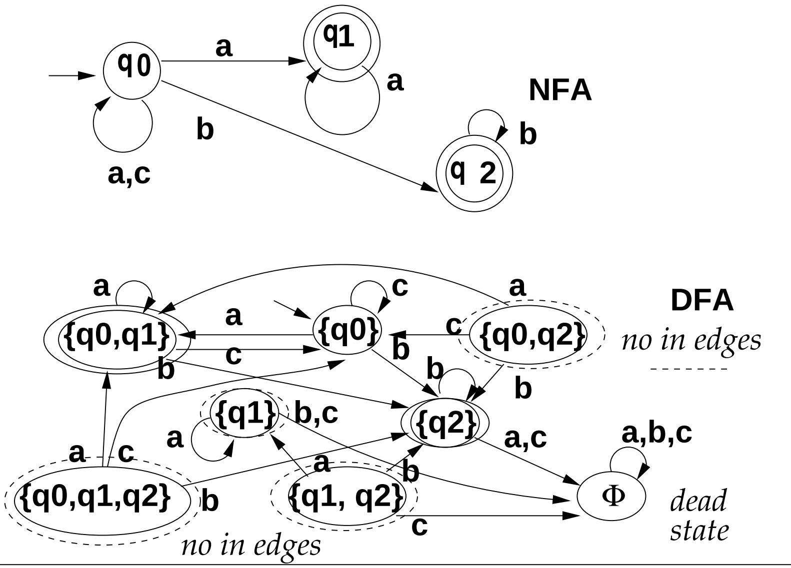
Example NFA and Equivalent DFA



Example

L = {w over {a1,a2,a3} | w does not contain ai
for some i}

NFA DFA

Regular Expressions & Languages

Regular expressions are a formalism for describing patterns in strings *(used in tools like Emacs, Perl,...)*

Language of a regular expression is the set of strings described

a a in Σ	{a}
ε	$\{\varepsilon\}$
ϕ	ϕ
$(R1 \cup R2)$ (R1, R2 reg. exp)	$L(R1) \cup L(R2)$
$(R1 \circ R2)$	$L(R1) \circ L(R2)$
$(R1)^*$	$(L(R1))^*$

What kind of definition is this?

Regular Expressions, Examples over {a,b}

$(a \cup b)$ $(a)^*$ $(a \cup b)^*$

All these parentheses are clumsy--can omit but then must know what order to carry out operations

$a \cup b^*$???

Precedence of operations: $*$, $\circ$, $\cup$

(associate to left)

but ( ) must always be followed if included!

Shorthand: use Σ to stand alphabet, ex. $a \cup b$

Omit $\circ$ where understood

$(\Sigma\Sigma\Sigma)^*$ $\Sigma^* bb \Sigma^*$ $a^* b$ $a^* \phi$ $a^* \varepsilon$

Lexical Analysis (Scanning) in Compilers

First phase of the compiler: **strings** $\longrightarrow$ **tokens**

Programming-language dependent

Tokens can include

keywords	**if**	**then**	**end**
operators	**<**	**>**	**=**
identifiers			
literals	**decimal integer, character**		

One regular expression to describe each of the token types

Qu: How recognize tokens?

Can use NFA or DFA

What is the Regular Expression?

A *Java identifier* **must start with a letter, underscore, or $, subsequent characters can also contain digits**

(Letters are A-Z, a-z, digits are 0-9)

A *Java decimal integer literal* **consists of a sequence of digits, without a leading zero, and with an optional L or l (indicating it is a long integer)**

It must have at least one digit.

A *Java floating point literal* **can have the following parts: a decimal integer literal, a decimal point, a fraction (decimal integer literal), an exponent (E or e followed by an optional sign and a decimal integer)**

It must have either . or an exponent.