PHPでバイナリ変換プログラミング

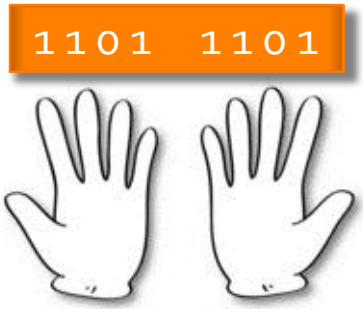
～ openpear/IO_Bit の紹介と応用事例 ～

“よや” <yoya@awm.jp>

自己紹介

- 六本木の方で**Web関連**の仕事をしてます

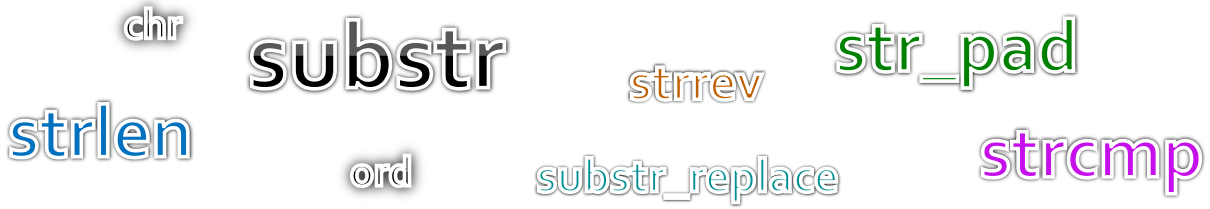
- **バイナリ**変換プログラミングが趣味です



神は人に
２進数を
与えたもう

発表題目

- (pure) PHP で**バイナリ**変換プログラミング
  - ここで云う pure は PHP の**標準関数**だけでという意味

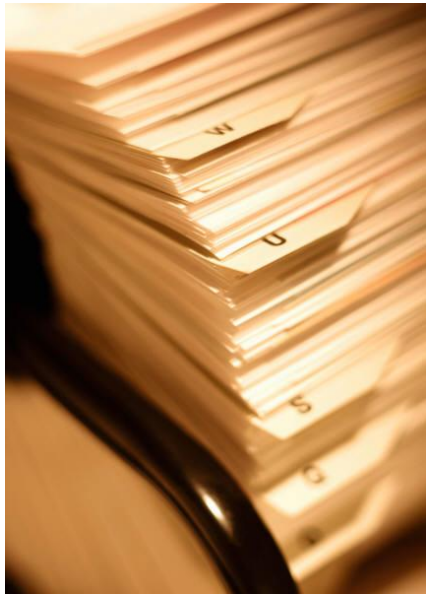


  - Web 開発に**飽きてきた**人向けの発表
    - 前半はバイナリのおさらいです。

。。ていうか PHP はホント楽です。~~JavaScript でのバイナリ処理に比べれば。。~~

発表内容

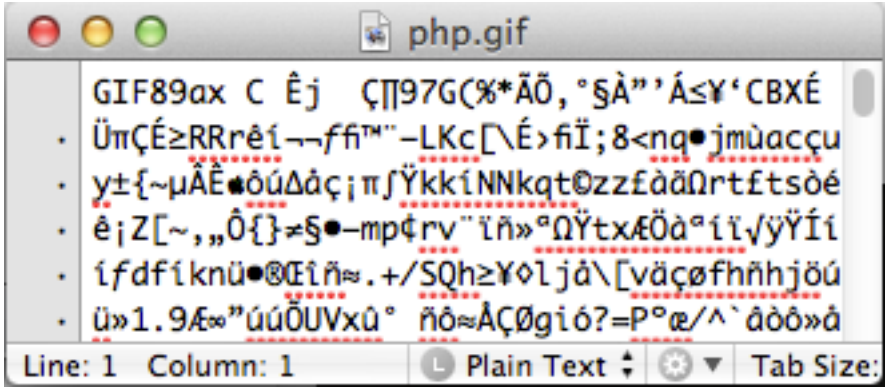
- バイナリについて
- ビット(Bit)とバイト(Byte)について
- PHP でバイト(Byte)処理
- PHP でビット(Bit)処理
- openpear/IO_Bit パッケージの紹介
- IO_Bit の応用事例 (IO_SWF, IO_Zlib)



バイナリについておさらい

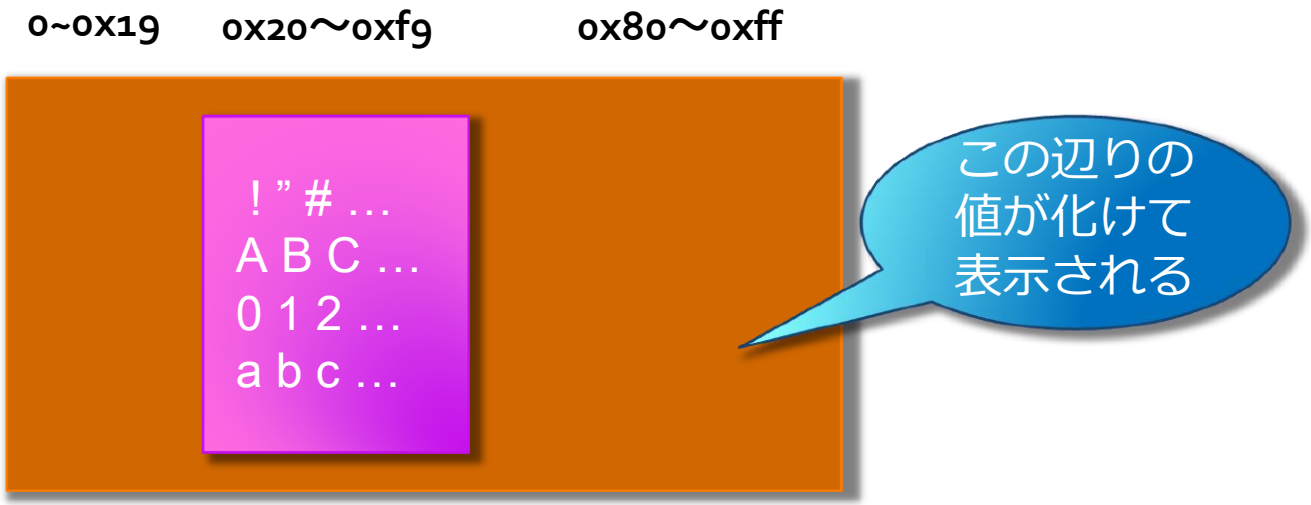
- バイナリって何？
  - 本来は、コンピュータが処理し易い 0,1 の２進値データ
  - 一般的には、**テキスト以外のデータ** (狭義のバイナリ)
    - エディタで開いて文字化けするデータ

GIFファイル
(php.git)



バイナリとテキスト

- １バイトで0～255の値を表現できるけど、テキストはその一部しか使わない。(日本語の話は棚に置きます)



- バイナリの方がより多くの情報を詰められる

バイナリの実例

- バイナリの種類
  - マルチメディア系ファイル (JPEG, PNG, MPEG, AVI...) 
  - 実行ファイル (exe, a.out, jar, ...) 
  - ネットワーク上の通信データ (TCP, ISDN, ...) 
  - 暗号化されたデータ (zip, gzip, ...) 

色々ありますネ！

バイナリ処理の目的

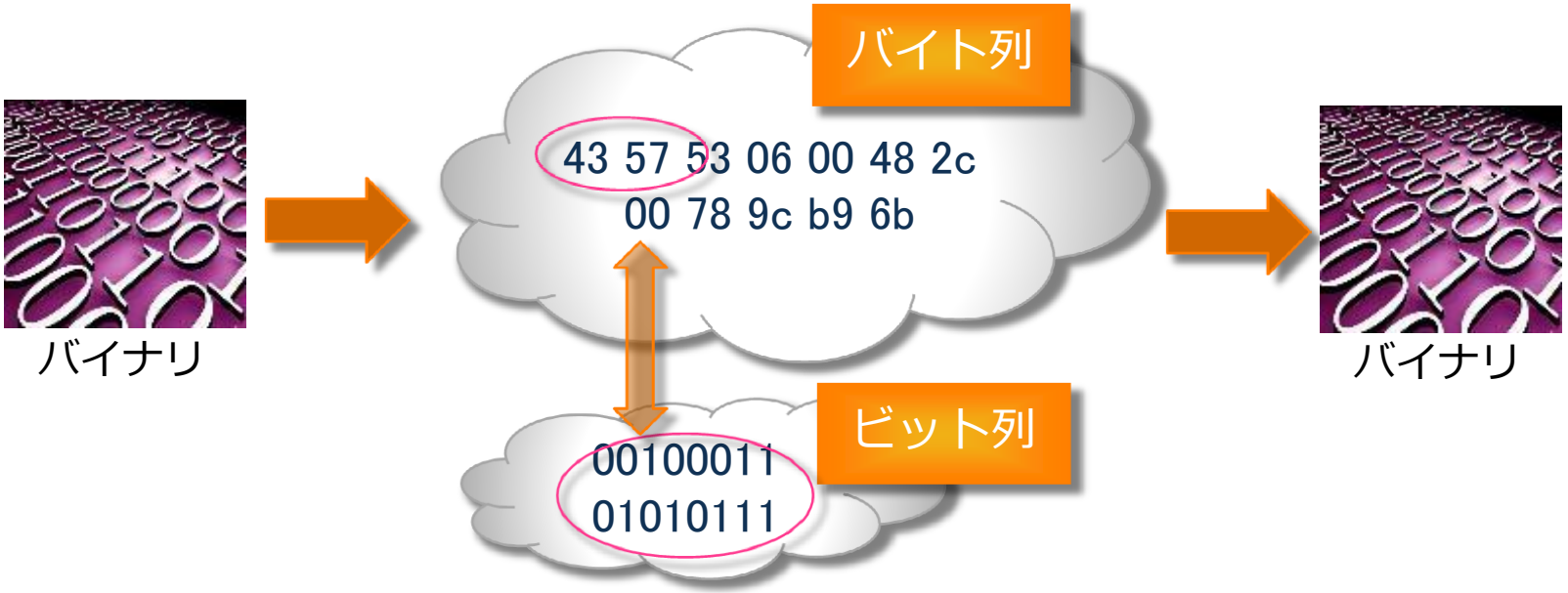
- Web サービスではテキスト以外に画像/動画データや、場合によると生の通信データを扱う事がある
- それらのデータを独自に変換する事で
  - より多くの種類のクライアント端末でサービスが受けられたり

  - 通信データ量を減らせたり出来るかもしれない

- エディタで開いて読めないから諦める。のだと勿体ない

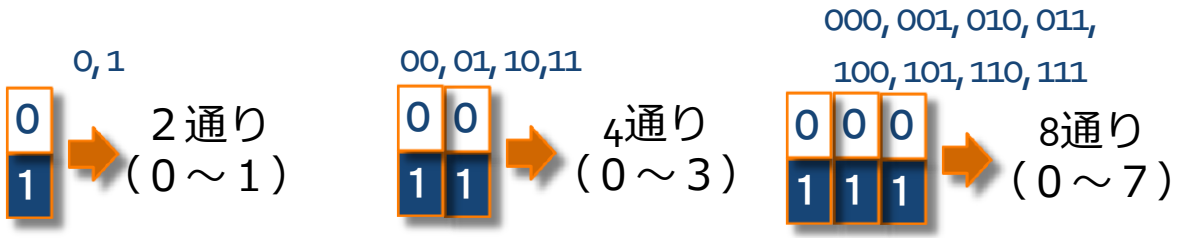
ビット(Bit)とバイト(Byte)

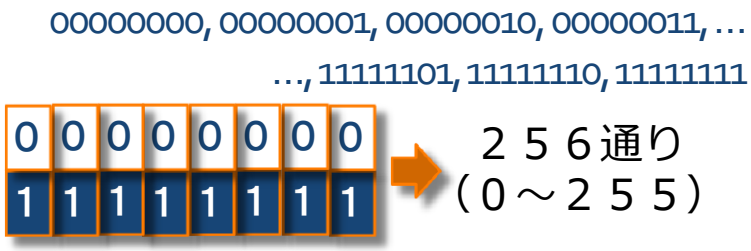
- コンピュータの処理(入出力や編集等)する単位
- バイナリ処理はこれらの単位でデータを操作する



ビット(Bit)

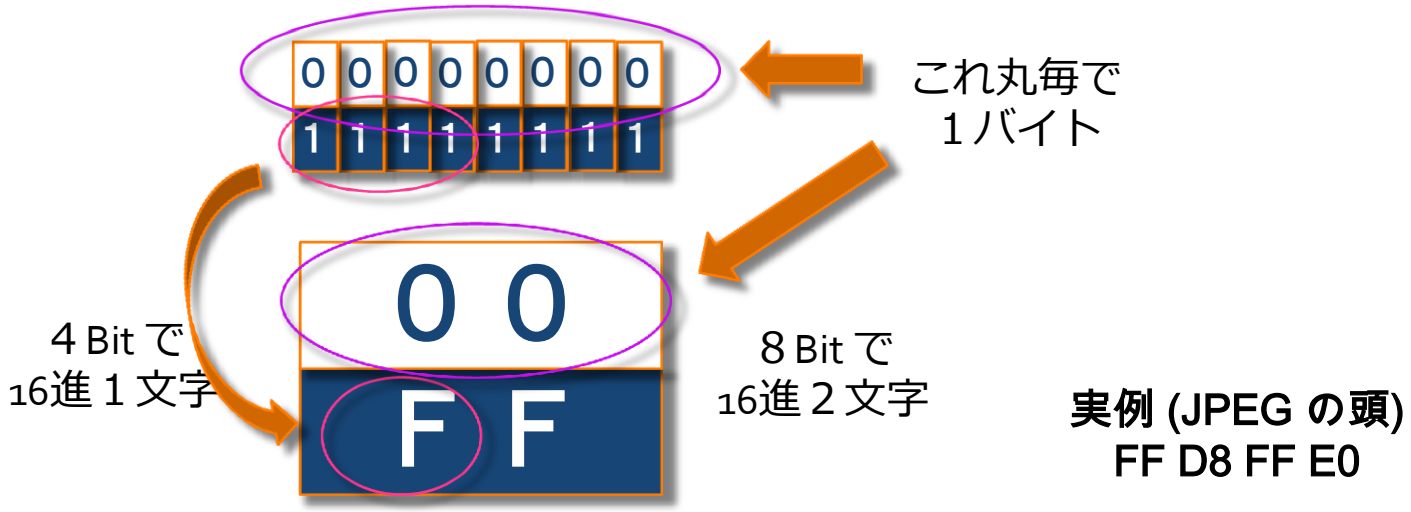
- ビット(Bit)とは
  - ０と１を用いて２通りの状態を表現したもの
  - ビットを並べると４通り８通りと表現の幅が広がる





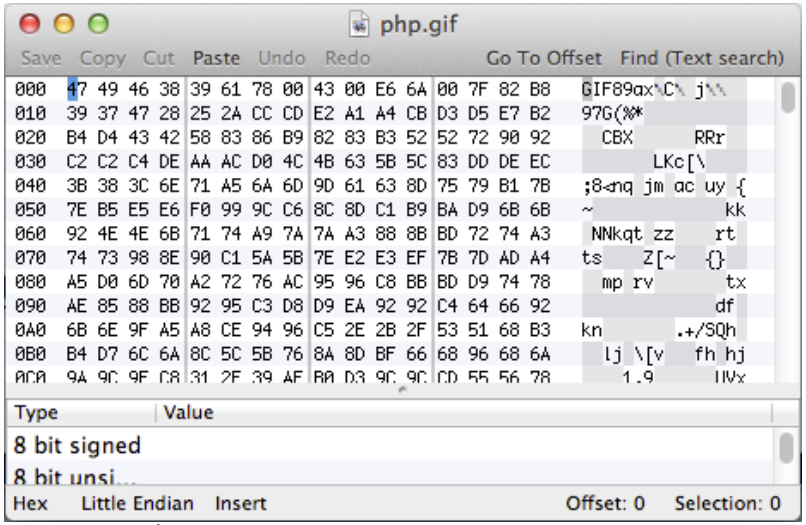
バイト(Byte)

- バイト(Byte)とは
  - 元々は、(欧米の)１文字を表すのに必要なビットの集まり
  - 狭義には**ビットを8つ**まとめた単位
  - １６進数で表現する事が多い(バイナリエディタの表示)



バイナリエディタを使う

- バイナリエディタ諸々
  - Macintosh なら 0xED、Windows なら Stirling, Bz Editor



- 手動で弄るのが**面倒**になったら PHP の出番です。

ここから本題

PHP

&

Byte

PHP とバイナリと String型

- PHP の String 型を使えばバイナリ処理が出来る
  - PHP は String 型に対し文字としての特別な事をしない

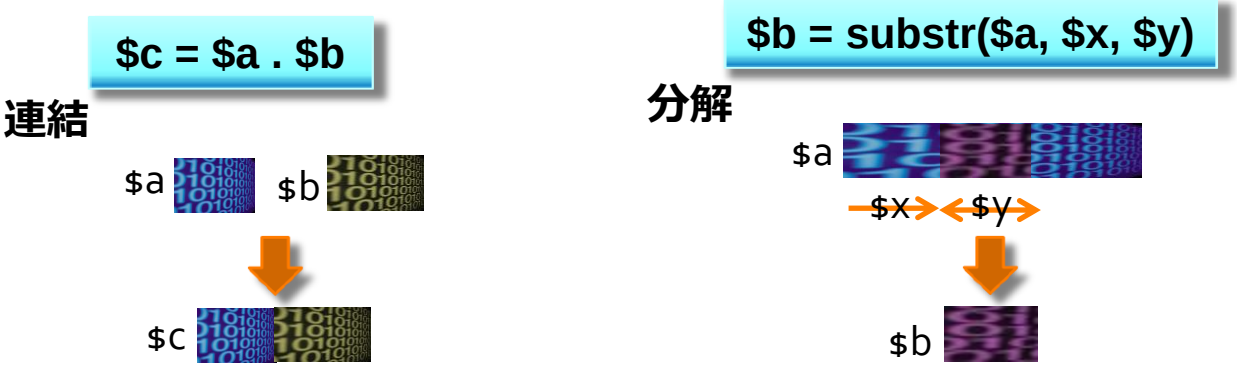
> PHP における文字列型は、**バイトの配列と整数値 (バッファ長) で実装**されています。バイト列を文字列に変換する方法については何の情報も持っておらず、完全にプログラマ任せとなっています。

    - http://www.php.net/manual/ja/language.types.string.php#language.types.string.details
  - 8bitスルー。 ¥0 終端もない ← この辺りの心配は無用
  - つまり、バイト(Byte)単位の処理は PHP でも簡単

String 型でバイト処理 (1/5)

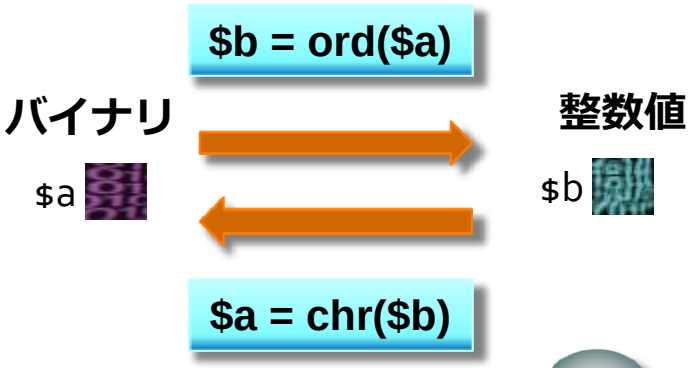
- ファイル入出力

- 連結と分解



String 型でバイト処理 (2/5)

+ バイナリと整数値との相互変換



実行例

```
$a = 'Yoya';
$b = ord($a[0]);
echo $b;
```

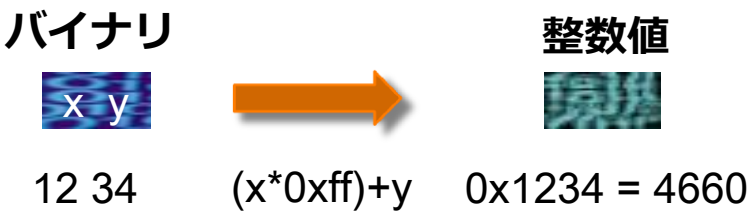
実行結果
⇒ 89
= (0x59)

```
$b = 89; $c = 111;
$a = chr($b).chr($c);
echo $a;
```

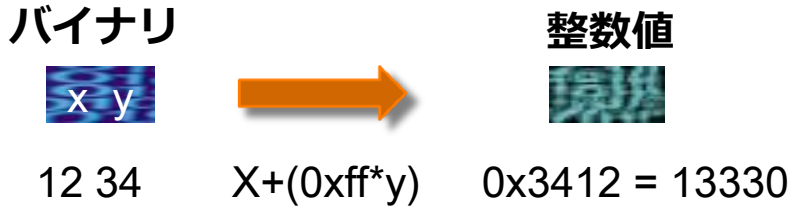
実行結果
⇒ Yo

String 型でバイト処理 (3/5)

- ２バイト以上のバイナリと整数値の相互変換
  - Big Endian (MSB First)

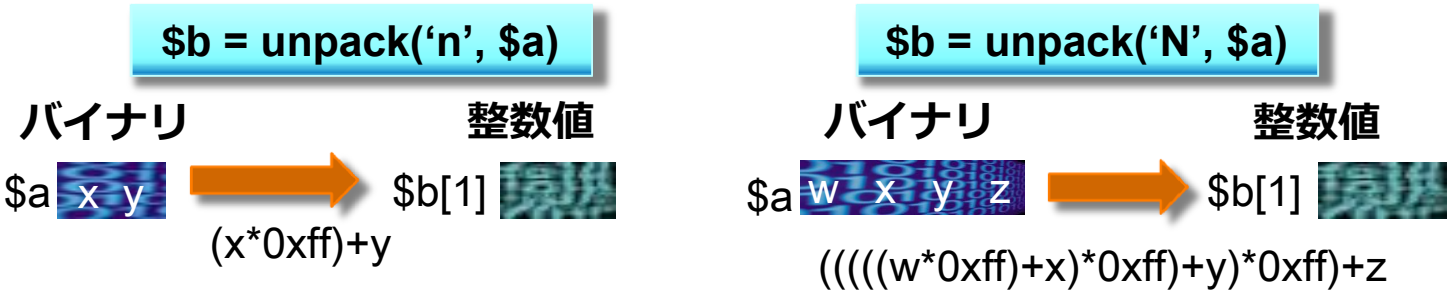


  - Little Endian (LSB first)



String 型でバイト処理 (4/5)

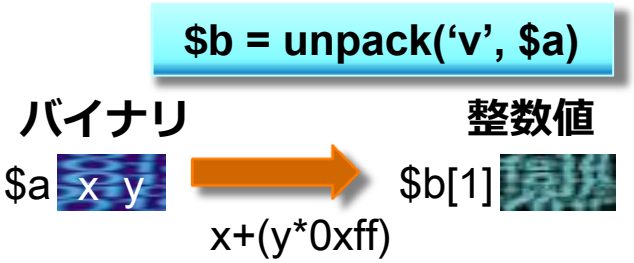
+ バイナリと整数値の相互変換 (Big Endian)

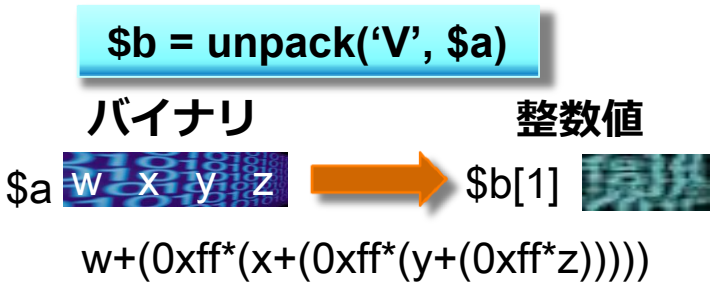


+ pack で逆変換

String 型でバイト処理 (5/5)

+ バイナリと整数値の相互変換 (Little Endian)





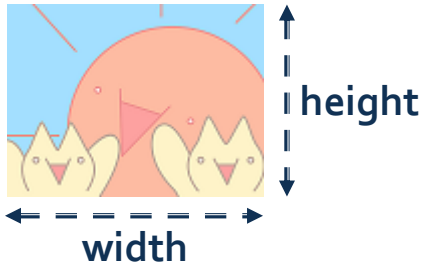
+ pack で逆変換

バイト処理の注意点

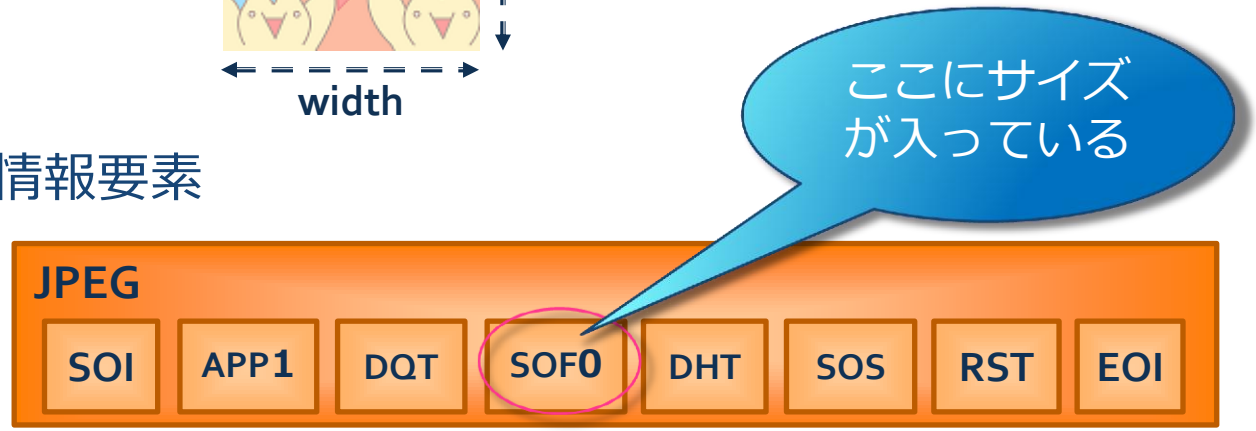
- $a[0]
  - 文字列を配列のように参照すると、($a の 0 番目の数値でなく)、**0番目の文字を切り出したもの**が取得できる。
  - $a[0] は substr($a, 0, 1) と同じ (C言語出身者は多分ココで躓く)
- unpack('N', ... と 'V'の PHP bug
  - N, V は unsigned long (32ビット値)のはずだが、**実際は signed long(符号付き)の値**が出てくる
  - 負の値が出てきたら 4294967296 を足して補正
  - **pack は正でも負でも受理**してくれる。

バイト処理の実例 JPEG分解 (1/2)

- 例えば) JPEG 画像のサイズを抜き出す
  - http://www.w3.org/Graphics/JPEG/ ← 仕様はココ

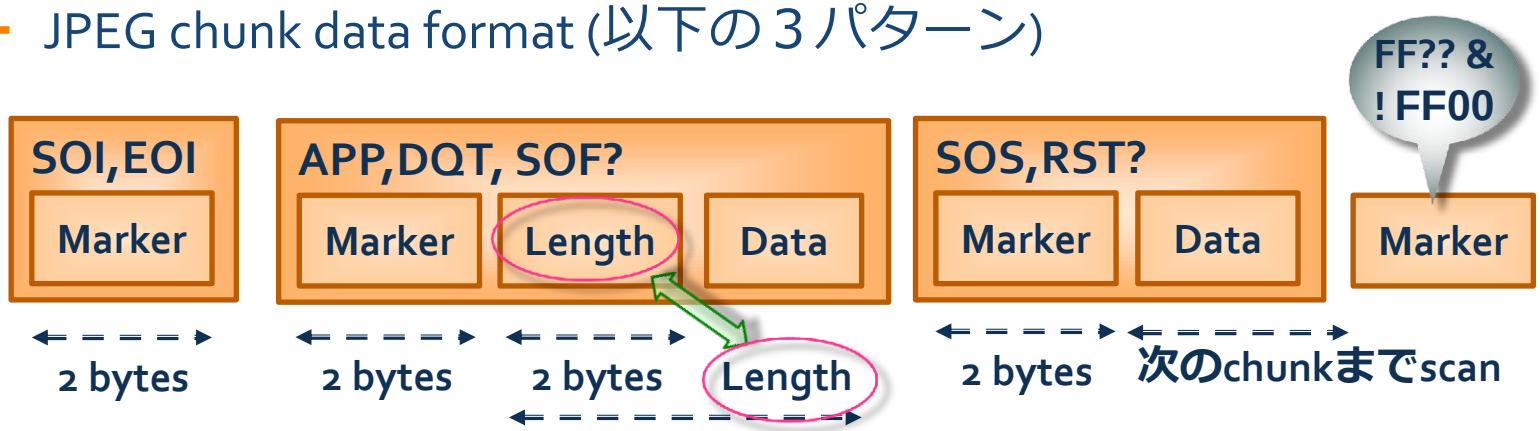


- JPEG 情報要素

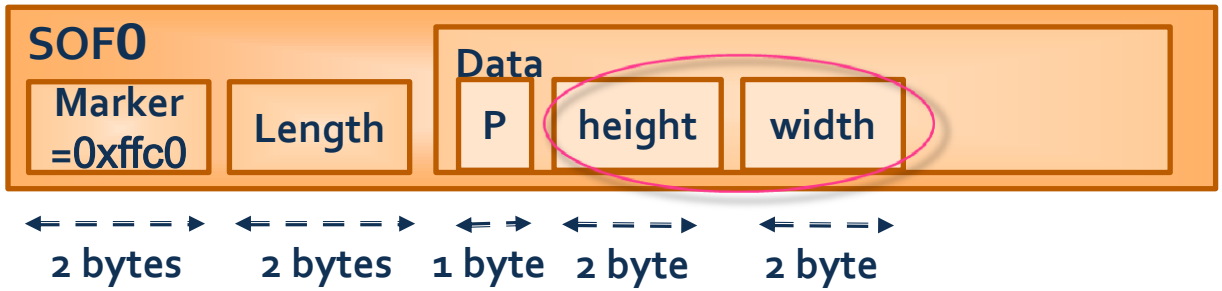


バイト処理の実例 JPEG分解 (2/2)

- JPEG chunk data format (以下の３パターン)



- SOF の中身



バイト処理の実例 JPEG分解 (3/3)

- Code

  Sample

```
$data = file_get_contents($argv[1]);  // jpeg file input
for ($i = 1 ; $i < strlen($data) ; $i++) {
   switch(ord($data[$i++])) { // chunk marker
     case 0xD8: case 0xD9: // SOI or EOI
       break;                    // skip
     default:
       $len = unpack('n', substr($data, $i, 2));
       $i += $len[1];
       break;                    // skip
     case 0xC0: // SOF0
       $sof0 = unpack('CP/nH/nW', substr($data, $i + 2, 5));
       echo "width:{$sof0['W']} heigth:{$sof0['H']}¥n";
       exit (0);                 // OK
   }
}
```

ここからビットの話

PHP

&

Bit

PHP でビット処理

- 論理演算で自前で処理する

PHP でビット読み出し (Read)

PHP でビット書き込み (Write)

ビット処理の別の方法

- ビットシフト

IO_Bit の紹介

- Openpear ～ IO_Bit

  - http://openpear.org/package/IO_Bit

ビット処理のユーティリティです。いちいち、pack v したり、incremental に offset を処理するのが面倒だという人向けです。利用に制限はかけません。コピーも改変もご自由にどうぞ。MIT ライセンスにしました。

IO_Bit の概要

IO_Bit の使い方

IO_Bit の応用例

IO_SWF の概要

IO_SWF の使い方

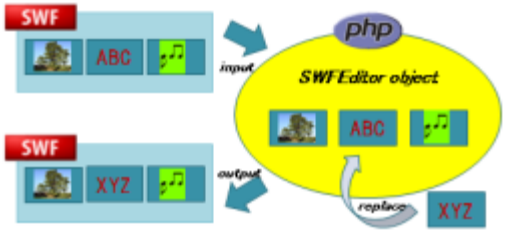
IO_SWF の応用例

IO_Zlib の概要

IO_Zlib の使い方と応用

エクスキューズ

- 実は、IO_SWF は SWFEditor というPHP拡張に似た機能があります。
  - http://sourceforge.jp/projects/swfed/



- 更に、IO_Zlib は標準関数に gzuncompress があります。
  - http://php.net/manual/ja/function.gzuncompress.php

- IO_Zlib はあくまでサンプルという事で。

要望

- 他の言語で、これに似た発表があれば教えて下さい。

以上

- ご清聴ありがとうございました