



Scanner: generic scanners and direct coding

Appendix to Lecture 2 (4), pages 56-59

Christoph Kessler, IDA,
Linköpings universitet, 2008.

Generic scanner, interpreting the table



state	digit	blank	other	accept
0	1	0	E	0
1	1	2	E	0
2	-	-	-	1
E	E	0	E	0

// global data structures:

int table [Nstates][Nchars]
= ... (read in or initialize)

```
typedef struct {
    int tokentype;
    union {
        int ival; float fval; double dval; ...
        symboltable *stptr;
    } tokenval;
} *Token;
```

// scanner routine, called from parser:

```
Token getNextToken( void )
{
    Token t = new_Token();
    int state = 0;
    while ( 1 ) {
        char ch = getc ( inputfile );
        int oldstate = state;
        state = table [ state ][ ch ]; // transition
        // update t->tokenval with ch as appropriate
        accumulate ( ch, state, t );
        if ( is_error_state( state ) )
            error_handler( oldstate, ch, ... );
        else if ( is_accepting_state( state ) ) {
            t->tokentype = tokentype( state );
            return t;
        }
    }
}
```

TDOB44 / TDDD16, C. Kessler, IDA, Linköpings universitet, 2008.

2a.2

Scanner fragment with direct coding



```
Token getNextToken( void )
{
    char ch = getc( inputfile );
    char idstr [ ... ]; // lexeme buffer for identifiers
    t = new_Token();
    while ( is_blank(ch) ) // eat whitespace
        ch = getc( inputfile );
    if ( is_letter(ch) ) { // identifier:
        while ( is_letter(ch) || is_digit(ch) ) {
            append( ch, idstr );
            ch = getc( inputfile );
        }
        if ( is_blank(ch) )
            t->tokenval.stptr =
                sytab_lookup( idstr );
        else error( ... );
    }
    ...
    else if ( is_digit(ch) ) { // int-constant:
        int ivalue = ch - '0';
        ch = getc( inputfile );
        while ( is_digit(ch) ) {
            ivalue *= 10;
            ivalue += ch - '0';
            ch = getc( inputfile );
        }
        if ( is_blank(ch) )
            t->tokenval.ival = ivalue;
        else error( ... );
    }
    ...
    else {
        // others (single-char. symbols):
        t->tokentype = chtyp[ ch ];
        if ( t->tokentype == 0 )
            error( ... );
    }
    return t;
}
```

TDOB44 / TDDD16, C. Kessler, IDA, Linköpings universitet, 2008.

2a.3