

How to get pure PHP language parser


hnw

a.k.a. Yoshio HANAWA(塙 与志夫)

PHPMatsuri発表資料

Who am I?

- id:hnw
- 細かすぎるツッコミ担当



php.net | support | documentation | report a bug | advanced search | search howto | statistics | login

go to bug id or search bugs for

Showing 1-13 of 13

ID#	Date	Last Modified	Package	Type	Status	PHP Version	OS	Summary	Assigned
52826 (edit)	2010-09-13 09:32 UTC	2010-09-13 09:42 UTC	Bzip2 Related	Bug	Closed	5.3.3	any	phpinfo is incorrect for bzip2 stream wrapper	aharvey
48375 (edit)	2009-05-24 16:19 UTC	2009-07-16 01:07 UTC	SPL related	Bug	Assigned	5.2.9	*	AppendIterator::append() never ends, which has SplFileObject as inner iterator	colder
48254 (edit)	2009-05-13 03:47 UTC	2009-05-14 01:42 UTC	Arrays related	Bug	Closed	5.2.9	*	Inconsistent handling of huge numerical keys for array	mattwil
47836 (edit)	2009-03-30 07:40 UTC	2009-06-07 19:32 UTC	Scripting Engine problem	Bug	Closed	5.2.9	*	array operator [] inconsistency when the array has PHP_INT_MAX index value	
47752 (edit)	2009-03-23 05:40 UTC	2009-11-25 10:41 UTC	Filter related	Bug	Closed	5.2.9	*	FILTER_VALIDATE_INT doesn't allow "+0" and "-0"	pajoye
47745 (edit)	2009-03-21 23:34 UTC	2009-03-31 10:06 UTC	Filter related	Bug	Closed	5.2.9	*	FILTER_VALIDATE_INT doesn't allow minimum integer	dmitry
47399 (edit)	2009-02-16 00:05 UTC	2009-02-16 01:58 UTC	mbstring related	Bug	Closed	5.2.9RC2	any	mb_check_encoding returns true with illegal SJIS string	moriyoshi
47370 (edit)	2009-02-12 16:22 UTC	2009-05-15 17:10 UTC	Documentation problem	Doc	Closed	5.2.9	*	array_unique has backward compatibility problem, and SORT_REGULAR is confusing	
46932 (edit)	2008-12-23 06:18 UTC	2009-09-15 06:52 UTC	Documentation problem	Doc	Closed	5.2.8	any	strtotime() has inconsistency between 'next Monday' and '+1 Monday'	
46930 (edit)	2008-12-22 23:09 UTC	2009-05-03 14:45 UTC	Documentation problem	Doc	Closed	5.3.0alpha3	any	5.3.0alpha3's strtotime() returns inconsistent result with some relative items	derick
46926 (edit)	2008-12-22 13:55 UTC	2009-06-25 04:33 UTC	Performance problem	Bug	Bogus	5.2.8	*	mktime() is slow when it gets very small 3rd argument	derick
46478 (edit)	2008-11-04 12:56 UTC	2009-12-22 05:50 UTC	Feature/Change Request	Req	Closed	5.2.6	*	htmlentities() uses obsolete mapping table for character entity references	moriyoshi
45712 (edit)	2008-08-05 02:39 UTC	2008-08-07 08:36 UTC	Variables related	Bug	Closed	5.2CVS, 5.3CVS, 6CVS (2008-08-05)	*	\$something == NaN returns true with 5.3, false with 5.2.*	tony2001

Showing 1-13 of 13

Who am I?

- id:hnw
 - 細かすぎるツッコミ担当
- 株式会社ディノ
 - スポンサー！



Who am I?

- id:hnw
 - 細かすぎるツッコミ担当
- 株式会社ディノ
 - スポンサー！
- WEB+DB PRESS
 - 「PHP転ばぬ先の杖」連載中



Idea sketch

- Goal : FindBugs for PHP
 - find bugs for PHP program
- We need pure PHP parser against PHP.

Idea sketch

- lexer + parser +
event dispatcher(observer pattern)
- hooks all node of parse tree, so we
can use detailed information.

lexer/parser

- lexer(字句解析器)
 - splits source code to tokens.
- parser(構文解析器)
 - build parse tree from token sequence.

lexer/parser of PHP

- lexer
 - Zend/zend_language_scanner.l
- parser
 - Zend/zend_language_parser.y

lexer/parser emulation

- emulate PHP using pure PHP
- lexer
 - `token_get_all()`
- parser
 - KMyacc + PHP patch

KMyacc

- KMyacc is a YACC implementation.
<http://www005.upp.so-net.ne.jp/kmori/kmyacc/>
- It supports multi-languages, generating not only C parser, but also Javascript & Perl one.
- PHP patch by @bto & @moriyoshi.

How to parse PHP using PHP

- zend_language_parser.phpy
 - YACC source for php parser.
 - generated from original parser.y
 - adds hook point to all node with “Symfony Evnet Dispatcher”.

Conclusion

- KMyacc is great tool.
 - We can use pure PHP parser.
- @bto++
- @moriyoshi++

ご清聴

ありがとうございます

ございました