The History of PCCTS

The Purdue Compiler-Construction Tool Set

Terence Parr
Parr Research Corporation
Minneapolis, Minnesota
and
University of Minnesota
Army High Performance Computing Research Center

[Updated 8-7-94]

The PCCTS project began as a parser-generator project for a graduate course at Purdue University in the Fall of 1988 taught by Hank Dietz— translator-writing systems. Under the guidance of Professor Dietz, the parser generator, ANTLR (originally called YUCC), continued after the termination of the course and eventually became the subject of Terence Parr's Master's thesis. Originally, lexical analysis was performed via ALX which was soon replaced by Will Cohen's DLG in the Fall of 1989 (DFA-based lexical-analyzer generator, also an offshoot of the graduate translation course).

The alpha version of ANTLR was totally rewritten resulting in 1.00B. Version 1.00B was released via an internet newsgroup (comp.compilers) posting in February of 1990 and quickly gathered a large following. 1.00B generated only LL(1) parsers, but allowed the merged description of lexical and syntactic analysis. It had rudimentary attribute handling similar to that of YACC and did not incorporate rule parameters or return values; downward inheritance was very awkward. 1.00B-generated parsers terminated upon the first syntax error. Lexical classes (modes) were not allowed and DLG did not have an interactive mode.

Upon starting his Ph.D. at Purdue in the Fall of 1990, Terence Parr began the second total rewrite of ANTLR. The method by which grammars may be practically analyzed to generate $LL(k)$ lookahead information was discovered in August of 1990 just before his return. Version 1.00 incorporated this algorithm and included the AST mechanism, lexical classes, error classes, and automatic error recovery; code quality and portability were higher. In February of 1992 1.00 was released via an article in SIGPLAN Notices. Peter Dahl, Ph.D. candidate, and Professor Matt O'Keefe (both at the University of Minnesota) tested this version extensively. Dana Hoggatt (Micro Data Base Systems, Inc.) came up with the idea of error grouping (strings attached to non-terminals) and tested 1.00 heavily.

Version 1.06 was released in December 1992 and represented a large feature enhancement over 1.00. For example, rudimentary semantic predicates were introduced, error messages were significantly improved for k>1 lookahead and ANTLR parsers could indicate that lookahead fetches were to occur only when necessary for the parse (normally, the lookahead "pipe" was constantly full). Russell Quong joined the project in the Spring of 1992 to aid in the semantic predicate design. Beginning and advanced tutorials were created and released as well. A make-file generator was included that sets up dependencies and such correctly for ANTLR and DLG.

Very few 1.00 incompatibilities were introduced (1.00 was quite different from 1.00B in some areas).

1.10 was released on August 31, 1993 and incorporated bug fixes, a few feature enhancements and a major new capability — an arbitrary lookahead operator (syntactic predicate), `(alpha)?beta`. This feature was co-designed with Professor Russell Quong also at Purdue. To support infinite lookahead, a preprocessor flag, ZZINF_LOOK, was created that forced the ANTLR() macro to tokenize all input prior to parsing. Hence, at any moment, an action or predicate can see the entire input sentence. The predicate mechanism of 1.06 was extended to allow multiple predicates to be hoisted; the syntactic context of a predicate was also moved along with the predicate.

In February of 1994, SORCERER (a simple tree-parser generator) was released. This tool allows the user to parse child-sibling trees by specifying a grammar rather than building a recursive-descent tree walker by hand. Work towards a library of tree transformations is underway. Aaron Sawdey at The University of Minnesota became a second author of SORCERER after the initial release.

On April 1, 1994, PCCTS 1.20 was released. This was the first version to actively support C++ output. It also included important fixes regarding semantic predicates and (..)+ subrules. This version also introduced token classes, the "*not*" operator, and token ranges.

On June 19, 1994, SORCERER 1.00B9 was released. Gary Funck of Intrepid Technology joined the SORCERER team and provided very valuable suggestions regarding the "transform" mode of SORCERER.

On August 8, 1994, PCCTS 1.21 was released. It mainly cleaned up the C++ output and included a number of bug fixes.

From the 1.21 release forward, the maintenance and support of all PCCTS tools will be primarily provided by Parr Research Corporation, Minneapolis MN---an organization founded on the principles of excellence in research and integrity in business; we are devoted to providing really cool software tools. Please see file PCCTS.FUTURE for more information. All PCCTS tools currently in the public domain will continue to be in the public domain.

Looking towards the future, a graphical user-interface is in the design phase. This would allow users to view the syntax diagram representation of their grammars and would highlight nondeterministic productions. Parsing can be traced graphically as well. This system will be built using a multiplatform window library. We also anticipate the introduction of a sophisticated error handling mechanism called "parser exception handling" in a near future release.

Currently, PCCTS is used at over 1000 known academic, government, and commercial sites in 37 countries. Of course, the true number of users is unknown due to the large number of ftp sites.

Credits

Idea/Tool	Coder	Co-designer(s)
ANTLR 1.00A	Terence Parr	Hank Dietz
ALX	Terence Parr	Hank Dietz
ANTLR 1.00B	Terence Parr	Hank Dietz, Will Cohen
DLG 1.00B	Will Cohen	Terence Parr, Hank Dietz
NFA Relabelling	Will Cohen	
$LL(k)$ analysis	Terence Parr	Hank Dietz
ANTLR 1.00	Terence Parr	Hank Dietz, Will Cohen
DLG 1.00	Will Cohen	Terence Parr, Hank Dietz
ANTLR 1.06	Terence Parr	Will Cohen, Russell Quong, Hank Dietz
DLG 1.06	Will Cohen	Terence Parr, Hank Dietz
ANTLR 1.10	Terence Parr	Will Cohen, Russell Quong
ANTLR 1.20	Terence Parr	Will Cohen, Russell Quong
ANTLR 1.21	Terence Parr	Russell Quong
DLG 1.10	Will Cohen	Terence Parr
DLG 1.20	Will Cohen	Terence Parr
DLG 1.21	Terence Parr	
Semantic predicates	Terence Parr	Russell Quonq
Syntactic predicates	Terence Parr	Russell Quonq
SORCERER 1.00A	Terence Parr	
SORCERER 1.00B	Terence Parr	Aaron Sawdey
SORCERER 1.00B9	Terence Parr	Aaron Sawdey, Gary Funck