

程序代写代做 CS编程辅导

FIT2014 Theory of Computation



Lecture 6

Regular Expressions

WeChat: cstutorcs

Assignment Project Exam Help

slides by Graham Farr

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

based in part on previous slides by David Albrecht

QQ: 749389476

COMMONWEALTH OF AUSTRALIA  
<https://tutorcs.com>

Copyright Regulations 1969

Warning

This material has been reproduced and communicated to you by or on behalf of Monash University  
in accordance with s113P of the Copyright Act 1968 (the Act).

The material in this communication may be subject to copyright under the Act.

Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice.

# Overview

程序代写代做 CS编程辅导



- ▶ Some Problems
- ▶ Applications of Regular Expressions

WeChat: cstutorcs

- ▶ Regular Expressions
- ▶ Regular Languages

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

## Some Problems

程序代写代做 CS编程辅导



- ▶ Find all the files which contain subject course codes.
- ▶ Find all the e-mail addresses in a set of mail files.
- ▶ Change the way comments in programs are formatted in your web pages.
- ▶ Using web server access files, record how many times each page is visited, and how many times each link is used.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# Applications of Regular Expressions

程序代写代做 CS编程辅导



► Useful way to describe simulations.

► Used in several programs:

► Editors: vi, emacs

► Filters: egrep, sed, gawk

► Programming languages: Java, C, Perl

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# Filters

程序代写代做 CS编程辅导



egrep

- ▶ A program which searches a pattern described by a regular expression.

sed

WeChat: cstutorcs

- ▶ A program which enables stream editing of files

Assignment Project Exam Help

awk, nawk, gawk

Email: tutorcs@163.com

- ▶ Programming languages which enable text manipulation.

QQ: 749889476

<https://tutorcs.com>

# Programming Languages

程序代写代做 CS编程辅导



JFlex, flex, lex

- ▶ Languages used to generate analysers.

CUP, bison, yacc

- ▶ Languages used to generate compilers.

Perl

- ▶ A powerful scripting language, developed in the 1980s by Larry Wall.

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Regular Expressions for Small Languages

程序代写代做 CS编程辅导

The empty language

$\emptyset$

Language  $\{\epsilon\}$  consisting only of the empty word

$\epsilon$

Language  $\{w\}$  consisting only of the single word  $w$

$w$

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

E.g.:

the language  $\{abbab\}$

QQ: 749389476

abbab

<https://tutorcs.com>



# Alternatives, Grouping

## Alternatives

are indicated by  $\cup$ .

E.g.:

$1 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9$

is a regular expression for:

$\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

## Grouping

is indicated by  $( )$ .

E.g.:

$QQ:(a|b|c)(e \cup g)$

is a regular expression for:

$\{abe, abg, bae, bag\}$

程序代写代做 CS编程辅导



WeChat: zstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: (2493894176)

<https://tutorcs.com>



# Finite Languages

程序代写代做 CS编程辅导

consist of finite number of words

E.g.:



$\{abaaba, abbbbba, abbaba\}$

WeChat: cstutorcs

Regular Expression:

Assignment Project Exam Help

$abaaba \cup abbbbba \cup abbaba$

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

... or:

QQ: 749389476

$ab(aa \cup bb \cup ba)ba$

<https://tutorcs.com>

# Kleene Star

程序代写代做 CS编程辅导

Matching an expression *zero or more times* is indicated by  $*$



For example:

$a^*$  represents  $\{\epsilon, a, aa, aaa, aaaa, \dots\}$

$(ab)^*$  represents  $\{\epsilon, ab, abab, ababab, \dots\}$

$ab^*$  represents  $\{a, ab, abb, abbb, abbbb, \dots\}$

Note:  $ab^* \neq (ab)^*$

<https://tutorcs.com>

程序代写代做 CS编程辅导



$$\begin{aligned}(aa \cup bb)^* &= (aa \cup bb)^0 \cup (aa \cup bb)^1 \cup (aa \cup bb)^2 \cup \dots \\ &= \varepsilon \cup (aa \cup bb) \cup (aa \cup bb)(aa \cup bb) \cup \dots\end{aligned}$$

represents:

$\{\varepsilon, aa, bb, aaaa, aabb, bbaa, bbbb, aaaaaa, aaaabb, aabbaa, \dots\}$

Assignment Project Exam Help

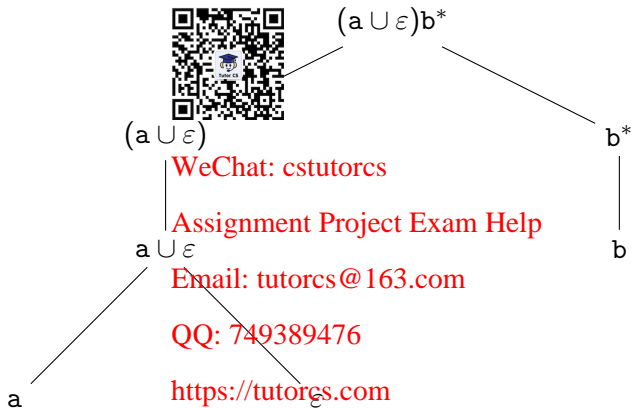
Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Parse Tree

程序代写代做 CS编程辅导



WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Definition

程序代写代做 CS编程辅导

1.  $\emptyset$  and  $\varepsilon$  are regular expressions.
2. All letters in the alphabet are regular expressions.
3. If  $R$  and  $S$  are regular expressions, then so are:



(i)  $(R)$

(ii)  $RS$

(iii)  $R \cup S$

(iv)  $R^*$

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Regular Language

程序代写代做 CS编程辅导



A language which can be described by a *regular expression* is called a **regular language**.

WeChat: cstutorcs

If a word belongs to the language described by a regular expression, then we say it is **matched** by the regular expression.

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

## Example: EVEN-EVEN

程序代写代做 CS编程辅导

Recall:



EVEN-EVEN = { All strings a and b each occur an *even* number of times }  
= {  $\epsilon$ , aa, bb, aaaa, aabb, abab, abba, ... }.

WeChat: cstutorcs

Assignment Project Exam Help

Regular Expression:

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

$(aa \cup bb \cup (ab \cup ba)(aa \cup bb)^*(ab \cup ba))^*$

QQ: 749389476

<https://tutorcs.com>

## Things to think about ...

程序代写代做 CS编程辅导

Is the set of all English words (from a standard dictionary) a regular language?

Is DOUBLEWORD (see Lecture 1) a regular language?

Is PALINDROME a regular language?

Is the set of all grammatical English sentences a regular language?

How would you determine, for a given string and regular expression, whether the string matches the regular expression?

<https://tutorcs.com>



WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476



## Example: Floating Point Number

程序代写代做 CS编程辅导



A floating point number has one or more digits, which may begin with a minus sign (—), and which may contain a decimal point.

WeChat: cstutorcs

E.g.,

0 1.2 — 3 — 4.675 002 023.50

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Sequence of Digits

One Digit

0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9

Two Digits

(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)

WeChat: cstutorcs

Three Digits

(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)

Email: tutores@163.com

One or more Digits

QQ: 749389476

(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)(0 ∪ 1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9)\*

<https://tutorcs.com>

程序代写代做 CS编程辅导



# Sequences of Digits

程序代写代做 CS编程辅导

Digit

$$D = (0 \cup 1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9)$$



Two Digits

DD or  $D^2$   
WeChat: cstutorcs

Three Digits

Assignment Project Exam Help

DDD or  $D^3$

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

One or more Digits

QQ: 749389476  
DD\*

<https://tutorcs.com>

# Numbers

程序代写代做 CS编程辅导

One Digit

$$D = (0 \cup 1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9)$$

Nonnegative Integers

$$N = DD^*$$



eg: 1 123  
WeChat: cstutorcs

1209 002 020

Integers

Assignment Project Exam Help

$$Z = N \cup (-N)$$

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

Floating Point Number

QQ: 749389476

$$F = Z \cup (Z.) \cup (.N) \cup (-.N) \cup (Z.N)$$

<https://tutorcs.com>

## Other Notations

程序代写代做 CS编程辅导

alternative notation

$R \cup S$

$0 \cup 1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9$

any letter a to z

$RR^*$

$\epsilon \cup R$



WeChat: cstutorcs

$R | S$

$[0-9]$

$[a-z]$

$R^+$

$R?$

*Be careful!* Many variations exist. Tools that use regexps differ in many details:

- ▶ whether or not they use the above special meanings of  $+$  and  $?$ ;
- ▶ how they handle the parentheses and vertical bar for alternatives
  - ▶ sometimes  $(\dots | \dots)$
  - ▶ sometimes  $\backslash(\dots \backslash | \dots \backslash)$
- ▶ how they use full stop (often represents any non-newline character);
- ▶ how they represent newline characters
- ▶ ...

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Revision

程序代写代做 CS编程辅导

- ▶ Regular Expressions
- ▶ Definitions
- ▶ How to use them to define Languages
- ▶ Regular languages



WeChat: cstutorcs

Read Sipser, §1.3, pp 63–66.

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

Additional Reading

QQ: 749389476

- ▶ Jeffrey E.F. Friedl, *Mastering Regular Expressions: Powerful Techniques for Perl and Other Tools*, O'Reilly, 1997. <https://tutorcs.com>