Nondeterministic Finite Automata (NFA)

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

CS 536

Previous Lecture

Scanner: converts a sequence of characters to a sequence of tokens

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Scanner implemented using FSMs

FSM: DFA or NFA

This Lecture

Assignment Project Exam Help

NFAs from a formal perspective

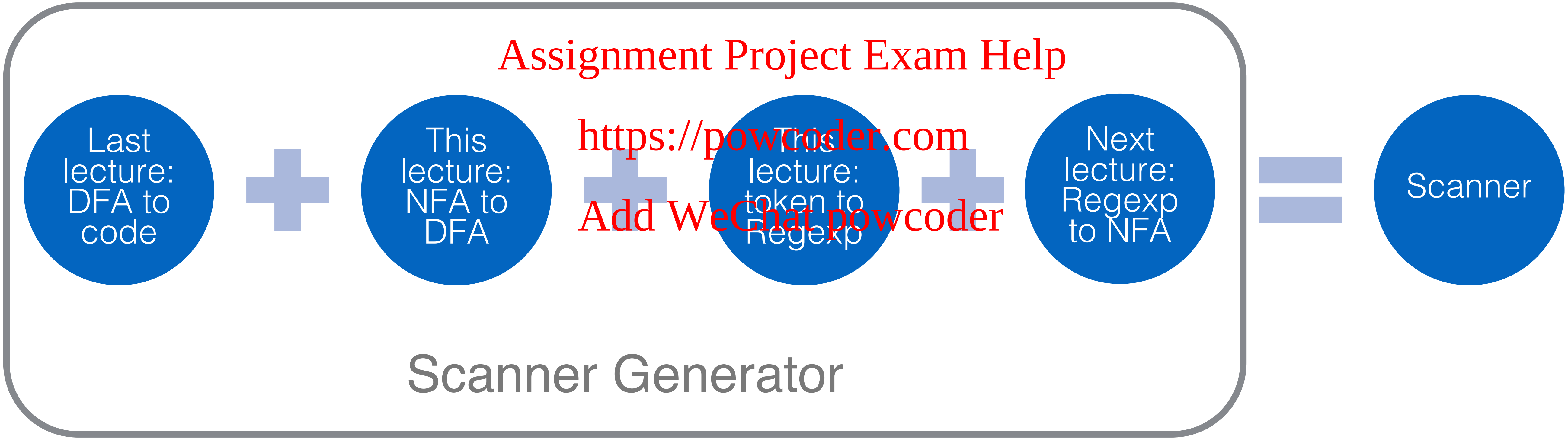
https://powcoder.com

Theorem: NFAs and DFAs are equivalent

Add WeChat powcoder

Regular languages and Regular expressions

Creating a Scanner


Last lecture: DFA to code
This lecture: NFA to DFA
This lecture: token to Regexp
Next lecture: Regexp to NFA
Scanner
Scanner Generator

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

NFAs, formally

$$M \equiv (Q, \Sigma, \delta, q, F)$$

finite set of states

Assignment Project Exam Help

https://powcoder.com

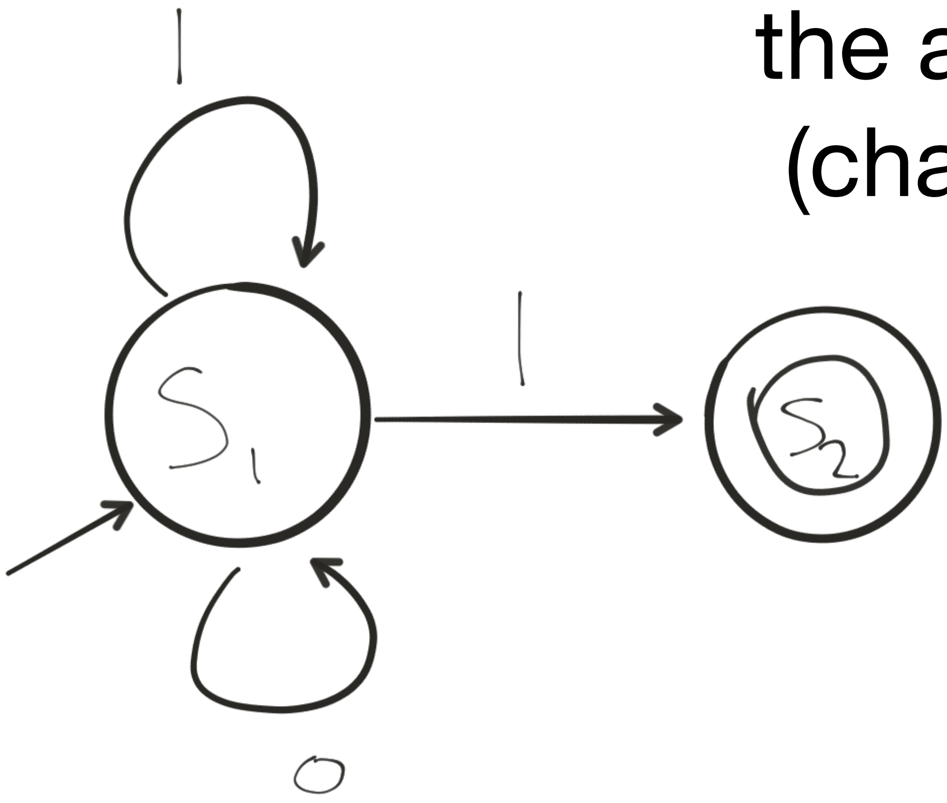
Add WeChat powcoder

the alphabet (characters)

final states $F \subseteq Q$

start state $q \in Q$

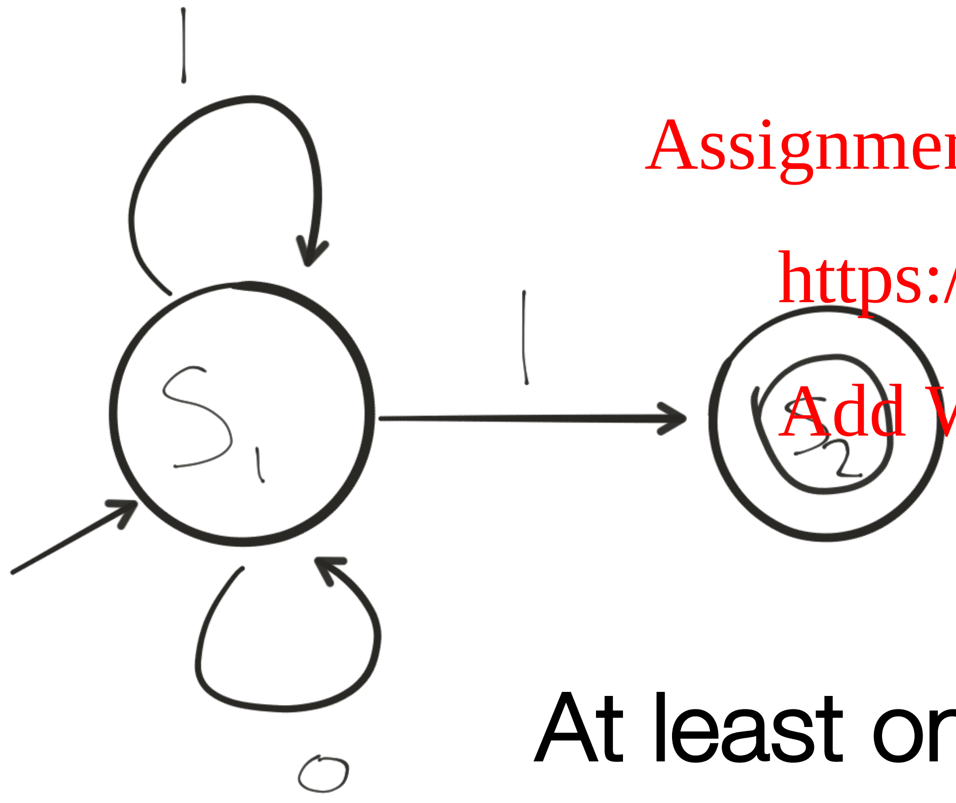
transition function $\delta : Q \times \Sigma \to 2^Q$



	0	1
s1	{s1}	{s1, s2}
s2		

NFA

To check if string is in *L*(*M*) *of* NFA *M*, simulate **set** of **choices** it could make



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

	1	1	1
s1	s2	st	st
s1	s1	s2	st
s1	s1	s1	**s2**
s1	s1	s1	s1

At least one sequence of transitions that:

Consumes all input (without getting stuck)

Ends in one of the final states

NFA and DFA are Equivalent

Two automata M and M’ are equivalent iff L(M) = L(M’)

Lemmas to be proven

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

✓ **Lemma 1**: Given a DFA M, one can construct an NFA M’ that recognizes the same language as M, i.e., L(M’) = L(M)

Lemma 2: Given an NFA M, one can construct a DFA M’ that recognizes the same language as M, i.e., L(M’) = L(M)

Proving Lemma 2

Lemma 2: Given an NFA M, one can construct a DFA M' that recognizes the same language as M, i.e., L(M') = L(M)

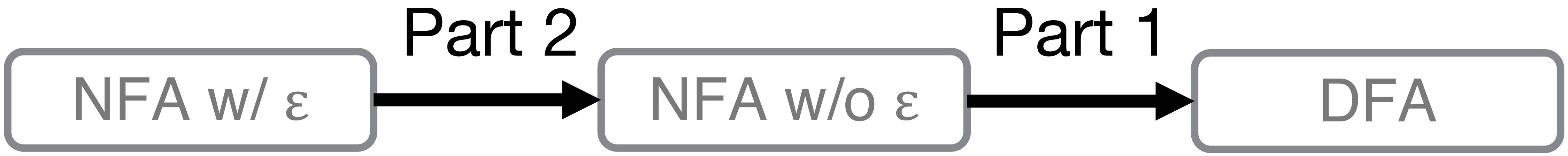
Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Part 1: Given an NFA M without ε–transitions, one can construct a DFA M' that recognizes the same language as M

Part 2: Given an NFA M with ε–transitions, one can construct an NFA M' without ε–transitions that recognizes the same language as M



NFA w/o ε–Transitions to DFA

NFA M to DFA M’

Assignment Project Exam Help

https://powcoder.com

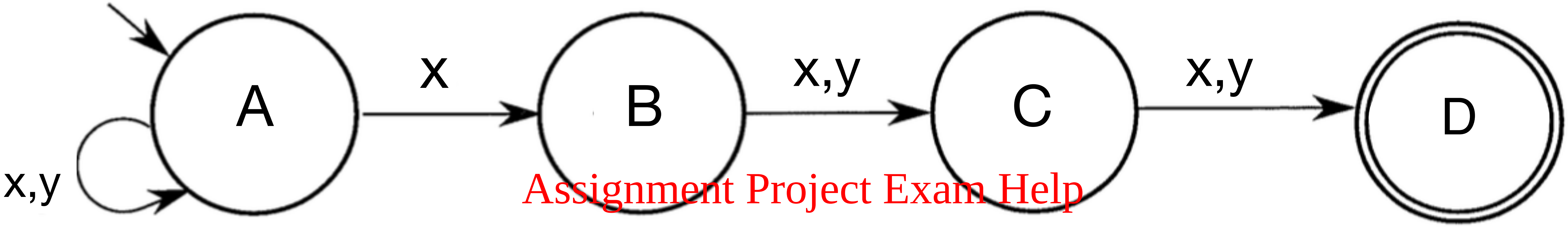
Add WeChat powcoder

Intuition: Use a single state in M’ to simulate sets of states in M

M has |Q| states

M’ can have only up to $2^{|Q|}$ states

NFA w/o ε–Transitions to DFA

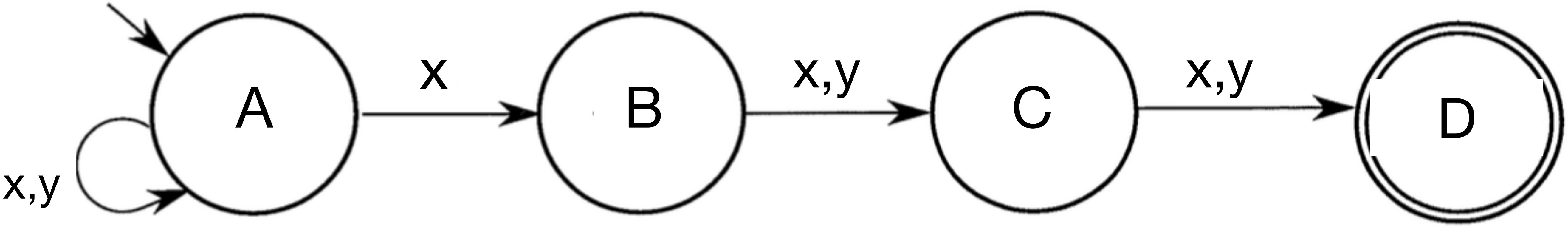


Assignment Project Exam Help

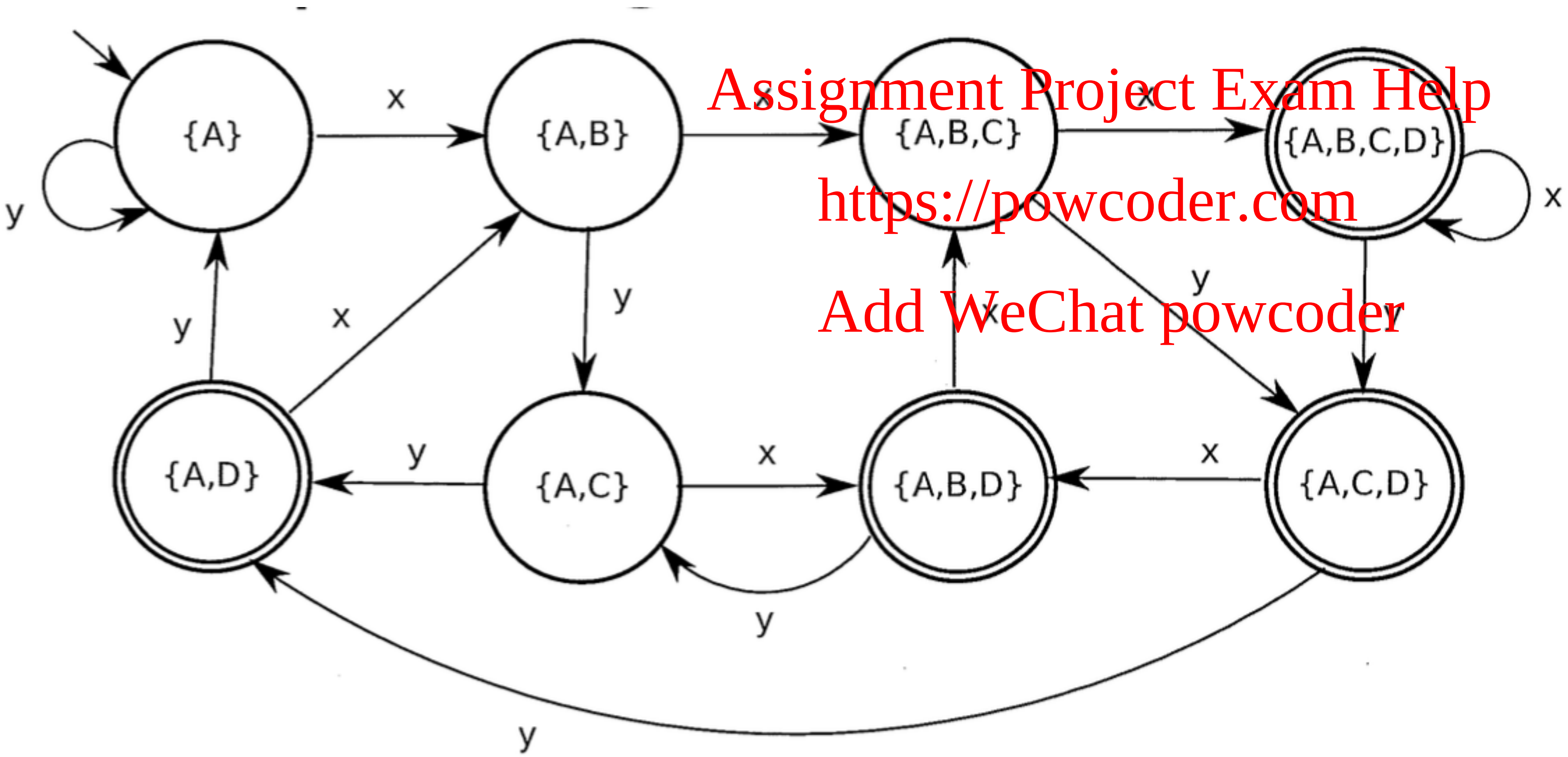
https://powcoder.com

Add WeChat powcoder

	X	Y
A	{A, B}	{A}
B	{C}	{C}
C	{D}	{D}
D	{}	{}



Build new DFA M’ where Q’ = 2^Q



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

	x	y
A	{A, B}	{A}
B	{C}	{C}
C	{D}	{D}
D	{}	{}

To build DFA: Add an edge from state S on character c to state S’ if S’ represents the set of all states that a state in S could possibly transition to on input c

Proving Lemma 2

Lemma 2: Given an NFA M, one can construct a DFA M’ that recognizes the same language as M, i.e., L(M’) = L(M)

Assignment Project Exam Help

https://powcoder.com

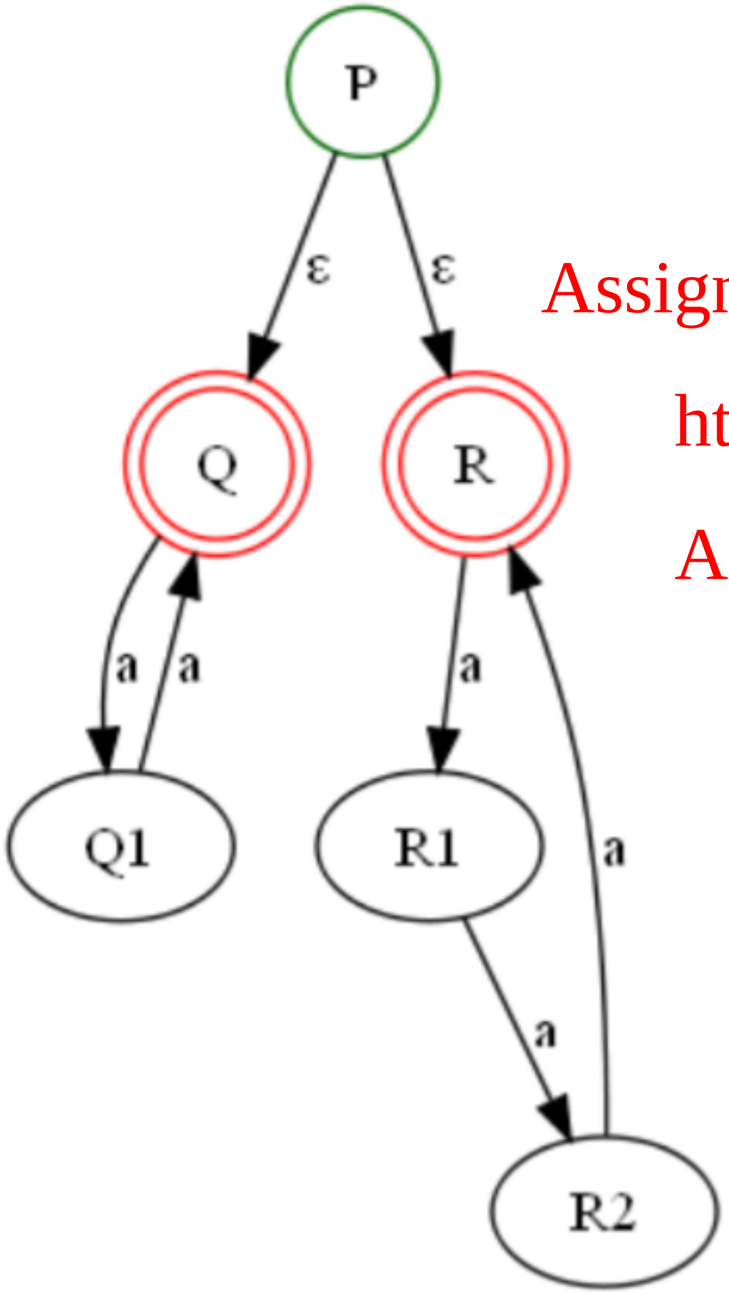
Add WeChat powcoder

✓ **Part 1**: Given an NFA M without ε–transitions, one can construct a DFA M’ that recognizes the same language as M

Part 2: Given an NFA M with ε–transitions, one can construct an NFA M’ without ε–transitions that recognizes the same language as M

ε-transitions

E.g.: x^n, where n is even **or** divisible by 3



Useful for taking union of two FSMs

In example, left side accepts even n; right side accepts n divisible by 3

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Eliminating ε-transitions

We want to construct ε-free NFA M’ that is equivalent to M

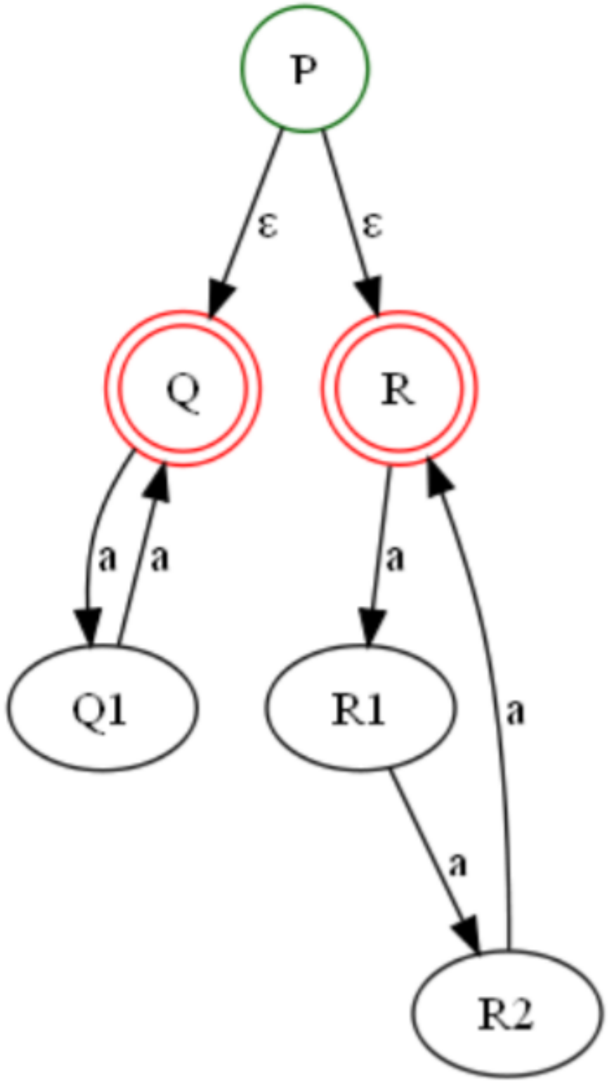
Definition: Epsilon Closure

eclose(s) = set of all states reachable from s using zero or more epsilon transitions

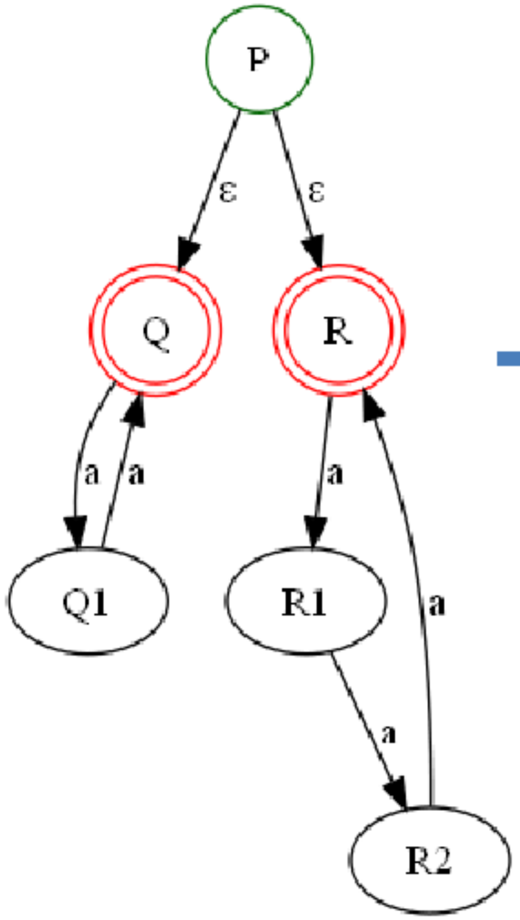
Assignment Project Exam Help

https://powcoder.com

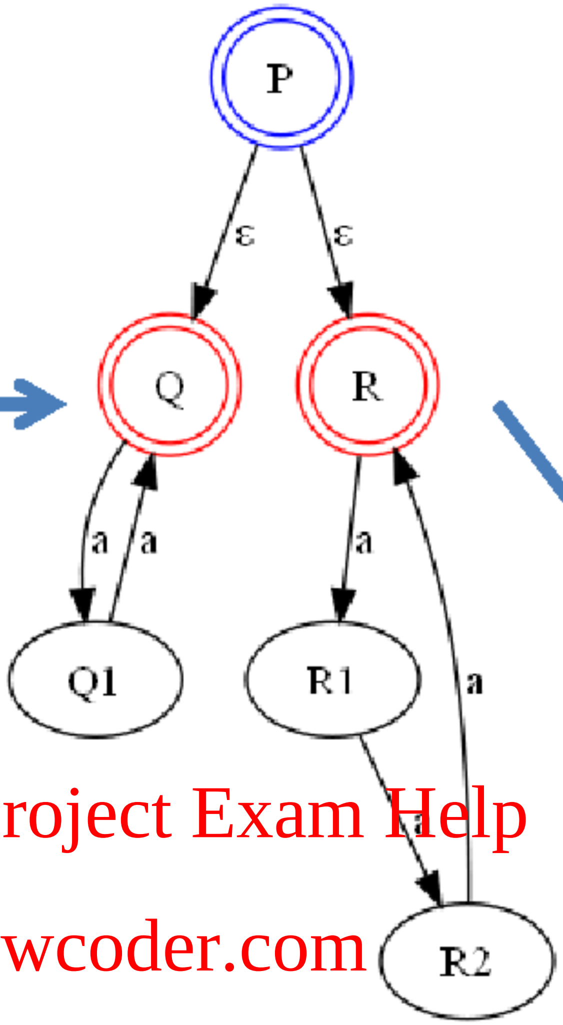
Add WeChat powcoder



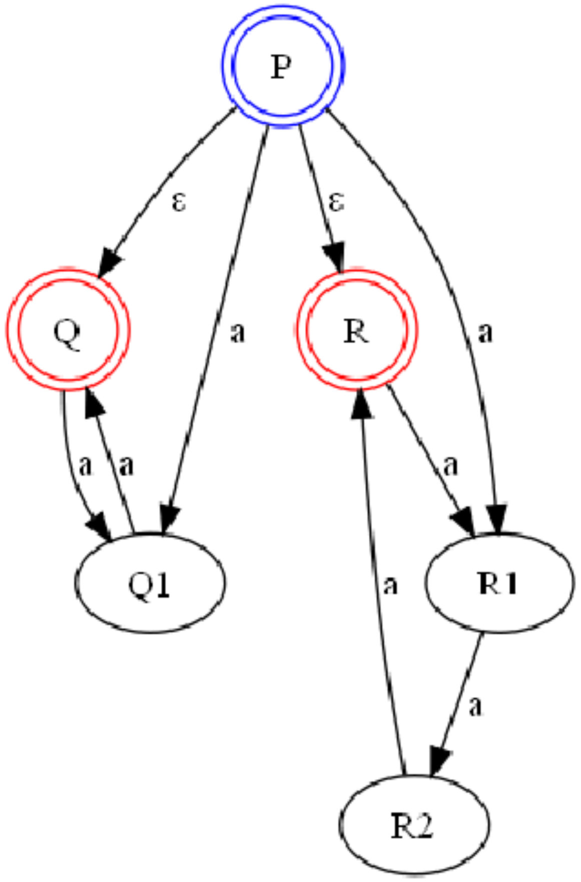
	eclose
P	{P, Q, R}
Q	{Q}
R	{R}
Q1	{Q1}
R1	{R1}
R2	{R2}



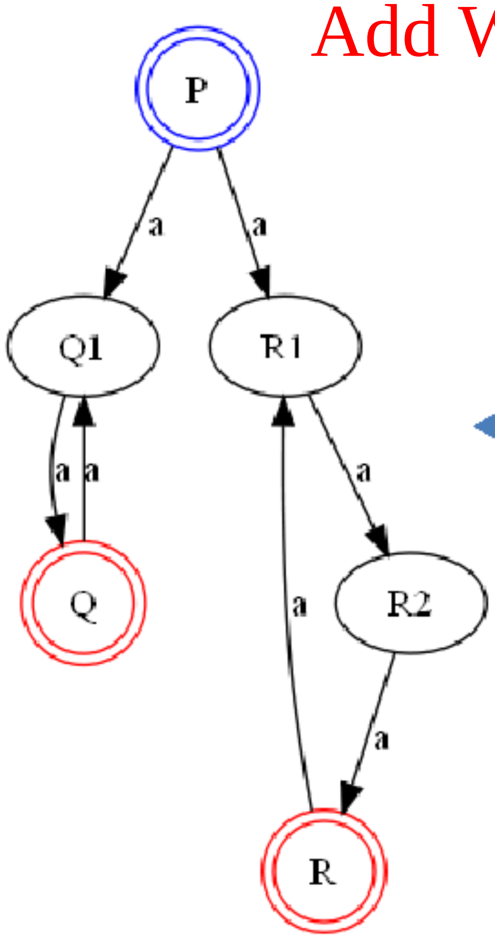
Make p an accepting state of N' iff ECLOSE(p) contains an accepting state of N.



Add an arc from p to q labeled a iff there is an arc labeled a in N from some state in ECLOSE(p) to q.



Delete all arcs labeled ε.



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Proving Lemma 2

Lemma 2: Given an NFA M, one can construct a DFA M’ that recognizes the same language as M, i.e., L(M’) = L(M)

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

✓ **Part 1**: Given an NFA M without ε–transitions, one can construct a DFA M’ that recognizes the same language as M

✓ **Part 2**: Given an NFA M with ε–transitions, one can construct an NFA M’ without ε–transitions that recognizes the same language as M

Summary of FSMs

DFAs and NFAs are equivalent

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

- An NFA can be converted into a DFA, which can be implemented via the table-driven approach

ε-transitions do not add expressiveness to NFAs

- Algorithm to remove ε-transitions

Regular Languages and Regular Expressions

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Regular Language

Any language recognized by an FSM is a regular language

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Examples:

- Single-line comments beginning with //
- Integer literals
- {ε, ab, abab, ababab, abababab, …. }
- C/C++ identifiers

Regular Expression

A pattern that defines a regular language

Assignment Project Exam Help

Regular language: set of (potentially infinite) strings

https://powcoder.com

Add WeChat powcoder

Regular expression: represents a set of (potentially infinite) strings by a single pattern

$\{\varepsilon$, ab, abab, ababab, abababab, …. $\} \Leftrightarrow (ab)^*$

Why do we need them?

Each token in a programming language can be defined by a regular language

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Scanner-generator input: one regular expression for each token to be recognized by scanner

Regular expressions are inputs to a scanner generator

Regular Expression

operands: single characters, epsilon

Assignment Project Exam Help

operators: from low to high precedence

https://powcoder.com

Add WeChat powcoder

“or”: a | b

“followed by”: a.b, ab

“Kleene star”: a* (0 or more a-s)

Regular Expression

Conventions:

aa is a . a

a+ is aa*

letter is a|b|c|d|…|y|z|A|B|…|Z

digit is 0|1|2|…|9

not(x) all characters except x

. is any character

parentheses for grouping, e.g., (ab)* is {ε, ab, abab, ababab, … }

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Regexp, example

Precedence: * > . > |

digit | letter letter

(digit) | (letter . letter)

one digit, or two letters

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

digit | letter letter*

(digit) | (letter . (letter)*)

one digit, or one or more letters

digit | letter+

Regexp, example

Hex strings

start with 0x or 0X

Assignment Project Exam Help

followed by one or more hexadecimal digits

https://powcoder.com

optionally end with l or L

Add WeChat powcoder

0(x|X)hexdigit+(L|l|ε)

where hexdigit = digit|a|b|c|d|e|f|A|…|F

Regexp, example

Integer literals: sequence of digits preceded by optional +/-

Assignment Project Exam Help

https://powcoder.com

Example: -543, +15, 0007

Add WeChat powcoder

Regular expression

(+|-|ε)digit+

Regexp, example

Single-line comments

- Example: // this is a comment

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Regular expression

- //(not(‘\n’))*’\n’

Regexp, example

C/C++ identifiers: sequence of letters/digits/underscores; cannot begin with a digit; cannot end with an underscore

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Example: a, _bbb7, cs_536

Regular expression

letter | (letter|_)(letter|digit|_)*(letter|digit)

Recap

Regular Languages

- Languages recognized/defined by FSMs

Assignment Project Exam Help

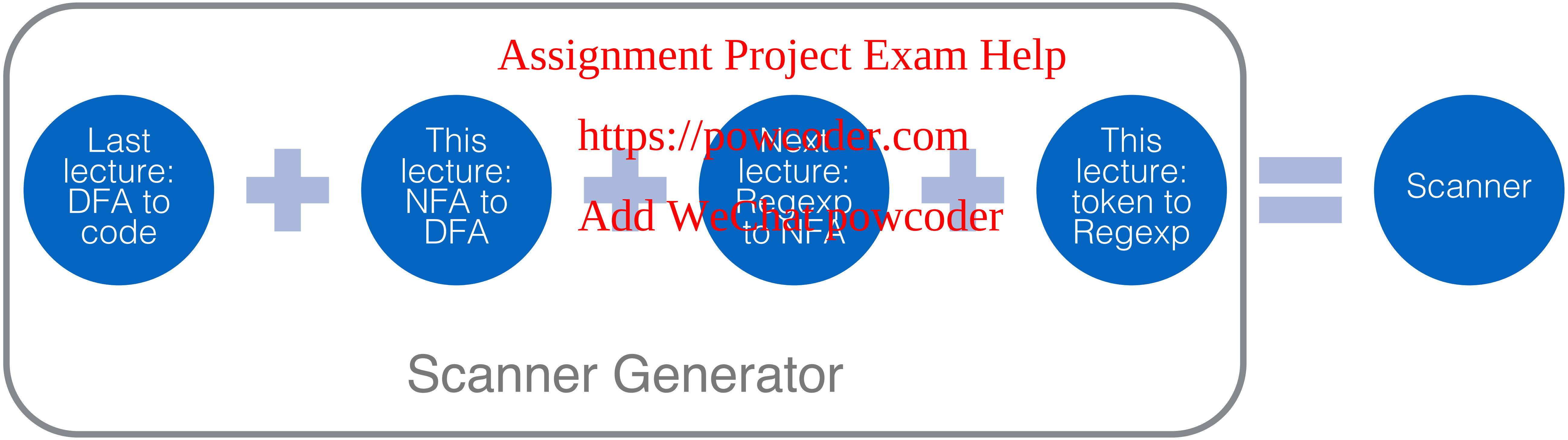
https://powcoder.com

Add WeChat powcoder

Regular Expressions

- Single-pattern representations of regular languages
- Used for defining tokens in a scanner generator

Creating a Scanner



Last lecture: DFA to code

This lecture: NFA to DFA

Next lecture: Regexp to NFA

This lecture: token to Regexp

Scanner

Scanner Generator

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder