Educational Compiler Project — COMPO
Grammar correction

Tanguy BERTHOUD & Yohann MARTIN

December 13, 2022

Institut Supérieur d'Électronique de Paris

1 Grammar

All words in uppercase are the tokens returned by the lexical analyzer (scanner).
All words in lowercase are the variables managed by the syntactic analyzer (parser). The document variable is the starting symbol of the grammar.

```
document:
    block_list


block_list:
    block BLANK_LINE block_list
    | block


block:
    H1 TEXT
    | H2 TEXT
    | H3 TEXT
    | H4 TEXT
    | H5 TEXT
    | H6 TEXT
    | HR
    | BLOCK_CODE NEWLINE TEXT NEWLINE BLOCK_CODE
    | paragraph
    | quote
    | svg


text:
    TEXT
    | BOLD text BOLD
    | ITALIC text ITALIC
    | UNDERLINE text UNDERLINE
    | STRIKETHROUGH text STRIKETHROUGH
    | INLINE_CODE text INLINE_CODE
    | LBRACKET text RBRACKET LPAREN TEXT RPAREN
    | EXCLAM LBRACKET text RBRACKET LPAREN TEXT RPAREN


line:
    text line
    | text


paragraph:
    line NEWLINE paragraph
    | line
```

```
quote:
    QUOTE line NEWLINE quote
```

1.1   SVG extension

```
svg:
    XSVG_BEGIN svg_coord COMMA svg_coord NEWLINE svg_instruction_list
    XSVG_END


svg_instruction_list:
    svg_instruction NEWLINE svg_instruction_list
    | ε


svg_instruction:
    LINE svg_coord svg_coord svg_attribute
    | POLYLINE svg_coord_list svg_attribute
    | POLYGON svg_coord_list svg_attribute svg_attribute
    | CIRCLE svg_coord NUMBER svg_attribute svg_attribute
    | ELLIPSE svg_coord NUMBER NUMBER svg_attribute svg_attribute
    | RECT svg_coord NUMBER NUMBER svg_attribute svg_attribute
    | XSVG_TEXT svg_coord STR svg_attribute svg_attribute
   svg_attribute


svg_coord:
    NUMBER COMMA NUMBER


svg_coord_list:
    svg_coord svg_coord_list
    | svg_coord


svg_attribute:
    STR
    | ε
```

2 Tokens

Here is the correspondence of the tokens to their value in markdown:

- BLANK_LINE : \n\n
- H1 : #
- H2 : ##
- H3 : ###
- H4 : ####
- H5 : #####
- H6 : ######
- TEXT : anything that is not from the other tokens are considered as text
- HR : - - -
- BLOCK_CODE : ```
- NEWLINE : \n
- BOLD : **
- ITALIC : *
- UNDERLINE : _ _
- STRIKETHROUGH : ~~
- INLINE_CODE : `
- LBRACKET : [
- RBRACKET :]
- LPAREN : (
- RPAREN :)
- EXCLAM : !
- QUOTE : >
- XSVG_BEGIN : ```xsvg:
- COMMA : ,
- XSVG_END : ```
- LINE : line
- POLYLINE : polyline
- CIRCLE : circle

- ELLIPSE : ellipse

- RECT : rect

- XSVG_TEXT : text

- NUMBER : -?[0-9]+