Parser Generators are Hard

Sander Wiig

May 19, 2023

Contents

1 Proof of man's Hubris

1.1 How we got here

I believe it was Mark Twain who once said, "*Challenges make life interesting*"; by that standard, my life has been very interesting.

The project's original goal was to create a metalanguage for defining other languages. The plan was to have an underlying semantics and then to let the user define a syntax and give bindings between that syntax and the underlying semantics. It turns out that this is somewhat difficult to achieve.
Therefore, the project goal was scaled back to just a parser generator that outputs a parser that can parse a CFG grammar and an AST for said grammar.

It turns out this also presents some challenges. An AST is very much just a parse tree but with all the superfluous information, such as parenthesis, semicolons, and other non-relevant information, removed so that all that remains is the information necessary to preserve the semantics and structure of the program. The problem then becomes, What information is relevant? Because of this, it was decided instead to output a parse tree rather than an AST.

1.2 The Parser Generator

The parser generator works by first parsing a `.ebnf` file. The `.ebnf` file contains a CFG grammar that is defined in a format inspired by EBNF. The EBNF parser outputs a AST representation of the grammar that is then fed to the code generator.

1.3 Program internals

The library consists of two sub-libraries,

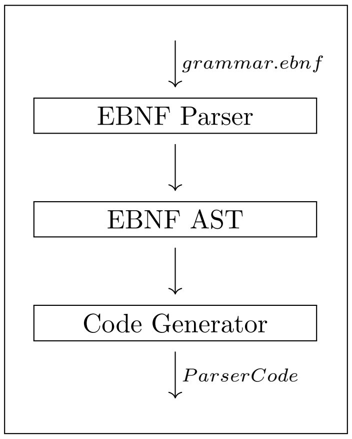

Figure 1: The Parser Generator Process

- **EBNF** - The `EBNF` library is responsible for parsing the user-defined grammar and creating an AST for it.
- **Generator** - The `Generator` takes an AST generated by the `EBNF` library and generates a Megaparsec parser that can parse a user-defined grammar.

EBNF

The `EBNF` library consists of the following files

- `EBNF.hs` - defines the AST for a grammar.
- `E`

2 Does it work?

2.1 "There are 56 consecutive parentheses"

2.2 Why it doesn't work

2.3 How to make it work

3 How to Use it