

An Introduction to ANTLR

Clinton Roy

January 27, 2008

1 Introduction

- Humbug
- CSIRO
- Antlrworks

2 Overview

- ANother Tool for Language Recognition
- Lexer & Parser generator
- Substitute for lex and yacc
- BSD Licence
- Working towards a 3.1 release
- Reference book by Terence Parr
- My involvement

3 Typical Problems

- configuration file parsing
- syntax highlighting
- Domain Specific Languages
- interpreter
- translation
- transformation
- compiler frontend (backend?)

4 Users

- Hibernate
- IntelliJ
- Sun
- Eclipse
- ...

5 Features

- Generates easy to follow code
- Multiple generation targets
- LL(*) parsing algorithm
- Great error reporting and recovery
- Combined lexer & parser generator
- From hacky scripts to compilers
- Well designed runtime library: common operations are directly supported, less common operations are supported by sensible overriding points

6 Basic Parser Design

- Lexing and Parsing
- Parser profiling
- string equality vs integer equality

7 Parsing Algorithms

- Right to Left, bottom up.
- Left to Right, top down.
- Lookahead: $LL(k)$ vs $LL(*)$, C example

8 Inputs and Outputs

- Input
 - string
- Outputs
 - boolean (syntax validation)
 - token stream (filtering)
 - tree (semantic validation)
 - template (translation)
 - string (transformation)

9 Closing the Loop

- Input
 - tree
- Outputs
 - tree (optimisation)
 - template (translation)

10 Other Features

- Dynamic Scope support
- Island grammars
- Fuzzy matching

11 In the Next Version

- Tree rewriting
- Grammar importation
- AST error nodes

12 Problems

- Lexing idiosyncrasies
- Documentation