

어휘분석기 구현

1891179 김현학

Character Set

ASCII Printable Characters

(character code 32-127)

Character Definition

- **Letter**

- A(65) - Z(90)
- a(97) - z(122)
- '-'(45)
- '.'(46)
- '_'(95)

- **Digit**

- 0(65) - 9(90)

Character Definition

- **Delimiter (Special Symbols)**

- @ : / ? #
- ! \$ & =

- **Ignore**

- ASCII Printable Characters
(character code 32-127)
except all above

Token Definition

SPECIAL FORM

- KEYWORD, DELIMITER (GEN, SUB)

Token definition - SPECIAL FORM

- **KEYWORD**

- **HTTP** >> **TOKEN_HTTP**
- **HTTPS** >> **TOKEN_HTTPS**
- **FILE** >> **TOKEN_FILE**
- **MAILTO** >> **TOKEN_MAILTO**

Default: Save as Lower case

Token definition - SPECIAL FORM

- DELIMITER - GEN

- @ >> TOKEN_UserInfo
- : >> TOKEN_Detail
- / >> TOKEN_Path
- ? >> TOKEN_Query
- # >> TOKEN_Fragment

Token definition - SPECIAL FORM

- DELIMITER - SUB

- ! >> TOKEN_Exclamation
- \$ >> TOKEN_Dollar
- & >> TOKEN_Ampersand
- = >> TOKEN_Equal

Token Definition

GENERAL FORM

- STRING, DIGIT

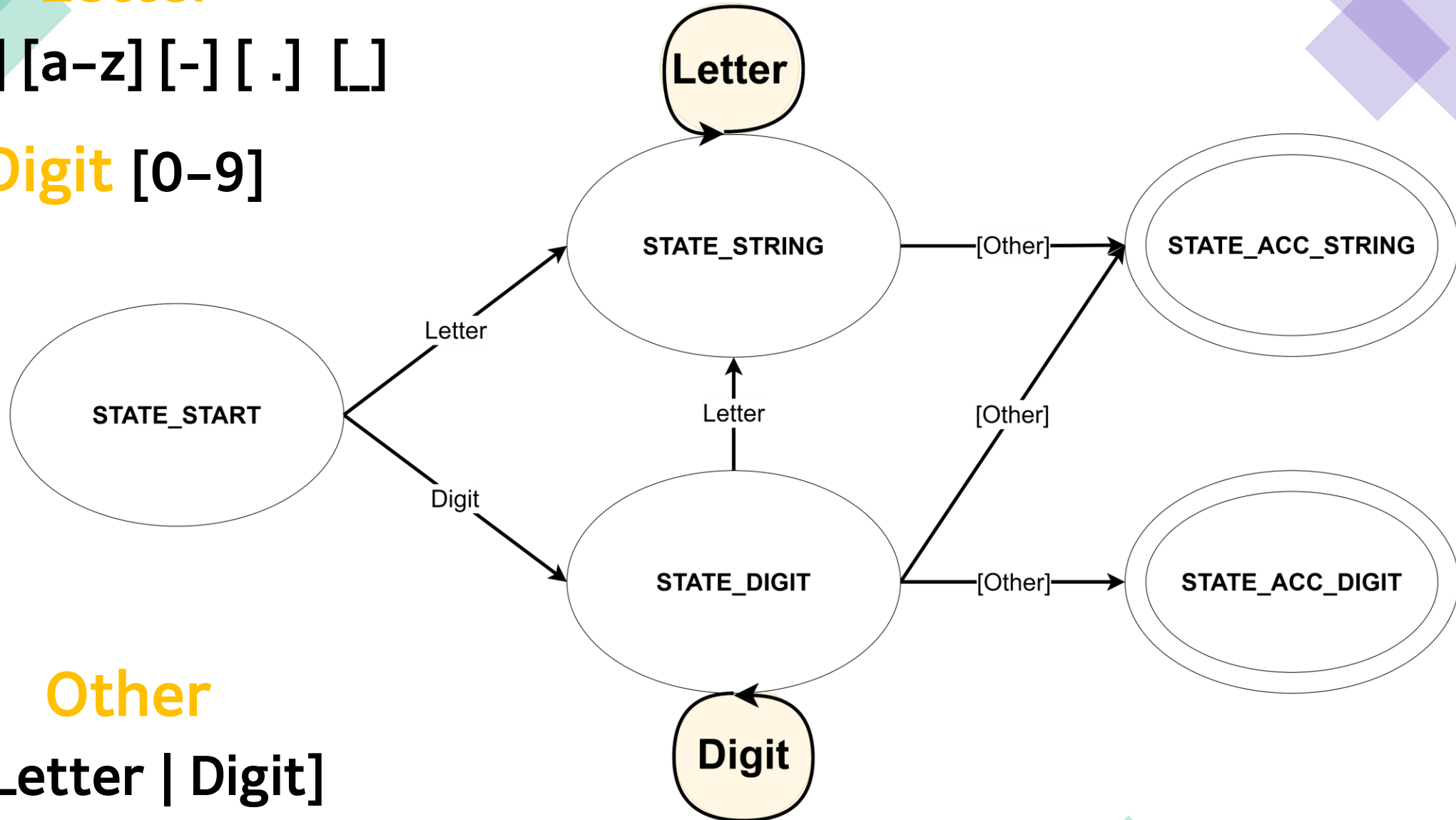
Letter

[A-Z][a-z][-][.][_]

Digit [0-9]

Other

~[Letter | Digit]



State

Naming & Numbering

```
Enum STATE_NUMBER
```

```
Char* stateName [ ]
```

STATE

- STATE_ERROR = -1,
- STATE_START = 0,
- STATE_STRING,
- STATE_DIGIT,
- STATE_ACC_STRING = 10,
- STATE_ACC_DIGIT = 11,
- STATE_ACC_UserInfo = '@',
- STATE_ACC_Detail = ':',
- STATE_ACC_Path = '/',
- STATE_ACC_Query = '?',
- STATE_ACC_Fragment = '#',
- STATE_ACC_Exclamation = '!',
- STATE_ACC_Dollar = '\$',
- STATE_ACC_Ampersand = '&',
- STATE_ACC_Equal = '='

Token

Naming & Numbering

Enum TOKEN_NUMBER

Char* tokenName []

TOKEN

- TOKEN_UNDEFINED = -1,
- TOKEN_STRING = 0,
- TOKEN_DIGIT = 1
- TOKEN_HTTP = 10,
- TOKEN_HTTPS = 11,
- TOKEN_FILE = 12,
- TOKEN_MAILTO = 13,
- TOKEN_UserInfo = '@',
- TOKEN_Detail = ':',
- TOKEN_Path = '/',
- TOKEN_Query = '?',
- TOKEN_Fragment = '#',
- TOKEN_Exclamation = '!',
- TOKEN_Dollar = '\$',
- TOKEN_Ampersand = '&',
- TOKEN_Equal = '='