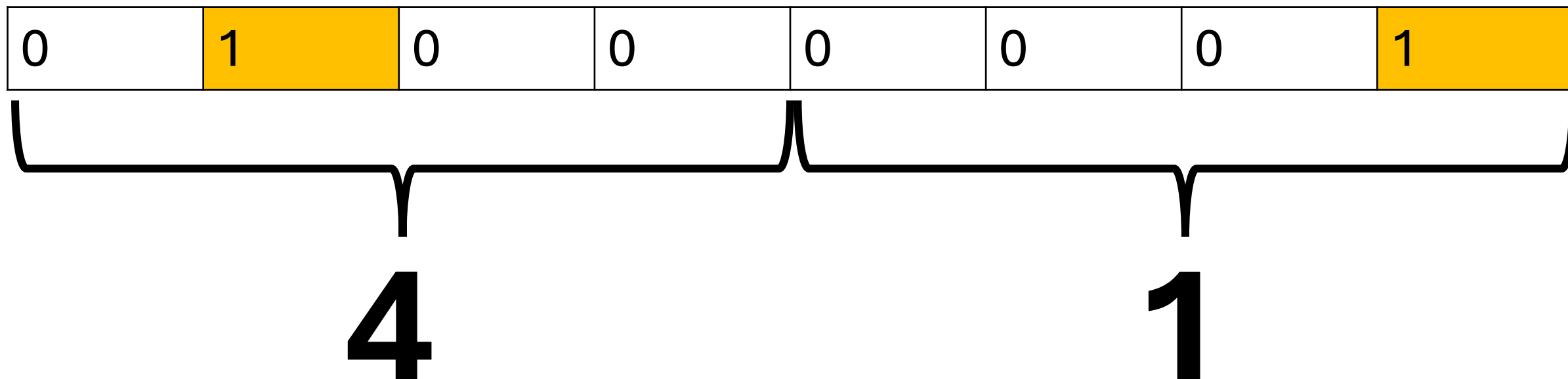
$$64 \times 1 + 1 \times 1 = 65$$

16進数

2進数4桁をまとめて**16進数**にして数値を表記することが多い。

| 2進数 | 16進数 | 10進数 |
|------|------|------|
| 0000 | 0 | 0 |
| 0001 | 1 | 1 |
| 0010 | 2 | 2 |
| 0011 | 3 | 3 |
| 0100 | 4 | 4 |
| 0101 | 5 | 5 |
| 0110 | 6 | 6 |
| 0111 | 7 | 7 |
| 1000 | 8 | 8 |
| 1001 | 9 | 9 |
| 1010 | a | 10 |
| 1011 | b | 11 |
| 1100 | c | 12 |
| 1101 | d | 13 |
| 1110 | e | 14 |
| 1111 | f | 15 |

2進数から16進数への変換



16進数と10進数の変換

16進数から10進数へ

16進数を10進数に変換するには、各桁に16の倍数を掛けたものを足す。

16進数

| | | | |
|---|---|---|---|
| 5 | 6 | c | e |
|---|---|---|---|

↓ 4096 (16^3) 倍 ↓ 256 (16^2) 倍 ↓ 16 倍 ↓ 1倍

4096×5 $+ 256 \times 6$ $+ 16 \times 12$ $+ 14$

10進数

22222

16進数と10進数の変換

10進数から16進数へ

10進数を16進数に変換する方法：

1.10進数を直接16進数に変換

2.10進数を2進数に変換してから16進数に変換

10進数を直接16進数に変換

元の10進数を割れなくなるまで16で割り、割ったときの余りを逆に拾っていく。

10進数

16) 64206 (余り)

16) 4012 ... 14 → e

16) 250 ... 12 → c

16) 15 ... 10 → a

16) 0 ... 15 → f

↑下から上に読む

16進数

face

10進数を2進数に変換してから16進数に変換

元の10進数を2進数に変換する。

| | | |
|----|--------------|---------|
| 2) | <u>64206</u> | (余り) |
| 2) | <u>32103</u> | 0 |
| 2) | <u>16051</u> | 1 |
| 2) | <u>8025</u> | 1 |
| 2) | <u>4012</u> | 1 |
| 2) | <u>2006</u> | 0 |
| 2) | <u>1003</u> | 0 |
| 2) | <u>501</u> | 1 |
| 2) | <u>250</u> | 1 |
| 2) | <u>125</u> | 0 |
| 2) | <u>62</u> | 1 |
| 2) | <u>31</u> | 0 |
| 2) | <u>15</u> | 1 |
| 2) | <u>7</u> | 1 |
| 2) | <u>3</u> | 1 |
| 2) | <u>1</u> | 1 |
| 2) | <u>0</u> | 1 |

↑ 下から上に読む

10進数2進数に変換してから16進数に変換

余りを下から上に読むと、

1111101011001110

これを1の位 (右側)から4桁毎に区切る。

1111, 1010, 1100, 1110

| | | | |
|---|---|---|---|
| ↓ | ↓ | ↓ | ↓ |
| f | a | c | e |

Rubyプログラムでの扱い

Rubyプログラム
では、プログラム
中に書く数値の先
頭に付ける記号
(**接頭辞**)で何進
法か指示できる。

| 接頭辞 | 意味 |
|-----|------|
| なし | 10進数 |
| 0b | 2進数 |
| 0 | 8進数 |
| 0x | 16進数 |

Rubyプログラムでの扱い

10 : 10進数の10

0b1010 : 2進数の1010

0xa : 16進数のa

プログラム中でどの表記で書いても計算機内では10進数の10として扱われる。

printfの%フォーマット

| %フォーマット | 何進法で出力されるか |
|---------|------------|
|---------|------------|

| | |
|----|-----|
| %b | 2進数 |
|----|-----|

| | |
|----|-----|
| %o | 8進数 |
|----|-----|

| | |
|----|------|
| %d | 10進数 |
|----|------|

| | |
|----|------|
| %x | 16進数 |
|----|------|

今回の内容

- 2進数と16進数
- **文字コード**
- 文字列のエンコーディング

文字コード

計算機で利用できる文字には全て番号がついている。この番号のことを**文字コード**という。

ASCII (アスキー)コード

アルファベット、数字、記号などには全て**ASCII**コード

(American Standard Code for Information Interchange)が割り当てられている。

ASCIIコード表

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|-----|----|----|----|-----|
| 0 | NUL | SOH | STX | ETX | EOT | ENQ | ACK | BEL | BS | HT | LF | VT | FF | CR | SO | SI |
| 1 | DLE | DC1 | DC2 | DC3 | DC4 | NAK | SYN | ETB | CAN | EM | SUB | ESC | FS | GS | RS | US |
| 2 | | ! | " | # | \$ | % | & | ' | (|) | * | + | , | - | . | / |
| 3 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 4 | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 5 | P | Q | R | S | T | U | V | W | X | Y | Z | [| \ |] | ^ | _ |
| 6 | ` | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o |
| 7 | p | q | r | s | t | u | v | w | x | y | z | { | | } | ~ | DEL |

制御文字

代表的な制御文字

- 0x00 NUL: 空文字 (文字列の終端を表すことが多い)
- 0x07 BEL: 端末のベルを鳴らす
- 0x08 BS: バックスペース (Back Space)
- 0x09 HT: 水平タブ (Horizontal Tab)
- 0x0A LF: 改行 (Line Feed)
- 0x0D CR: 復帰 (Carriage Return; リターン)
- 0x1B ESC: エスケープ (Escape)

キーボードから入力するときは、文字コードに0x40を足した文字をCtrlキーを押しながらタイプする。例えば、0x09のHTは0x40を足した0x49がIなので、C-iが対応するキーとなる。₂₁

今回の内容

- 2進数と16進数
- 文字コード
- 文字列のエンコーディング

文字列のエンコーディング

計算機で日本語文字列など、ASCII文字以外の文字列を表現する方法は複数ある。その方式のことを**エンコーディング**という。

代表的なエンコーディング

日本語を処理するエンコーディング

JISコード (iso-2022-jp):

主に電子メールで用いられる。

日本語EUCコード (euc-jp):

日本語に特化したプログラム処理を作成する場合に主に用いられる。

UTF-8コード (utf-8):

多国語を同時表現する場合に用いられる。

日本語エンコーディングの変換

kconvライブラリを用いると、文字列のエンコーディング変換を簡単に行える。

kconvライブラリ

プログラムで

require 'kconv'

とすると、文字列で以下のメソッドが利用できる。

| | |
|------------|----------------------------|
| 文字列.tojis | 元の文字列をJISコードに変換した文字列を返す |
| 文字列.toeuc | 元の文字列を日本語EUCコードに変換した文字列を返す |
| 文字列.toutf8 | 元の文字列をUTF-8コードに変換した文字列を返す |

まとめ

- 計算機はデータを**2進数**の状態です保存する。
- 計算機内では文字に付された番号（**文字コード**）を用いて文字を扱う。
- 文字列を表現する方式のことを**エンコーディング**という。