

The PREV Language Specification

Boštjan Slivnik

March 18, 2016

1 Lexical structure

The programs in the PREV programming language are written in 7-bit ASCII character set (all other characters are invalid). A single character LF denotes the end of line regardless of the text file format.

The (groups of) symbols of the PREV programming language are:

- **Symbols:**

`+ & = : , }]) . / == >= > <= < @ % * != ! { [(| - ^`

- **Constants:**

- *Integer constants:* An integer constant is a sequence of decimal digits optionally prefixed by a sign, i.e., “+” or “-”, denoting a 64-bit signed integer, i.e., from the interval $[-2^{63}, 2^{63} - 1]$.
- *Boolean constants:* `true false`
- *Character constants:* A character constant consists of a single character name within single quotes. A character name is either a character with an ASCII code from the interval $[32, 126]$ (but not a backslash, a single or a double quote) or an escape sequence. An escape sequence starts with a backslash character followed by a backslash (denoting a backslash), a single quote (denoting a single quote), a double quote (denoting a double quote), “`\t`” (denoting TAB), or “`\n`” (denoting LF).
- *String constants:* A string constant is a possibly empty finite sequence of character names within double quotes. A character name is either a character with an ASCII code from the interval $[32, 126]$ (but not a backslash, a single or a double quote) or an escape sequence. An escape sequence starts with a backslash character followed by a backslash (denoting a backslash), a single quote (denoting a single quote), a double quote (denoting a double quote), “`\t`” (denoting TAB), or “`\n`” (denoting LF).
- *Pointer constant:* `null`
- *Void constant:* `none`

- **Type names:**

`integer boolean char string void`

- **Keywords:**

`arr else end for fun if then ptr rec typ var where while`

- **Identifiers:** An identifier is a nonempty finite sequence of letters, digits and underscores that starts with a letter or an underscore.

Additionally, the source might include the following:

- **White space:** Characters CR, LF, TAB, or space.
- **Comments:** A comment is a sequence of character that starts with an octothorpe character, i.e., “#”, and ends with the LF character (regardless of the text file format).

To break the source file into individual symbols, the first-match-longest-match rule must be used.

2 Syntactic structure

The concrete syntax of the PREV programming language is defined by an LR(1) grammar. Nonterminal and terminal symbols are written in *italic* and *typewriter* fonts, respectively. Terminal symbols `IDENTIFIER`, `INTEGER`, `BOOLEAN`, `CHAR` and `STRING` denote (all) identifiers, integer constants, boolean constants, character constants and string constants, respectively. The start symbol of the grammar is *Program*. The LR(1) grammar contains the following productions:

Note that the LR(1) grammar generates certain sentential forms which are prohibited by semantics.