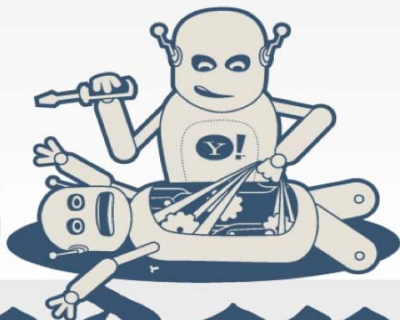


OPEN
HACK
2009



London

PHP and Internationalisation

YAHOO!
DEVELOPER NETWORK



Web 2.0



Mobile
Social
Open

london

YAHOO! DEVELOPER NETWORK

Where is Social?

- US-Centric

Where is Social?

Any-Nation-Centric



London

YAHOO! DEVELOPER NETWORK



Where is Social?

- Multi-culture-centric?
- What is the Social Web in other cultures?



Who is Open?

- Mashups

- Combine applications
- Combine data
- New from old



- Global

- 64.2% online do not speak English



Who is Open?

LOVES YOU
flickr™

Connecté en tant que Rev Dan Catt (2 nouveaux) Aide Déconnexion

Accueil La visite Inscription Explorer

Chercher un lieu Rechercher



The World / France / Île-de-France /

Paris



3° C
low of -1°

10.21 (CET)
mercredi, le 14 novembre

Rechercher (par exemple architecture, urbanisme, forêt)

Photos géniales

OK



underground symmetry II by phil h

Photos intéressantes

Récentes





Yahoo! and PHP



YAHOO! 中国雅虎

服务分类 餐饮 租买房 家政 打折
通讯 邮箱 雅虎通 无线 软件

今日奥运 进入奥运首页

12日盘点-击剑突破历史 传统强项展王者风范

- [策划] 仲满创造奇迹 夺中国男子击剑首金 廖辉力举第13金
- [篮球] 姚明5犯离场 男篮加时惜败西班牙 女足2-0昂首进8强
- [体操] 体操男团洗刷雅典耻辱 夺中国第10金 赛后集体流泪
- [金牌] 陈若琳/王鑫夺中国第11金 高峻混双无缘奥运三连冠
- [策划] 奥运会各国政要表情秀 50岁不老传奇-女剑侠栾菊杰

- [生活] 吃这些10天必瘦 路上拍到布什车队
- [娱乐] 杨钰莹身价暴跌 郎昆妻子高位截瘫
- [明星] 章子怡称永不入美国籍 李宁全家福
- [资讯] 隐身衣将问世 奥运明星状态大预测
- [证券] 沪指跌12点 深跌后哪些股疯狂反扑
- [汽车] 5款车不宜家用 大奔不能超车桑塔纳
- [房产] 穗惊现最低首付 楼市迷局年底揭晓
- [财经] 7月CPI涨6.3% 王楠跃升亿万富婆

推广 佳能摄影PK赢大奖! 翰威特2009年最佳雇主网络提名 奥运生活服务尽在雅虎口碑





From text to bits

- A “glyph” is a single visual unit of text.
- A “character” is a single logical unit of text.
- A “code point” is an integer assigned to a character.
- A “character set” is an organized collection of characters with code points.
- A “character encoding” is a mapping from a sequence of code points (characters) to a sequence of code units.
- A “code unit” is a single logical unit of storage (like a byte, wchar_t, int16_t, etc.)

À
U+00C0

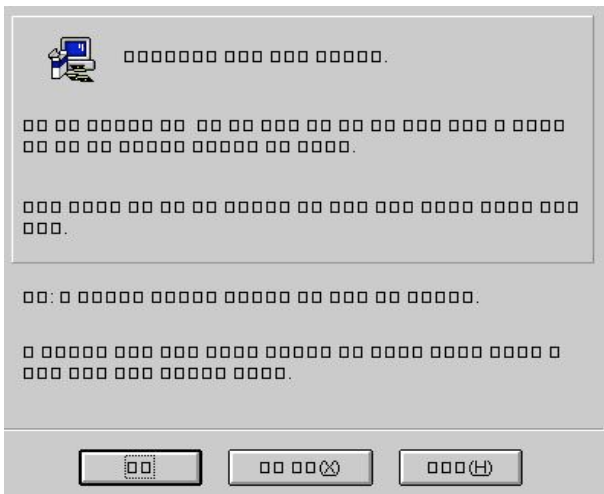
!	"	#	\$	%	&	'	()	*	+	,	-	.	/	0	1	2	3	4
5	6	7	8	9	:	<	=	>	?	@	A	B	C	D	E	F	G	H	
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	[\
]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
q	r	s	t	u	v	w	x	y	z	{		}	~		ı	ç	£	¥	
ı	§	~	©	ª	«	»	®	™	°	±	²	³	´	µ	¶	·	,	.	'
°	»	¼	½	¾	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í
Î	Ï	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß	à	á
â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï	ð	ñ	ò	ó	ô	õ
ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ	Ă	ă	Ą	ą	Ȧ	ȧ	Ć	ć	Ĉ	ĉ

0xC3 0x80 (utf-8)

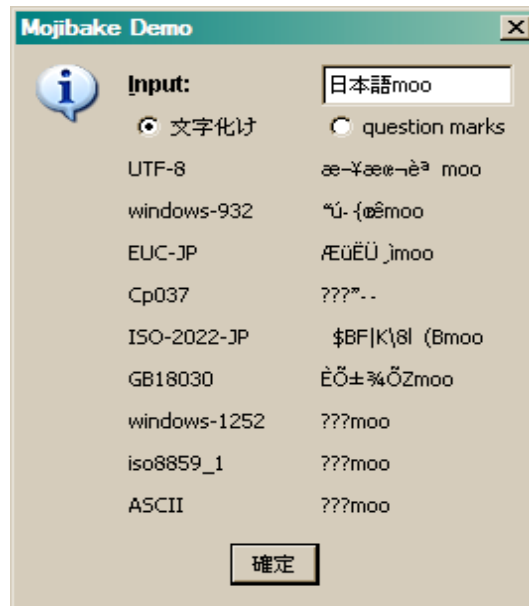


Common Encoding Problems

Tofu hollow boxes



Mojibake garbage characters



Question Marks (conversion not supported)





Examples of Mojibake

http://www.marketwatch.com - **ÍAÇÍÐ: Ìà òíðããò á ÑØÀ òííáíáúá èíááêñú ìääàòò ìà òííá ìèáíí ìííùè ááíèàì** - MarketWatch

MarketWatch

FRONT PAGE **MARKETS**

Global Markets Emerging Markets Earnings Watch

LATEST NEWS [DE:IFX] J.P. Morgan: Infineon was previously underweight

ÍAÇÍÐ: Ìà òíðããò á ÑØÀ òííáíáúá èíááêñú ìääàòò ìà òííá ìèáíí ìííùè ááíèàì

By MarketWatch
Last update: 11:42 a.m. EST Feb. 10, 2009 | [Comments: 2](#)

10 òááðãèÿ /Dow Jones/. Ìà òíðããò á ÑØÀ áí àòíðíèè òííáíáúá èíááêñú ìääàòò ìííèá òííá èáè ìèíèòð òèíáííá Òèíòè Ááéóíáð ìá+áè ìðááíòááèáíèá ìèáíá ìííùè ááíèàì. Èáçíá+áéíðáí ìðááííòááèò 1 òðèí áíèè á ðáíèáò ÿóíè ìðíãðáìí; ìá+íáðñÿ ìðááíðáíèá á æèçíú "ííòááòíðáòòáííè ìðíãðáìí ðúíèá æèèÿÿ".

DJIA óíáè ìà 284,03 ìóíèòá, èèè ìà 3,43%, áí 7986,84 ìóíèòá. Áèòèè Bank of America Corp. (BAC), ìáííáí èç òáð èííòèòóòíá, èíòíðúá áíèæíú á ìáèáíèüðáé ñòáííáíè áúèáðáòò ìò ìðááèòáèüíðááííè ìðíãðáìí ìííùè, ìíááðáááèè ìà 9,4%.

S&P 500 ìíèçèèñÿ ìà 3,71% áí 837,66 ìóíèòá; èíòèðíáèè áèòèè òèíáííáíáí ñáèòíðá òíáèè ìà 4,8%.

After the Bell E-newsletter

Receive a daily email with closing levels of the major indexes and important closing market news commentary.

Enter your email address:

MOST POPULAR

READ **E-MAILED** **EDITOR'S PICK**

1. [Yet more evidence we're in a bear market](#)
2. [Banks may need more capital but AIG reports](#)
3. ['Smart money' starts to bail on stocks'](#)
4. [Asian oil firms' gains built on slippery ground](#)
5. [Chrysler is battleground for Wall Street](#)
6. [Australia holds policy interest rate steady](#)
7. [Interest rates, like Wall Street, must run](#)
8. [Australian dollar rises as RBA stands pat](#)
9. [Palm's future riding on the mysterious](#)

E-mail
Print
Disable Live Quotes
Subscribe to RSS
Yahoo! Buzz

[Recommend this story](#)

[Save and tag this story](#)





Common causes of Mojibake

- Over Conversion
- Under Conversion
- Wrong Conversion
- No Conversion

γĬĀ×ŌŌ° | Ā°Āç/4bÑŌ | ŌÉĀÑ¼a'ò

Ō÷Ō° ŌĐĬĀĬŌ÷Ō°

2008Āa10ŌĀ22ÈŌ ,ĤĀŌĀĀŌĬ±±×¼Ē±¼a09:14±±¼©Ē±¼a 17:14ç±Ĭ

FIGHT MOJIBAKE!!!

ŌĐ°ŭĐĀĬĀ×ŌŌ

ĬĐ°ŭ¼ç¼ÇŌŌ×éŌĬĒ³òĀ

• ŌĐ°ŭÑŌĐŌĀĀĤ¼Ĭ¼ÇŌ

• ĒĀ¼ç±ŌμĐ»Ĥ¼Ĭ¼ĬĀ×Ō

• ŌĐ°ŭÑŌŌŌŌŪĬ°ŌŌ±ĀŌ

ŌĀĀŭÇĀĬÇ°ĀēĤĬĀ

ŸĒĐ-

ŸŌμ¼°Ō,Đμ¼ĒĒĬĀ»ĒĒĒ-ĬĀÇ°ĀēĤĬĀĬĀĒ

unicode.corp.yahoo.com

YAHOO!



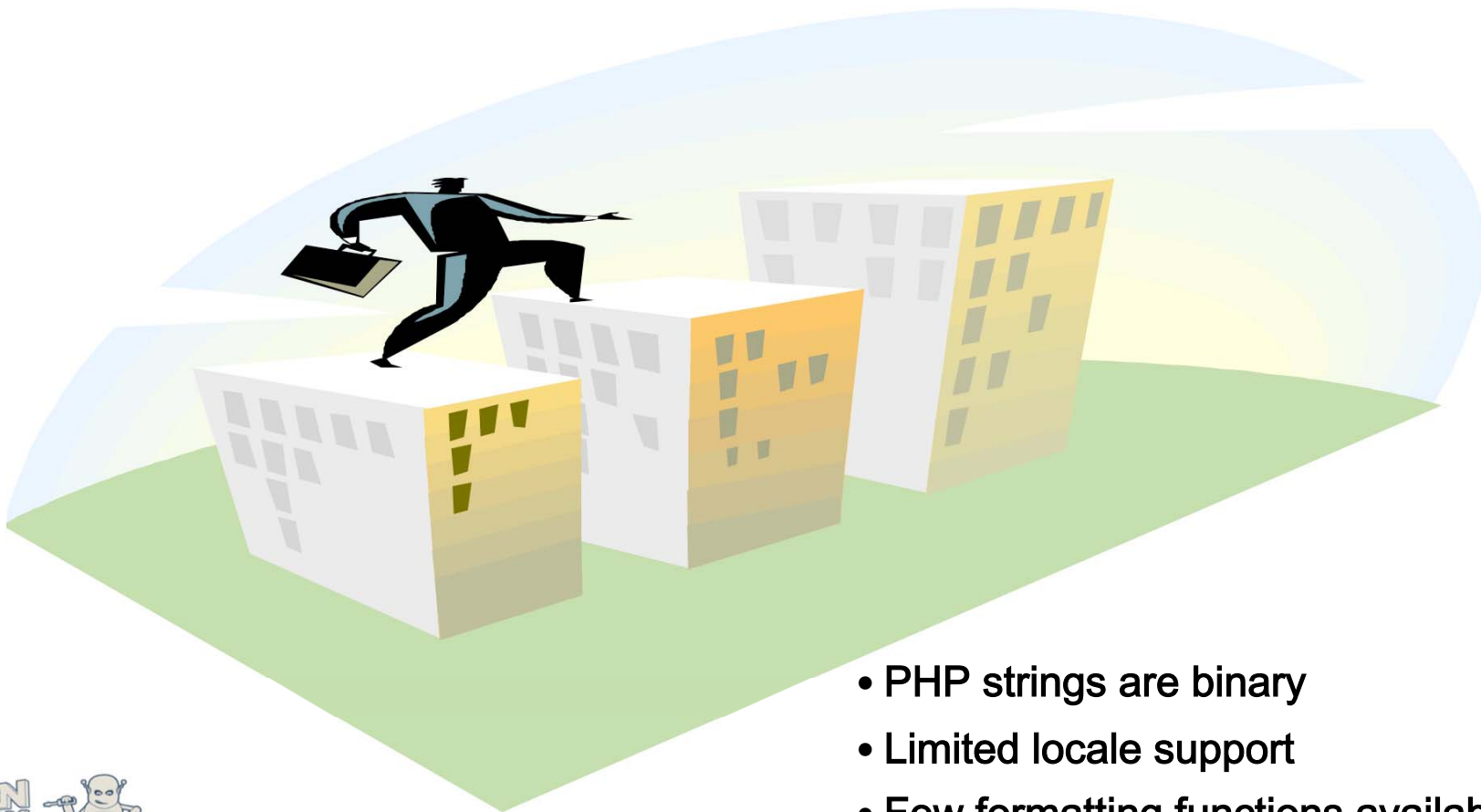
PHP

PHP is a widely-used general-purpose scripting language that is especially suited for Web development and can be embedded into HTML.

```
<html>
<?php
echo 'Hello World!';
?>
</html>
```



Internationalization Challenges with PHP

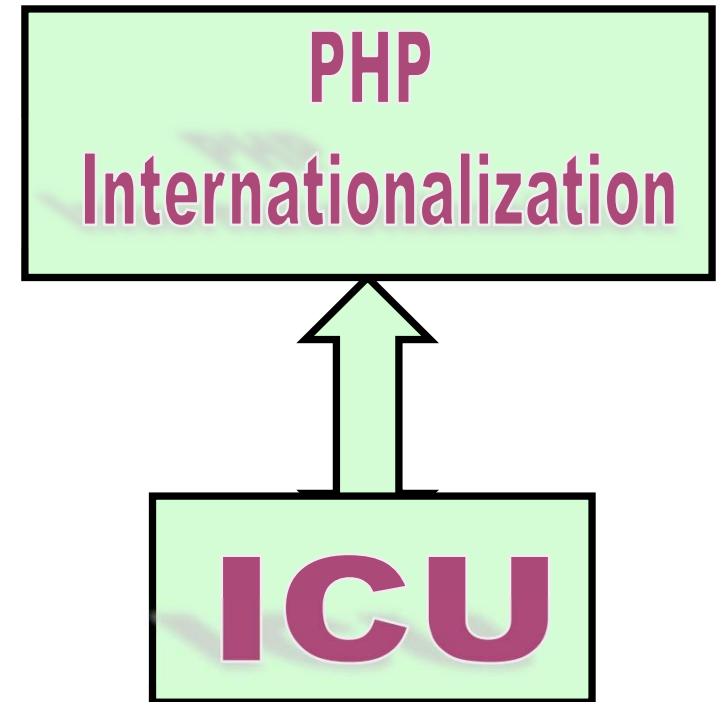


- PHP strings are binary
- Limited locale support
- Few formatting functions available



Strategies

- Implementing Unicode support
 - Write it ourselves?
 - Or use an existing library?
- Decision – Use ICU library for Unicode support
 - Advantages
 - Concerns



International Components
for Unicode

Unicode

Journey of text data

- Detection
- Conversion
- Correction
- Mojibake



Loftoon

YAHOO! DEVELOPER NETWORK



Unicode (ISO 10646)

Unicode is a character set that supports virtually *all the world's languages* and writing systems.

- Code space of up to 0x10FFFF characters (about 1.1 million)
- Unicode and ISO 10646 are maintained in sync
 - Unicode is maintained by an industry consortium.
 - ISO 10646 is maintained by the ISO.
 - Yahoo! is a member of the Unicode Consortium.



Organized into blocks by script

0000

C0 Controls and Basic Latin

	000	001	002	003	004	005	006	007
0	NUL	DLE	SP	0	@	P	`	p
1	SOH	DC1	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3	#	3	C	S	c	s
4	EOT	DC4	\$	4	D	T	d	t
5	ENQ	NAK	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(8	H	X	h	x
9	HT	EM)	9	I	Y	i	y
A	LF	SUB	*	:	J	Z	j	z
B	VT	ESC	+	;	K	[k	{
C	FF	FS	,	<	L	\	l	
D	CR	GS	-	=	M]	m	}
E	SO	RS	.	>	N	^	n	~
F	SI	US	/	?	O	_	o	DEL

C1 Controls and Latin-1 Supplement

	008	009	00A	00B	00C	00D	00E	00F
0	XXX	DCS	NS SP	°	À	Ð	à	ð
1	XXX	PU1	¡	±	Á	Ñ	á	ñ
2	BPH	PU2	¢	²	Â	Ò	â	ò
3	NBH	STS	£	³	Ã	Ó	ã	ó
4	IND	CCH	¤	´	Ä	Ô	ä	ô
5	NEL	MW	¥	µ	Å	Õ	å	õ
6	SSA	SPA	¦	¶	Æ	Ö	æ	ö
7	ESA	EPA	§	·	Ç	×	ç	÷
8	HTS	SOS	¨	¸	È	Ø	è	ø
9	HTJ	XXX	©	¹	É	Ù	é	ù
A	VTS	SCI	ª	º	Ê	Ú	ê	ú
B	PLD	CSI	«	»	Ë	Û	ë	û
C	PLU	ST	¼	½	Ì	Ü	ì	ü
D	RI	OSC	½	¾	Í	Ý	í	ý
E	SS2	PM	®	¾	Î	Þ	î	þ
F	SS3	APC	¿		Ï	ß	ï	ÿ

Tifinagh

	2D3	2D4	2D5	2D6	2D7
0	ⵏ	ⵍ	ⵎ	ⵏ	
1	ⵐ	ⵑ	ⵒ	ⵓ	
2	ⵔ	ⵕ	ⵖ	ⵗ	
3	ⵘ	ⵙ	ⵚ	ⵛ	
4	ⵜ	ⵝ	ⵞ	ⵟ	
5	ⵠ	ⵡ	ⵢ	ⵣ	
6	ⵤ	ⵥ	ⵦ	ⵧ	
7	⵨	⵩	⵪	⵫	
8	⵬	⵭	⵮	ⵯ	
9	⵰	⵱	⵲	⵳	
A	⵴	⵵	⵶	⵷	
B	⵸	⵹	⵺	⵻	
C	⵼	⵽	⵾	⵿	
D	ⶀ	ⶁ	ⶂ	ⶃ	
E	ⶄ	ⶅ	ⶆ	ⶇ	
F	ⶈ	ⶉ	ⶊ	ⶋ	

Devanagari

	090	091	092	093	094	095	096	097
0		ऐ	ठ	र	ी	ॐ	ऋ	०
1	ँ	ऑ	ड	र	ं	ॠ		
2	ं	ओ	ढ	ल	ॡ	ॡ		
3	ः	ओ	ण	ळ	ॢ	ॢ		
4	अ	औ	त	ळ	ॣ	ॣ		
5	अ	क	थ	व	।	॥		
6	आ	ख	द	श	॥	०		
7	ह	ग	घ	ष	॥	१		
8	ई	घ	न	स	॥	क	२	
9	उ	ड	न	ह	॥	ख	३	
A	ऊ	च	प		॥	ो	ग	४
B	ऋ	छ	फ		॥	ो	ज	५
C	ॠ	ज	ब	॥	॥	ौ	ड	६
D	ॡ	झ	भ	॥	॥	०	ड	७
E	ॢ	ञ	म	॥	॥	०	फ	८
F	ॣ	ट	य	॥	॥	०	य	९



Unicode Encodings Compared

A (U+0041)

UTF-32: 0x00000041

UTF-16: 0x0041

UTF-8: 0x41

Æ (U+1251)

UTF-32: 0x00001251

UTF-16: 0x1251

UTF-8: 0xE1 0x89 0x91

À (U+00C0)

UTF-32: 0x000000C0

UTF-16: 0x00C0

UTF-8: 0xC2 0x80

Ψ (U+10338)

0x00010338

0xD800 0xDF38

0xF0 0x90 0x8C 0xB8



Graphemes example: Hindi

What is **Unicode**

यूनिकोड क्या है?

यू नि को ड

यू न किो ड

न + ि = नि



HTML and PHP

- **Set PHP or Apache to declare UTF-8 in HTTP Content-Type header**
- Declare UTF-8 in META tag header
- **Actually use UTF-8 as the encoding!!**

```
<?php
  header("Content-type: text/html; charset=UTF-8");
?>
```

```
<html>
```

```
<head>
```

```
  <meta http-equiv="Content-Type"
```

```
    content="text/html; charset=UTF-8" />
```

```
  <title>Fight 文字化け!</title>
```

```
</head>
```





Pecl/intl Extension in PHP 5.3 Core

- Open Source Project

<http://pecl.php.net/package/intl>

- Goal
- Features



Dual API: Procedural and OO

Procedural	Object Oriented
locale_get_script()	Locale::getScript()
collator_sort()	Collator::sort()





Features

These functions use ICU

Feature	PHP 5.2/5.3	PHP 6.0
Locale Handling	Y	Y
Date/time Formatting	Y	Y
Number/currency Formatting	Y	Y
Message/Choice Formatting	Y	Y
Normalization		
Grapheme clusters	Y	Y

Technical Issues



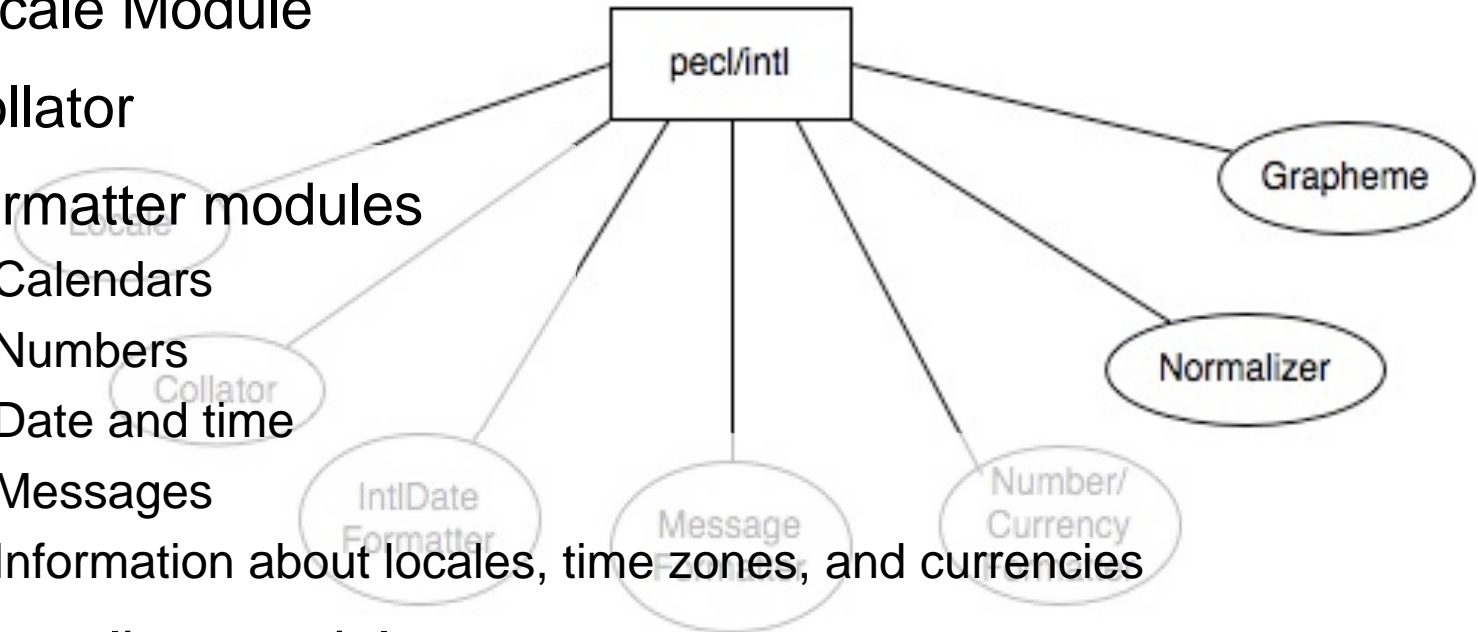
twitbot

YAHOO!. DEVELOPER NETWORK



Intl Modules

- Locale Module
- Collator
- Formatter modules
 - Calendars
 - Numbers
 - Date and time
 - Messages
 - Information about locales, time zones, and currencies
- Normalizer module
 - Normalize Text
 - Checks if text is in normalized form
- Grapheme module
 - Clusters, words, lines, and sentences.





Module – Locale

- Hierarchical fall-back (en-UK → en)
- RFC 4646 (Language tags)
- Root and default locales
- In addition to the ICU functionality: filterMatches, lookup, compose, parse

Based on RFC 4647 (Matching of Language Tags)



Locale::get – Example Usage

Examples to get different subtags:

```
<?php
echo Locale::getPrimaryLanguage('sl-Latn-IT-nedis') . " , ";
echo Locale::getScript('sl-Latn-IT-nedis') . " , ";
echo Locale::getRegion('sl-Latn-IT-nedis') . " , ";
echo Locale::getVariant('sl-Latn-IT-nedis');
?>
```

The output will be as the following:

```
sl , Latn , IT , NEDIS
```



Locale::getDisplay – Example Usage

Examples of getDisplay methods:

```
<?php
echo Locale::getDisplayLanguage('sl-Latn-IT-nedis', 'de');
echo Locale::getDisplayScript('sl-Latn-IT-nedis', 'de');
echo Locale::getDisplayRegion('sl-Latn-IT-nedis', 'de') . " , ";
echo Locale::getDisplayVariant('sl-Latn-IT-nedis', 'de') . "\n";

echo locale_get_display_name('sl-Latn-IT-nedis', 'de');
?>
```



The output will be as the following:

```
Slowenisch , Lateinisch , Italien , NEDIS
Slowenisch (Lateinisch, Italien, NEDIS)
```



Locale::lookup – Example Usage

Locale::lookup -- locale_lookup —
Searches the language tag list for the **best match** to the language

```
<?php
$arr = array(
    'de-DEVA',
    'de-DE-1996',
    'de',
    'de-De' );
echo Locale::lookup( $arr ,
    'de-DE-1996-x-prv1-prv2' , 'en_US' );
?>
```

The output is:

```
de-DE-1996
```



Locale::filterMatches – Example Usage

**Locale::filterMatches -- locale_filter_matches –
Checks if a language tag filter matches with locale**

```
<?php  
echo (Locale::filter_matches('de-DEVA', 'de-DE')) ?  
"Matches" : "Does not match";  
echo ' ; '  
echo (Locale::filter_matches('de-DE-1996', 'de-DE')) ?  
"Matches" : "Does not match";  
?>
```

The above example will output:

```
Does not match ; Matches
```



Module – IntlDateFormatter

- Used for date and time formatting /parsing
- Locale dependent
- Time zones from Olson database
- Still need to make sure other calendars are fully supported





IntlDateFormatter – Example Usage

Simple date formatting example:

```
<?php
fmt = new IntlDateFormatter( "en-US" ,IntlDateFormatter::FULL,
IntlDateFormatter::FULL,
'America/Los_Angeles',IntlDateFormatter::GREGORIAN );
echo "First output is " .fmt->format(0);//plain English

fmt = new IntlDateFormatter( "de-DE" ,IntlDateFormatter::FULL,
IntlDateFormatter::FULL,
'America/Los_Angeles',IntlDateFormatter::GREGORIAN );
echo "Second output is " .fmt->format(0);//Same in German
?>
```

Output is:

First output is Wednesday, December 31, 1969 4:00:00 PM PT

Second output is Mittwoch, Dezember 31.1969, 16:00 Uhr GMT-08.00



IntlDateFormatter Format with Pattern Example

Date formatting example with pattern specified:

```
$fmt = new IntlDateFormatter( "en-GB" ,IntlDate  
IntlDateFormatter::FULL, 'America/Los Angeles',  
IntlDateFormatter::GREGORIAN , "yyyy-MM-dd");  
echo "First output with pattern is " . $fmt->format($date);  
  
$fmt = new IntlDateFormatter( "de-DE" ,IntlDate  
IntlDateFormatter::FULL, 'America/Los Angeles',  
IntlDateFormatter::GREGORIAN , "yyyy-MM-dd");  
echo "Second output with pattern is " . $fmt->format($date);
```

```
م ٢:٢٨:١٧ ٢٠٠٩/٠٣/١٨  
18.03.2009 14:28:17  
18 mars 2009 14:28:17  
२००९ मार्च १८ १४:२८:१७  
2009/03/18 14:28:17  
2009. 3. 18. 오후 2:28:17  
2009/3/18 下午 2:28:17  
18 Mar 2009 14:28:17
```

Output is:

First output (en-GB) with pattern is 1969-12-31

Second output (de-DE) with pattern is 1969-12-31



Module – NumberFormatter

٧	٦	٥	٤	٣	٢	١	GR - 2009 (v 1.0)	Status
[Plan] نعم	[Plan] نعم	نعم [لا]	نعم [لا]	لا	نعم [لا]	نعم	لا	
<u>٢٥١٨٨٥٢</u>	<u>٢٥٧١٢٣٩</u>	<u>٢٥١٨٩٥٠</u>	<u>٢٥١٨٩٧٨</u>	يوجد	<u>٢٥١٨٨٠٠</u>			
FIXED	FIXED	NEW	NEW	-- لا تنطبق	NEW			

Formats numbers and currencies

Locale dependent

For currencies, extra currency string parameter

Exchange rate?



NumberFormatter – Example Usage

Number formatting:

```
<?php  
$fmt = numfmt_create( 'de_DE', NumberFormatter::DECIMAL );  
$data = numfmt_format($fmt, 1234567.891234567890000);  
?>
```

Output is:

1.234.567,891



NumberFormatter Currency Formatting – Example Usage

Currency formatting:

```
<?php
fmt = new NumberFormatter( 'de_DE', NumberFormatter::CURRENCY );
echo fmt->formatCurrency( 1234567.891234567890000, "EUR" )."\n";

fmt = new NumberFormatter( 'ru_RU', NumberFormatter::CURRENCY );
echo fmt->formatCurrency( 1234567.891234567890000, "EUR" )."\n";
?>
```

Output is:

```
1.234.567,89 €
1 234 567,89 €
```





Collator::sort – Example Usage

Collator_sort example:

```
<?php
$coll = collator_create( 'en_US' );
$arr  = array( 'at', 'às', 'as' );
var_export( $arr );

collator_sort( $coll, $arr );
var_export( $arr );
?>
```

Output is:

```
array (
  0 => 'at',
  1 => 'às',
  2 => 'as',
)

array (
  0 => 'as',
  1 => 'às',
  2 => 'at',
)
```



Collator::asort – Example Usage

Collator_asort example:

```
<?php
$coll = collator_create( 'en_US' );
$arr = array(
    'a' => '50',
    'c' => '100',
    'b' => '7'
);
collator_asort( $coll, $arr,
    Collator::SORT_NUMERIC );
var_export( $arr );

collator_asort( $coll, $arr,
    Collator::SORT_STRING );
var_export( $arr );
?>
```

Output is:

```
array (
  'b' => '7',
  'a' => '50',
  'c' => '100',
)

array (
  'a' => '50',
  'b' => '7',
  'c' => '100',
)
```



Module – Normalizer

Normalization is a process that involves transforming characters and sequences of characters into a formally-defined underlying representation

Example #2 OO example

```
<?php
$char_A_ring = "\xC3\x85"; // 'LATIN CAPITAL LETTER A WITH RING ABOVE' (U+00C5)
$char_combining_ring_above = "\xCC\x8A"; // 'COMBINING RING ABOVE' (U+030A)

$char_1 = Normalizer::normalize( $char_A_ring, Normalizer::FORM_C );
$char_2 = Normalizer::normalize( 'A' . $char_combining_ring_above, Normalizer::FORM_C );

echo urlencode($char_1);
echo ' ';
echo urlencode($char_2);
?>
```

Å

◦

The above example will output:

%C3%85 %C3%85

Å



Module – Grapheme

"Grapheme cluster" is the term the Unicode Standard uses for a Unicode character sequence that users usually perceive as a single character.

यू नि को ड
यू न कि को ड

In some writing systems, grapheme clusters using combining marks are the only way to represent many user-perceived characters, for example, Indic languages.



Module - Grapheme



"Grapheme cluster" is the term the Unicode standard uses for a Unicode character sequence that users usually perceive as a single

Examples

Example #1 grapheme_strlen() example

```
<?php
$char_a_ring_nfd = "a\xCC
\x8A"; // 'LATIN SMALL LETTER A WITH RING ABOVE' (U+00E5) normalization form "D"
$char_o_diaeresis_nfd = "o\xCC
\x88"; // 'LATIN SMALL LETTER O WITH DIAERESIS' (U+00F6) normalization form "D"

print grapheme_strlen( 'abc' . $char_a_ring_nfd . $char_o_diaeresis_nfd . $char_a_ring_nfd)
?>
```



The above example will output:

6



Module – MessageFormatter

Goal - to build all the messages that are seen by end users.

Example usage formatting a message:

```
<?php
$fmt = new MessageFormatter("en-US", "You have {0} {1,number,integer}.");
echo $fmt->format(array("opened", 34));

$fmt = new MessageFormatter("de-DE", "Sie haben (1,number,integer) Dateien
{0}.");
Echo $fmt->format(array("geöffnet", 34));
?>
```

The above example will output:

```
You have opened 34 files.
Sie haben 34 Dateien geöffnet.
```



Resource Bundles

PHP_rb

- Simple text files, always UTF-8
 - Similar to Unix message catalogs, Java resource bundles, etc.
- Loaded and cached at runtime
 - No build steps necessary
 - Uses open source SDBM database
- APIs for C, C++, and PHP
- Open Source
 - php.net (PHP)
 - github.org (C++)





Resource Bundles: Files

- Patterns

- {resource directory}/{**basename**}.pres

- {resource directory}/{**basename**}_{language}.pres

- Examples

- /home/y/var/yrb/profiles/**ui**.pres

- /home/y/var/yrb/profiles/**ui**_ko-KR.pres

- Fallbacks

- When loading bundle

- When looking up string





Resource Bundles: Core PHP API

- Core
 - *PropertyResourceBundle*(
 string **basename**, string **locale**)
 - string *get_string*(string **key**)
- Plus
 - string *get_actual_locale*()
 - string *get_keys*()
 - etc...



Hello World – Guten Tag Welt

```
<?php
```

```
// get language tag from the command line
```

```
$lang = $argc > 1 ? $argv[1] : "";
```

```
// get localized string from resource bundle and print it
```

```
$rb = new ResourceBundle(  
    "demos/hello/strings", $lang);
```

```
echo $rb->get_string("HELLO") . "\n";
```

```
?>
```



Hello World – Message File

```
# English message file  
HELLO = Hello World!
```

```
# Chinese message file  
HELLO = 你好世界!
```

```
#  
# Application root Resource Bundle.  
#  
SPACEPORT_HEADING=Spaceport:  
SPACEPORT_HEADING1=<b>Spaceport:</b><br>\n  
PRODUCTS_BUTTON=Products  
TEAMS_BUTTON=Teams  
ASSESSMENT_BUTTON=Assessments  
USERS_BUTTON=Users
```

```
#  
# 太空根資源束。 Zh-Hant-TW pres  
#  
SPACEPORT_HEADING=太空:  
SPACEPORT_HEADING1=的<b>太空:</b><br>\n  
PRODUCTS_BUTTON=產品  
TEAMS_BUTTON=團隊  
ASSESSMENT_BUTTON=評估  
USERS_BUTTON=用戶  
IS_REPORTING_TO= %s的報告:<br>\n  
IS_MANAGER_OF= %s是一個經理:<br>\n  
NO_TEAMS_FOUND=&nbsp;沒有球隊被發現! <br>\n  
NOT_A_MEMBER=&nbsp; %s是不屬於任何球隊在太空。<br>\n
```




YRB – meta data

```
<?php
    $prb = new ResourceBundle("sports/ui","en-US");
    $base = $prb->get_basename();
    echo "Basename is " . $base . "\n";
    $actual_loc = $prb->get_actual_locale();
    echo "Actual Locale is " . $actual_loc . "\n";
    $req_loc = $prb->get_requested_locale();
    echo "Requested Locale is " . $req_loc . "\n";
    $res = $prb->get_string("HELLO_WORLD");
    echo "Resource String for ID - HELLO_WORLD is " . $res . "\n";
```



YRB – get the keys

```
<?php
    $prb = new ResourceBundle("news/ui",
        "zh-Hans-CN");

    // gets the list of keys associated with resource bundle
    // "news/ui_zh-Hans-CN"

    // including fallback resource keys

    $arr = $prb->get_keys();

?>
```



More PHP Features

More in “Human Language and Character Encoding Support”
<http://uk.php.net/manual/en/refs.international.php>

Feature	
Transliterations	iconv
IDN and web addresses	idn
BIDI algorithm	FriBiDi
MultiByte Strings	mb* with UTF-8 and mbstring.func_overload = 7 in php.ini



PHP Resources

- PHP International Functions
<http://uk.php.net/manual/en/refs.international.php>
- Pecl/intl documentation <http://uk.php.net/intl>
- PHP 5.3 doc <http://www.php.net/manual/en/book.intl.php>
- ICU <http://www.icu-project.org/>
- RFCs <http://www.ietf.org/rfc.html>

OPEN
HACK
2009



London



Questions

and comments and feedback!



YAHOO!
DEVELOPER NETWORK



Photo Credits

www.flickr.com/photos/4real/2737847473.jpg

www.flickr.com/photos/30401414@N04/2909364745.jpg

www.flickr.com/photos/francistill/2352451333.jpg

www.flickr.com/photos/gaetanlee/118885175.jpg

www.flickr.com/photos/marcwathieu/2688549097.jpg

www.flickr.com/photos/gabrieljaquemet/1934926002.jpg

www.flickr.com/photos/theshanghaieye/261346205.jpg

www.flickr.com/photos/rajchettupuzha/177597956.jpg

www.flickr.com/photos/manganite/173253613.jpg

www.flickr.com/photos/ashleys-natural-images/2455299481.jpg

www.flickr.com/photos/psychonautfromatlantis/661507507.jpg

www.flickr.com/photos/lorenzocuppini/1830418930.jpg

www.flickr.com/photos/cloughridge/2530747104.jpg

NASA

www.baekdal.com/web2dna/ - convert web pages to DNA!

