

The basic task of the lexical analyzer is to scan the source-code strings into small, meaningful units, that is, into the tokens we talked about in Chapter 1.

Chapter 2 -- Lexical Analysis

2













• The lexical analyzer is best implemented as a finite state machine or a finite state automaton.

- Informally a <u>Finite-State Automaton</u> is a system that has a finite set of states with rules for making transitions between states.
- The goal now is to explain these two things in detail and bridge the gap from the first to the second.

Chapter 2 -- Lexical Analysis



3.1 State Diagrams and State Tables

- Def: <u>State Diagram</u> -- is a directed graph where the vertices in the graph represent states, and the edges indicate transitions between the states.
- Let's consider a vending machine that sells candy bars for 25 cents, and it takes nickels, dimes, and quarters.

Chapter 2 -- Lexical Analysis







state	Nickel	Dime	Quarter	Select	
0	1	2	5	0)
1	2	3	6	1	
2	3	4	6	2	
3	4	5	6	3	> Next state
4	5	6	6	4	
5	6	6	6	0*	
6	6	6	6	0*	J
	•				
		Chapter 2 I	exical Analysis.		12































Def: <u>a language over an alphabet Σ</u> is any set of strings made up only of the characters from Σ

Def: <u>L(M) -- the language accepted by M</u> is the set of all strings over Σ that are accepted by M

Chapter 2 -- Lexical Analysis





4. Nondeterministic Finite-State Automata

- So far, the behavior of our FSAs has always been predictable. But there is another type of FSA in which the state transitions are not predictable.
- In these machines, the state transition function is of the form:

N: Q x (Σ U { ε }) -> P(Q)

♦ Note: some authors use a Greek lambda, λ or A Chapter 2 -- Lexical Analysis 24





































• Informally a regular expression of an alphabet Σ is a combination of characters from Σ and certain operators indicating concatenation, selection, or repetition.

Chapter 2 -- Lexical Analysis

- \bullet b^{*} -- 0 or more b's (Kleene Star)
- ♦ b⁺ -- 1 or more b's

◆ | -- a|b -- choice







































- Identifier/Keyword differences
 - Accept everything as an identifier, and then look up keywords in table. Or pre-load the Symbol Table with Keywords.
- When you read an identifier, you read the next character in order to tell it was the end. You need to back up (put it back on the input stream). 49

Chapter 2 -- Lexical Analysis





- Lesk & Schmidt -- "LEX a lexical analyzer generator" in Unix Programmers Manual Vol. 2
- Mason & Brown -- *Lex & Yacc* -- O'Reilly & Associates.

Chapter 2 -- Lexical Analysis