Reference Implementations	Test Data Repository	Usage	Reference Implementation Notes	testregex Notes

AT&T Labs Research regex(3) regression tests

Glenn Fowler <gsf@research.att.com>

AT&T Labs Research - Florham Park NJ

testregex.c 2004-05-31 is the latest source for the AT&T Labs Research regression test harness for the X/Open regex pattern match interface. See **testregex**(1) for option and test input details. The source and test data posted here are license free.

testregex can:

- verify stability for a particular implementation in the face of source code and/or compilation environment changes
- verify standard compliance for all implementations
- provide a basis for discussions on what *compliance* means

See An Interpretation of the POSIX regex Standards for an analysis of the POSIX-X/Open **regex** standards.

Reference Implementations

testregex is currently built against these reference implementations:

NAME	LABEL	AUTHORS
AT&T ast	A	Glenn Fowler and Doug McIlroy
bsd	B	
Bell Labs	D	Doug McIlroy
old gnu	G	
gnu	H	Isamu Hasegawa
irix	I	
boost	J	John Maddock
regex++	M	John Maddock
pcre perl compatible	P	Philip Hazel
rx	R	Tom Lord
spencer	S	Henry Spencer
libtre	T	Ville Laurikari
unix caldera	U	

Test Data Repository

basic.dat	basic regex(3) -- all implementations should pass these
categorize.dat	implementation categorization
nullsubexpr.dat	null (...)* tests
leftassoc.dat	left associative catenation implementation must pass these
rightassoc.dat	right associative catenation implementation must pass these
forcedassoc.dat	subexpression grouping to force associativity
repetition.dat	explicit vs. implicit repetitions

Usage

To run the **basic.dat** tests:

```
testregex < basic.dat
```

If the local implementation hangs or dumps on some tests then run with the **-c** option. The **-h** option lists the test data format details. The test data files exercise all features; the test harness detects and ignores features not supported by the local implementation.

Reference Implementation Notes

D: diet libc

The diet libc implementation is currently omitted because it fails all but one **basic.dat** test.

P: PCRE

The **P** implementation emulates **perl**(1) and is not X/Open compliant by design. The main differences are:

- **P** *leftmost-first* matching as opposed to the X/Open *leftmost-longest*.
- **REG_EXTENDED** patterns only.

However, the **P** package regression tests, and **perl**(1) features creeping into other implementations, make it reasonable to include here.

testregex Notes

Extensions to the standard terminology are derived from the AT&T implementation, unified under **<regex.h>** with these modes:

MODE	FLAGS	DESCRIPTION

BRE	0	basic RE
ERE	REG_EXTENDED	egrep RE with perl (...) extensions
ARE	REG_AUGMENTED	ERE with ! negation, <> word boundaries
SRE	REG_SHELL	sh patterns
KRE	REG_SHELL|REG_AUGMENTED	ksh93 patterns: ! @ (| &) { }
LRE	REG_LITERAL	fgrep patterns

and a few flags to handle **fnmatch**(3):

regex FLAG	fnmatch FLAG
REG_SHELL_ESCAPED	FNM_NOESCAPE
REG_SHELL_PATH	FNM_PATHNAME
REG_SHELL_DOT	FNM_PERIOD

The original `testregex.c` was done by Doug McIlroy at Bell Labs. The current implementation is maintained by Glenn Fowler <gsf@research.att.com>.

Glenn Fowler
Information and Software Systems Research
AT&T Labs Research
Florham Park NJ
January 05, 2006