# CS 4240 Phase 1

## David Zhang

#### February 23, 2014

## 1 Lexical Rules

Pre-parse the input to remove comments with the following DFA. The resulting NOT-COMMENT tokens are concatenated together.

Everything under the token column represents either generated tokens or custom table actions on the character buffer used to produce the text associated with tokens.

| start state  | symbol                      | next state   | token       |
|--------------|-----------------------------|--------------|-------------|
| START        | $\Sigma - \{/, "\}$         | START        | NOT-COMMENT |
| START        | "                           | STRING       |             |
| START        | /                           | SLASH        |             |
| STRING       | $\Sigma - \{\setminus, "\}$ | STRING       |             |
| STRING       |                             | STRING-SLASH |             |
| STRING       | ,,                          | START        | NOT-COMMENT |
| STRING-SLASH | $\Sigma$                    | STRING       |             |
| SLASH        | $\Sigma - \{*\}$            | START        | NOT-COMMENT |
| SLASH        | *                           | COMMENT      |             |
| COMMENT      | $\Sigma - \{*\}$            | COMMENT      |             |
| COMMENT      | *                           | COMMENT-END  |             |
| COMMENT-END  | $\Sigma - \{*,/\}$          | COMMENT      |             |
| COMMENT-END  | *                           | COMMENT-END  |             |
| COMMENT-END  | /                           | START        | COMMENT     |

The DFA for uncommented code.

Note, any time backtracking is mentioned, it essentially is the same as treating the current state as the start state and doing the corresponding transitions or token generations. This is included to simplify the table by removing duplication of the start state transitions.

Note, all ids are later matched character by character with keywords to determine if they are keywords.

Note, drop character for the error state means that the last read character is ignored, and the state remains unchanged.

| start state | symbol | next state | token  |
|-------------|--------|------------|--------|
| START       | +      | START      | PLUS   |
| START       | -      | START      | MIN    |
| START       | *      | START      | MULT   |
| START       | /      | START      | DIV    |
| START       | =      | START      | EQ     |
| START       |        | START      | LPAREN |
| START       |        | START      | RPAREN |
| START       | ,      | START      | COMMA  |

| START             | &                              | START             | AND                 |
|-------------------|--------------------------------|-------------------|---------------------|
| START             | _                              | START             | OR                  |
| START             |                                | START             | LSQUARE             |
| START             |                                | START             | RSQUARE             |
| START             | :                              | START             | SEMI                |
| START             | (                              | LANGLE            |                     |
| START             | \<br>\<br>\                    | RANGLE            |                     |
| START             | :                              | COLON             |                     |
| START             | 0-9                            | INT-LIT           |                     |
| START             | "                              | STRING-LIT        |                     |
| START             | a-zA-Z                         | ID                |                     |
| START             | whitespace                     | START             | ignore              |
| START             | others                         | ERROR             | drop character      |
| LANGLE            | $\Sigma - \{=, \rangle\}$      | START             | LESS, backtrack     |
| LANGLE            | >                              | START             | NOTEQ               |
| LANGLE            | =                              | START             | LESSEQ              |
| RANGLE            | $\Sigma - \{=\}$               | START             | GREATER, backtrack  |
| RANGLE            |                                | START             | GREATEREQ           |
| COLON             | =                              | START             | ASSIGN              |
| COLON             | $\Sigma - \{=\}$               | START             | COLON, backtrack    |
| INT-LIT           | 0-9                            | INT-LIT           |                     |
| INT-LIT           | $\Sigma - 0 - 9$               | START             | INT-LIT, backtrack  |
| ID                | $a$ -z $A$ - $Z0$ - $9$ _      | ID                |                     |
| ID                | $\Sigma - a - zA - Z0 - 9_{-}$ | START             | ID, backtrack       |
| STRING-LIT        | $\Sigma - \setminus$           | STRING-LIT        |                     |
| STRING-LIT        | "                              | START             | STRING-LIT          |
| STRING-LIT        | \                              | STRING-LIT-SLASH  |                     |
| STRING-LIT-SLASH  | n                              | STRING-LIT        |                     |
| STRING-LIT-SLASH  | t                              | STRING-LIT        |                     |
| STRING-LIT-SLASH  | "                              | STRING-LIT        |                     |
| STRING-LIT-SLASH  |                                | STRING-LIT        |                     |
| STRING-LIT-SLASH  | ^                              | STRING-LIT-CTL    |                     |
| STRING-LIT-SLASH  | 0-9                            | STRING-LIT-CODE-1 |                     |
| STRING-LIT-SLASH  | whitespace                     | STRING-LIT-SPACE  | ignore 2 characters |
| STRING-LIT-SLASH  | others                         | ERROR             | drop character      |
| STRING-LIT-CTL    | @A-Z[\]^                       | STRING-LIT        |                     |
| STRING-LIT-CTL    | others                         | ERROR             | drop character      |
| STRING-LIT-CODE-1 | 0-9                            | STRING-LIT-CODE-2 |                     |
| STRING-LIT-CODE-1 | others                         | ERROR             | drop character      |
| STRING-LIT-CODE-2 | 0-9                            | STRING-LIT        |                     |
| STRING-LIT-CODE-2 | others                         | ERROR             | drop character      |
| STRING-LIT-SPACE  | whitespace                     | STRING-LIT-SPACE  | ignore              |
| STRING-LIT-SPACE  |                                | STRING-LIT        | ignore              |
| STRING-LIT-SPACE  | others                         | ERROR             | drop character      |
|                   |                                |                   |                     |

# 2 Grammar Rules

Raw Grammar:

| symbol                                       | rule   |
|--|--|
| $\langle \text{tiger-program} \rangle$       | let (declaration-segment) in (stat-seq) end  |
| $\langle declaration\text{-segment} \rangle$ | $\langle \text{type-declaration-list} \rangle \langle \text{var-declaration-list} \rangle \langle \text{funct-declaration-list} \rangle$ |

```
⟨type-declaration-list⟩
                                         ⟨type-declaration⟩ ⟨type-declaration-list⟩
⟨type-declaration-list⟩
                                         NULL
⟨var-declaration-list⟩
                                         (var-declaration) (var-declaration-list)
⟨var-declaration-list⟩
                                         NULL
\( \) funct-declaration-list \( \)
                                         \(\langle \text{funct-declaration} \rangle \text{funct-declaration-list} \rangle \)
                                         NULL
\( \) funct-declaration-list \( \)
⟨type-declaration⟩
                                         type id = \langle \text{type} \rangle;
\langle type \rangle
                                         ⟨type-id⟩
                                         array [ INTLIT ] of \( \text{type-dim} \) of \( \text{type-id} \)
\langle type \rangle
⟨type-dim⟩
                                         [ INTLIT ] (type-dim)
⟨type-dim⟩
                                         NULL
\langle \text{type-id} \rangle
                                         int
⟨type-id⟩
                                         string
\langle \text{type-id} \rangle
                                         id
(var-declaration)
                                         var (id-list) : (type-id) (optional-init);
\langle id-list \rangle
\langle id-list \rangle
                                         id, \langle id-list \rangle
⟨optional-init⟩
                                         NULL
(optional-init)
                                         := \langle const \rangle
\(\rangle \text{funct-declaration} \rangle \)
                                         function id ( \( \rangle \text{param-list} \rangle ) \( \rangle \text{ret-type} \rangle \text{ begin \( \stat-\text{seq} \rangle \text{ end} \);
(param-list)
                                         NULL
(param-list)
                                         ⟨param⟩ ⟨param-list-tail⟩
⟨param-list-tail⟩
                                         NULL
                                         , (param) (param-list-tail)
⟨param-list-tail⟩
                                         NULL
⟨ret-type⟩
⟨ret-type⟩
                                         : (type-id)
                                         id: \langle type-id \rangle
(param)
                                         ⟨stat⟩ ⟨stat-seq⟩
(stat-seq)
\langle \text{stat-seq} \rangle
                                         \langle stat \rangle
\langle stat \rangle
                                         \langle \text{lvalue} \rangle := \langle \text{expr} \rangle ;
\langle stat \rangle
                                         if \langle \exp r \rangle then \langle \operatorname{stat-seq} \rangle endif;
                                         if \langle \exp r \rangle then \langle \operatorname{stat-seq} \rangle else \langle \operatorname{stat-seq} \rangle endif;
\langle stat \rangle
                                         while \langle \exp r \rangle do \langle \operatorname{stat-seq} \rangle enddo;
\langle stat \rangle
                                         for id := \langle \exp r \rangle to \langle \exp r \rangle do \langle \text{stat-seq} \rangle enddo;
\langle stat \rangle
                                         \langle \text{opt-prefix} \rangle \text{ id } (\langle \text{expr-list} \rangle);
\langle stat \rangle
\langle stat \rangle
                                         break;
                                         return (expr);
\langle stat \rangle
                                         ⟨expr⟩ ⟨binary-operator⟩ ⟨expr⟩
\langle \exp r \rangle
\langle \exp r \rangle
                                         \langle const \rangle
                                         ⟨lvalue⟩
\langle \exp r \rangle
                                         - \langle \exp r \rangle
\langle \exp r \rangle
                                         (\langle \exp r \rangle)
\langle \exp r \rangle
(binary-operator)
(binary-operator)
(binary-operator)
                                         +
(binary-operator)
(binary-operator)
                                         =
(binary-operator)
                                         <
(binary-operator)
                                         >
(binary-operator)
                                         <=
(binary-operator)
                                         >=
(binary-operator)
                                         <>
(binary-operator)
                                         &
```

```
\langle \text{binary-operator} \rangle
(binary-operator)
                                                 :=
\langle \text{opt-prefix} \rangle
                                                 \langle lvalue \rangle :=
(opt-prefix)
                                                 NULL
\langle const \rangle
                                                 INTLIT
\langle const \rangle
                                                 STRLIT
\langle const \rangle
                                                 _{\mathrm{nil}}
                                                 \langle \exp r \rangle \langle \exp r - \operatorname{list-tail} \rangle
⟨expr-list⟩
\langle \text{expr-list} \rangle
                                                 NULL
                                                 , \langle \exp r \rangle \langle \exp r-list-tail\rangle
\langle \text{expr-list-tail} \rangle
\langle \text{expr-list-tail} \rangle
                                                 NULL
                                                 id \langle lvalue-tail \rangle
(lvalue)
                                                 [\langle \expr \rangle] \langle \text{lvalue-tail} \rangle
(lvalue-tail)
                                                 NULL
(lvalue-tail)
```

#### Modified Grammar:

| symbol   | rule   |  |
|--|--|--|
| $\langle \text{tiger-program} \rangle$                 | let (declaration-segment) in (stat-seq) end  |  |
| $\langle declaration\text{-segment} \rangle$           | \langle \text{type-declaration-list} \langle \text{funct-declaration-list} \rangle \text{funct-declaration-list} \rangle |  |
| $\langle \text{type-declaration-list} \rangle$         | $\langle \text{type-declaration} \rangle \langle \text{type-declaration-list} \rangle$                                   |  |
| $\langle \text{type-declaration-list} \rangle$         | NULL   |  |
| $\langle var-declaration-list \rangle$                 | (var-declaration) (var-declaration-list)   |  |
| $\langle var-declaration-list \rangle$                 | NULL   |  |
| $\langle \text{funct-declaration-list} \rangle$        | $\langle \text{funct-declaration} \rangle \langle \text{funct-declaration-list} \rangle$                                 |  |
| $\langle funct\text{-}declaration\text{-}list \rangle$ | NULL   |  |
| $\langle \text{type-declaration} \rangle$              | type $id = \langle type \rangle$ ;   |  |
| $\langle var\text{-declaration} \rangle$               | $  var \langle id-list \rangle : \langle type-id \rangle \langle optional-init \rangle ;$                                |  |
| $\langle \text{funct-declaration} \rangle$             | function id ( $\langle param-list \rangle$ ) $\langle ret-type \rangle$ begin $\langle stat-seq \rangle$ end;            |  |
| $\langle \mathrm{type} \rangle$                        | $\langle \text{type-id} \rangle$   |  |
| $\langle \mathrm{type} \rangle$                        | array [ INTLIT ] of $\langle \text{type-dim} \rangle$ of $\langle \text{type-id} \rangle$                                |  |
| $\langle \mathrm{type\text{-}dim} \rangle$             | $[INTLIT] \langle type-dim \rangle$  |  |
| $\langle \mathrm{type\text{-}dim} \rangle$             | NULL   |  |
| $\langle \mathrm{type\text{-}id} \rangle$              | id   |  |
| $\langle \mathrm{id}\text{-list} \rangle$              | id (id-list-tail)  |  |
| $\langle id$ -list-tail $\rangle$                      | $\mid$ , id $\langle$ id-list-tail $\rangle$   |  |
| $\langle \mathrm{id}	ext{-list-tail} \rangle$          | NULL   |  |
| $\langle { m optional\text{-}init}  angle$             | $ :=\langle \mathrm{const} \rangle$  |  |
| $\langle { m optional\text{-}init}  angle$             | NULL   |  |
| $\langle \mathrm{param\text{-}list} \rangle$           | $\langle param \rangle \langle param-list-tail \rangle$  |  |
| $\langle param-list \rangle$                           | NULL   |  |
| $\langle param-list-tail \rangle$                      | $\mid , \langle param \rangle \langle param-list-tail \rangle$   |  |
| $\langle param-list-tail \rangle$                      | NULL   |  |
| $\langle { m ret-type}  angle$                         | $ :\langle 	ext{type-id} \rangle$  |  |
| $\langle { m ret-type} \rangle$                        | NULL   |  |
| $\langle \mathrm{param} \rangle$                       | $  id : \langle type-id \rangle$   |  |
| $\langle \text{stat-seq} \rangle$                      | $\langle \text{stat} \rangle \langle \text{stat-seq-tail} \rangle$   |  |
| $\langle \text{stat-seq-tail} \rangle$                 | $\langle \text{stat} \rangle \langle \text{stat-seq-tail} \rangle$   |  |
| $\langle \text{stat-seq-tail} \rangle$                 | NULL   |  |
| $\langle \mathrm{stat} \rangle$                        | $  if \langle expr \rangle then \langle stat-seq \rangle \langle stat-if-tail \rangle$                                   |  |
| $\langle \mathrm{stat} \rangle$                        | while $\langle \exp r \rangle$ do $\langle \operatorname{stat-seq} \rangle$ enddo;                                       |  |
| $\langle \mathrm{stat} \rangle$                        | for id := $\langle \exp r \rangle$ to $\langle \exp r \rangle$ do $\langle \operatorname{stat-seq} \rangle$ enddo;       |  |
| $\langle \mathrm{stat}  angle$                         | break;   |  |

```
\langle stat \rangle
                                     return (expr);
\langle stat \rangle
                                     id (stat-func-or-assign)
⟨stat-func-or-assign⟩
                                     (\langle expr-list \rangle);
                                      \langle \text{lvalue-tail} \rangle := \langle \text{stat-assign} \rangle
⟨stat-func-or-assign⟩
⟨stat-if-tail⟩
                                     else (stat-seq) endif;
⟨stat-if-tail⟩
                                     endif;
(stat-assign)
                                     - (unaryminus) (stat-assign-tail)
                                     (\langle \exp r \rangle) \langle \operatorname{stat-assign-tail} \rangle
(stat-assign)
                                      \langle const \rangle \langle stat-assign-tail \rangle
⟨stat-assign⟩
(stat-assign)
                                     id \langle stat-assign-id \rangle
                                      (\langle expr-list \rangle);
(stat-assign-id)
                                      \langle lvalue-tail \rangle \langle stat-assign-tail \rangle
⟨stat-assign-id⟩
(stat-assign-tail)
                                      (expr-tail)
⟨stat-assign-tail⟩
                                      (orexpr-tail)
⟨stat-assign-tail⟩
                                      (andexpr-tail)
⟨stat-assign-tail⟩
                                      ⟨compare-tail⟩
⟨stat-assign-tail⟩
                                      ⟨term-tail⟩
\langle \exp r \rangle
                                      (orexpr) (expr-tail)
                                      ⟨orop⟩ ⟨orexpr⟩ ⟨expr-tail⟩
⟨expr-tail⟩
⟨expr-tail⟩
                                     NULL
                                      ⟨andexpr⟩ ⟨orexpr-tail⟩
\langle orexpr \rangle
                                      \langle andop \rangle \langle andexpr \rangle \langle orexpr-tail \rangle
(orexpr-tail)
(orexpr-tail)
                                     NULL
                                      ⟨compare⟩ ⟨andexpr-tail⟩
\langle and expr \rangle
\langle and expr-tail \rangle
                                      ⟨compop⟩ ⟨compare⟩ ⟨andexpr-tail⟩
(andexpr-tail)
                                     NULL
\langle compare \rangle
                                      ⟨term⟩ ⟨compare-tail⟩
                                      ⟨addop⟩ ⟨term⟩ ⟨compare-tail⟩
(compare-tail)
\langle compare-tail \rangle
                                     NULL
                                      ⟨factor⟩ ⟨term-tail⟩
\langle \text{term} \rangle
⟨term-tail⟩
                                      ⟨mulop⟩ ⟨factor⟩ ⟨term-tail⟩
⟨term-tail⟩
                                     NULL
                                      (unaryminus)
(factor)
⟨factor⟩
                                     - (unaryminus)
(unaryminus)
                                     (\langle \exp r \rangle)
⟨unaryminus⟩
                                      \langle const \rangle
(unaryminus)
                                      (lvalue)
                                     INTLIT
\langle const \rangle
\langle const \rangle
                                     STRLIT
\langle const \rangle
                                     nil
\langle \text{orop} \rangle
\langle andop \rangle
                                     &
\langle compop \rangle
\langle compop \rangle
                                      <>
\langle compop \rangle
                                      >
\langle compop \rangle
                                      <
\langle compop \rangle
                                      >=
\langle compop \rangle
                                      <=
(addop)
                                      +
\langle addop \rangle
                                      *
\langle \text{mulop} \rangle
\langle \text{mulop} \rangle
                                     ⟨expr⟩ ⟨expr-list-tail⟩
\langle \text{expr-list} \rangle
```

| $\langle \text{expr-list} \rangle$      | NULL   |
|---|--|
| $\langle \text{expr-list-tail} \rangle$ | , $\langle \exp r \rangle \langle \exp r - \operatorname{list-tail} \rangle$ |
| $\langle \text{expr-list-tail} \rangle$ | NULL   |
| $\langle lvalue \rangle$                | id $\langle lvalue-tail \rangle$   |
| (lvalue-tail)                           | $[\langle \expr \rangle] \langle \text{lvalue-tail} \rangle$                 |
| (lvalue-tail)                           | NULL   |
|   |  |

| non-terminal                            | first set   |
|---|---|
| $\langle lvalue-tail \rangle$           | $[,\epsilon$  |
| $\langle lvalue \rangle$                | id  |
| $\langle \text{expr-list-tail} \rangle$ | $ \; ,, \; \epsilon  $  |
| $\langle \text{expr-list} \rangle$      | $ \operatorname{first}(\langle \operatorname{expr} \rangle), \epsilon $ |
| $\langle \mathrm{mulop} \rangle$        | *, /  |
| $\langle \mathrm{addop} \rangle$        | +, -  |
| $\langle \text{compop} \rangle$         | =,<,>,<=,>=,<>  |
| $\langle andop \rangle$                 | &   |
| $\langle \text{orop} \rangle$           |   |
| $\langle \text{const} \rangle$          | nil, STRLIT, INTLIT   |
| (unaryminus)                            | (, nil, STRLIT, INTLIT, id  |
| (factor)                                | (, nil, STRLIT, INTLIT, id, -   |
| (term-tail)                             | $ *,/,\epsilon $  |
| $\langle \text{term} \rangle$           | (, nil, STRLIT, INTLIT, id, -   |
| $\langle \text{compare-tail} \rangle$   | $+, -, \epsilon$  |
| $\langle \text{compare} \rangle$        | (, nil, STRLIT, INTLIT, id, -   |
| (andexpr-tail)                          | =,<,>,<=,>=,<   |
| $\langle andexpr \rangle$               | (, nil, STRLIT, INTLIT, id, -   |
| (orexpr)                                | (, nil, STRLIT, INTLIT, id, -   |
| (orexpr-tail)                           | $ \&,\epsilon $   |
| $\langle \exp r \rangle$                | (, nil, STRLIT, INTLIT, id, -   |
| (expr-tail)                             | $ \cdot $ , $\epsilon$  |
| $\langle \text{stat-assign} \rangle$    | id, -, (, nil, STRLIT, INTLIT   |
| $\langle \text{stat-assign-id} \rangle$ | $[, (, *, /, +, -, =, <, >, <=, >=, <>, \&,  , \epsilon]$               |
| (stat-assign-tail)                      | *, / , +, -, =, <, >, <=, >=, <>, & ,  , \( \epsilon \)                 |
| (stat-if-tail)                          | else, endif   |
| (stat-func-or-assign)                   | (,:=,[  |
| $\langle \text{stat} \rangle$           | if, while, for, break, return, id                                       |
| $\langle \text{stat-seq} \rangle$       | if, while, for, break, return, id                                       |
| (stat-seq-tail)                         | if, while, for, break, return, id, $\epsilon$                           |
| $\langle param \rangle$                 | id  |
| $\langle \text{ret-type} \rangle$       | $ :,\epsilon $  |
| (param-list-tail)                       | $ \cdot $ ,, $\epsilon$   |
| (param-list)                            | $id, \epsilon$  |
| (optional-init)                         | $:=,\epsilon$   |
| (id-list-tail)                          | $ \cdot $ ,, $\epsilon$   |
| $\langle id$ -list $\rangle$            | id  |
| $\langle \text{type-id} \rangle$        | id  |
| $\langle \text{type-dim} \rangle$       | $[,\epsilon]$   |
| $\langle \text{type} \rangle$           | array, id   |
| (funct-declaration)                     | function  |
| (var-declaration)                       | var   |
| (type-declaration)                      | type  |
| (funct-declaration-list)                | function, $\epsilon$  |
| (var-declaration-list)                  | $  var, \epsilon  $   |
| . ,                                     | 1   |

```
 \begin{array}{c|c} \langle \text{type-declaration-list} \rangle & \text{type, } \epsilon \\ \langle \text{declaration-segment} \rangle & \text{function, var, type, } \epsilon \\ \langle \text{tiger-program} \rangle & \text{let} \\ \end{array}
```

Note, the dollar sign is the end of tokens symbol.

```
non-terminal
                            follow set
                            :=, *, /, +, -, =, <, >, <=, >=, <>, &
(lvalue-tail)
                            |, if, while, for, break, return, id, endif, end, & ,|, ), ,, , ], then, do, to, ;
(lvalue-tail)
(expr-list-tail)
\langle \text{expr-list} \rangle
⟨expr-tail⟩
                            if, while, for, break, return, id, endif, end, ), ,, , ], then, do, to, ;
                            if, while, for, break, return, id, endif, end, [, ), ,, , ], then, do, to, ;
(orexpr-tail)
(andexpr-tail)
                            if, while, for, break, return, id, endif, end, & ,|, ), ,, ,|, then, do, to, ;
                            if, while, for, break, return, id, endif, end
(compare-tail)
                            & ,|, ), ,, ,], then, do, to, ;, =, <, >, <=, >=, <>
(compare-tail)
(term-tail)
                            if, while, for, break, return, id, endif, end, & ,|, )
                            ,,,], then, do, to, ;,=,<,>,<=,>=,<>,+,-
⟨term-tail⟩
(stat-assign-tail)
                            if, while, for, break, return, id, endif, end
(stat-assign-id)
                            if, while, for, break, return, id, endif, end
(stat-seq-tail)
                            endif, end
⟨ret-type⟩
                            begin
(param-list-tail)
(param-list)
(optional-init)
                            function, in
(id-list-tail)
⟨type-dim⟩
                            of
\( \) funct-declaration-list \( \)
                            in
⟨var-declaration-list⟩
                            function, in
⟨type-declaration-list⟩
                            var, function, in
(declaration-segment)
                            in
```

Note, if there is no corresponding rule, then that means a parser error is generated.

Note, errors are handled by dropping tokens until a valid token is found.

symbol | next token | rule