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

based in part on previous slides by David Albrecht

QQ: 749389476

Overview

程序代写代做 CS编程辅导

▶ Some Problems

▶ Applications of Regular Expressions

▶ Regular Expressions

▶ Regular Languages

WeChat: cstutorcs

Assignment Project Exam Help

Email: 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

- Useful way to describe simple patterns.
- Used in several programs:
- Editors: `vi`, `emacs`
- Filters: `egrep`, `sed`, `gawk`
- Programming languages: JFlex, CUP, Perl

Filters

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

sed
▶ A program which enables stream editing of files.

awk, nawk, gawk
▶ Programming languages which enable text manipulation.

Programming Languages

JFlex, flex, lex

▶ Languages used to generate lexical analysers.

CUP, bison, yacc

▶ Languages used to generate compilers.

Perl

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

Regular Expressions for Small Languages

The empty language $\emptyset$

Language $\{\varepsilon\}$ consisting only of the empty word ε

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

E.g.:
the language $\{\texttt{abbab}\}$ $\texttt{abbab}$

Alternatives, Grouping

Alternatives
are indicated by ∪.
E.g.:

Grouping
is indicated by ( ).
E.g.:

程序代写代做 CS编程辅导

1 ∪ 2 ∪ 3 ∪ 4 ∪ 5 ∪ 6 ∪ 7 ∪ 8 ∪ 9
is a regular expression for:
$\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476
(a ∪ b)(e ∪ g)
is a regular expression for:
$\{abe, abg, bae, bag\}$

https://tutorcs.com

Finite Languages

consist of finite number of words

E.g.:

$$\{\text{abaaba}, \text{abbbba}, \text{abbaba}\}$$

Regular Expression:

$$\text{abaaba} \cup \text{abbbba} \cup \text{abbaba}$$

. . . or:

$$\text{ab}(\text{aa} \cup \text{bb} \cup \text{ba})\text{ba}$$

Kleene Star

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

For example:

a^* represents $\{\varepsilon, a, aa, aaa, aaaa, \ldots\}$

$(ab)^*$ represents $\{\varepsilon, ab, abab, ababab, \ldots\}$

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

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

Kleene Star

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

represents:

$$\{\, \varepsilon, aa, bb, aaaa, aabb, bbaa, bbbb, aaaaaa, aaaabb, aabbaa, \ldots \,\}$$
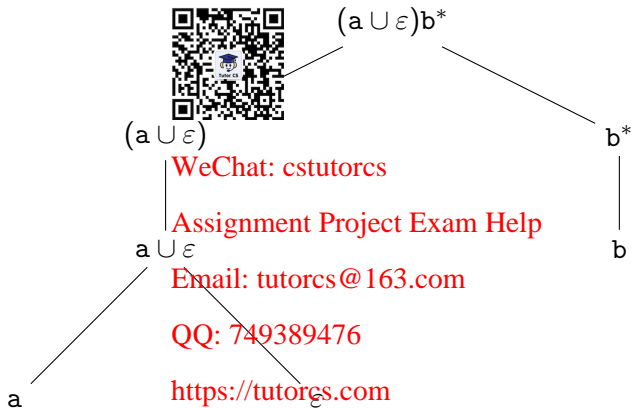
Parse Tree

程序代写代做 CS编程辅导

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

$(a \cup \varepsilon)$

WeChat: cstutorcs

Assignment Project Exam Help

$a \cup \varepsilon$

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

a

b^*

b

Definition

1. $\emptyset$ and ε are regular expressions.

2. All letters in the alphabet [...] expressions.

3. If R and S are regular expressions, then so are:

   (i) (R)

   (ii) RS

   (iii) $R \cup S$

   (iv) R^*

Regular Language

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

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

Example: EVEN-EVEN

Recall:

EVEN-EVEN $=$ { All strings $\qquad$ a and b each occur an *even* number of times }

$\qquad =$ $\{\varepsilon, \mathtt{aa}, \mathtt{bb}, \mathtt{aaaa}, \mathtt{aabb}, \mathtt{abab}, \mathtt{abba}, \ldots\}.$

Regular Expression:

$$(\mathtt{aa} \cup \mathtt{bb} \cup (\mathtt{ab} \cup \mathtt{ba})(\mathtt{aa} \cup \mathtt{bb})^{*}(\mathtt{ab} \cup \mathtt{ba}))^{*}$$

Things to think about . . .

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

Is DOUBLEWORD (see Lecture ?) 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?

Example: Floating Point Number

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

E.g.,

$$0 \quad 1.2 \quad -3 \quad -4.675 \quad 002 \quad 023.50$$

Sequence of Digits

One Digit

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

Two Digits

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

Three Digits

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

One or more Digits

$$(0 \cup 1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9)(0 \cup 1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9)^{*}$$

Sequences of Digits

Digit

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

Two Digits

$$DD \quad \text{or} \quad D^2$$

Three Digits

$$DDD \quad \text{or} \quad D^3$$

One or more Digits

$$DD^*$$

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

$$N = DD^*$$

e.g. 1 123 1209 002 020

WeChat: cstutorcs

Integers

Assignment Project Exam Help

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

Email: tutorcs@163.com

Floating Point Number

QQ: 749389476

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

https://tutorcs.com

Other Notations

alternative notation

$R \cup S$	$R \mid S$
$0 \cup 1 \cup 2 \cup 3 \cup 4 \cup \ldots \cup 7 \cup 8 \cup 9$	[0-9]
any letter a to z	[a-z]
RR^*	R^+
$\varepsilon \cup R$	$R?$

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 $(\cdots \mid \cdots)$

    ▶ sometimes $\backslash(\cdots \backslash \mid \cdots \backslash)$

▶ how they use full stop (often represents any non-newline character);

▶ how they represent newline characters

▶ $\ldots$

Revision

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

Read Sipser, §1.3, pp 63–66.

Additional Reading

- Jeffrey E.F. Friedl, *Mastering Regular Expressions: Powerful Techniques for Perl and Other Tools*, O'Reilly, 1997.