

Introduction to language theory and compiling

Project - Part 1

Université Libre de Bruxelles

Hakim Boulahya

October 25, 2017

1 Definition

- MUST
- OR
- CAN
- case-sensitive characters
- alpha numeric

2 Extended regular expressions

This section describes extended regular expressions used by the lexer. The shortcuts section describes simple EREs that are used to write the advanced ERE.

2.1 Shortcuts

1. `AlphaUpperCase` = `[A-Z]` : Match all uppercase letters
2. `AlphaLowerCase` = `[a-z]` : Match all lowercase letters
3. `Alpha` = `{AlphaUpperCase}|{AlphaLowerCase}` : Match alphabet case-sensitive characters
4. `Numeric` = `[0-9]` : Match digit characters
5. `AlphaNumeric` = `{Alpha}|{Numeric}` : Match alphabetic **OR** digit characters

2.2 Advanced

1. `Number = ([1-9]Numeric*)|0` : Match all numbers that **MUST** start with a non-zero digit and follow by 0 or more digits (first RE), **OR** match only zero digit (second RE). This RE correspond to the `[Number]` lexical unit.
2. `VarName = AlphaAlphaNumeric*` Match identifiers. Identifiers **MUST** starts with a alphabetic character and **CAN** be followed by 0 or more digits **OR** alphabetic characters. Correspond to the `[VarName]` lexical unit.
3. `Spaces =`
`s+`

3 All in yyinitial vs states for different grammars