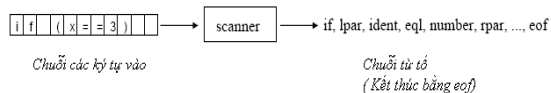


## Bài 5 Bộ phân tích từ vựng

1

### Nhiệm vụ của bộ phân tích từ vựng

- Phát hiện các từ tổ



- Bỏ qua các ký tự không cần thiết

- ☐ Khoảng trống
- ☐ Dấu tab
- ☐ Ký tự xuống dòng (CR,LF)
- ☐ Chú thích

2

### Từ tổ có cấu trúc cú pháp

```
ident = letter {letter | digit}.
number = digit {digit}.
if = "if" "if".
eql = "=" "=" .
...
```

- Tại sao không xử lý các luật này trong giai đoạn phân tích cú pháp ?

3

### Xử lý các luật từ vựng trong bộ phân tích cú pháp ?

- Làm cho bộ phân tích cú pháp trở nên quá phức tạp
  - ☐ Phân biệt tên và từ khoá
  - ☐ Phải có những luật phức tạp để xử lý chuỗi các ký tự không cần thiết (khoảng trống, tab, chú thích . . .)

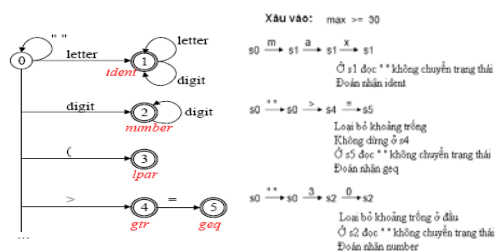
4

## Các từ tổ của KPL

- Số nguyên
- Định danh
- Từ khóa: begin, end, if, then, while, do, call, const, var, procedure, program, type, function, of, integer, char, else, for, to, array
- Hằng ký tự
- Dấu phép toán:
  - số học  
+ - \*/
  - so sánh  
= != < > <= >=
- Dấu phân cách  
( ) . : ; ( . . )
- Dấu phép gán :=

5

## Ôtômat hữu hạn của bộ PTTV



Sau mỗi từ tổ được nhận biết, bộ từ vựng lại quay lại trạng thái s0

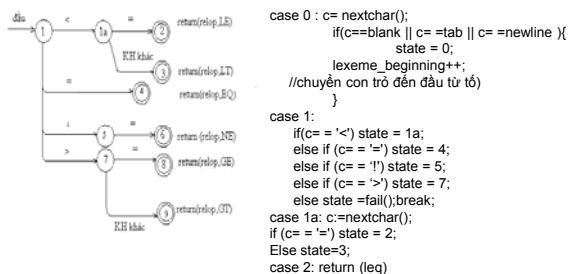
6

## Xử lý các loại từ tổ

```
switch(state)
{
  case 0 // Bỏ qua khoảng trống
  case 1 // Xử lý các phép toán quan hệ
  case 10
  // Xử lý tên
  case 13
  // Xử lý số
  ...
}
```

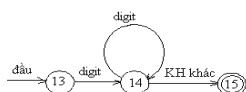
7

## Các phép toán quan hệ



8

## Số nguyên



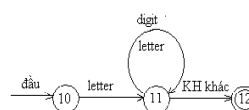
```

case 13: c = nextchar();
if(isdigit(c)) state = 14;
case 14: c=nextchar();
if(isdigit(c)) state = 14;
else state = 15; break;
case 15 : retract(1);
install_num();
return(num);

```

9

## Định danh



```

case 10:
c = nextchar();
if(isletter(c)) state =11;
else state = fail; break;
case 11:
c = nextchar();
if(isletter(c)) state =11;
else if (isdigit(c)) state = 11;
else state = 12; break;
case 12: retract(1) ;
install_id();
return (gettoken());

```

10

## Các thông tin trong bảng ký hiệu

- Tên: xâu ký tự
- Thuộc tính: tên kiểu, tên biến, tên thủ tục, tên hằng. . .
- Kiểu dữ liệu
- Phạm vi sử dụng
- Địa chỉ vùng nhớ, kích cỡ vùng nhớ
- . . .

11

## Xử lý tên / từ khoá

- Khởi đầu ch chứa ký tự đầu của tên
- Đọc tiếp các chữ cái và chữ số, lưu vào t.string
- Tìm trong bảng từ khóa (sử dụng giải thuật tìm kiếm nhị phân)
  - Nếu tìm thấy t.kind = số hiệu của từ khoá
  - Ngược lại t.kind = ident
- Kết thúc ch chứa ký tự đứng ngay sau tên hoặc từ khoá

12

## Cấu trúc dữ liệu

enum symbol

```
{  
    nul, ident, number,  
    plus, minus, times, slash,  
    eql, neq, lss, leq, grt, geq,  
    lparen, rparen, comma, semicolon, period, becomes,  
    quote, colon, lsquare, rsquare,  
    beginsym, endsym, ifsym, thensym, whilesym, dosym,  
    callsym, constsym, varsym, procsym, programsym, typesym,  
    funcsym, ofsym, integersym, charsym, elsesym, forsym,  
    tosym, arraysym  
};
```

13