МГТУ им. Баумана

Рубежный контроль №2

По курсу: "Анализ алгоритмов"

Регулярные выражения

Работу выполнил: Лумбунов Дмитрий, ИУ7-54

Преподаватели: Волкова Л.Л., Строганов Ю.В.

Москва, 2019

Оглавление

Введение

Цель работы: изучение возможностей регулярных выражений. Реализация валидации строк при использовании регулярных выражений и конечного автомата. В ходе рубежного контроля предстоит:

- Изучить регулярные выражения;
- реализовать алгоритм валидации строк формата даты с помощью регулярных выражений и конечного автомата.

1 | Аналитическая часть

Регулярные выражения (англ. regular expressions) — формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов (символов-джокеров, англ. wildcard characters). Для поиска используется строка-образец (англ. pattern, по-русски её часто называют «шаблоном», «маской»), состоящая из символов и метасимволов и задающая правило поиска. Для манипуляций с текстом дополнительно задаётся строка замены, которая также может содержать в себе специальные символы.

1.1 Использование регулярных выражений

Регулярные выражения используются некоторыми текстовыми редакторами и утилитами для поиска и подстановки текста. Например, при помощи регулярных выражений можно задать шаблоны, позволяющие:

- найти все последовательности символов «кот» в любом контексте, как то: «кот», «котлета», «терракотовый»;
- найти отдельно стоящее слово «кот» и заменить его на «кошка»;
- найти слово «кот», которому предшествует слово «персидский» или «чеширский»;
- убрать из текста все предложения, в которых упоминается слово кот или кошка.

Регулярные выражения позволяют задавать и гораздо более сложные шаблоны поиска или замены.

Результатом работы с регулярным выражением может быть:

- проверка наличия искомого образца в заданном тексте;
- определение подстроки текста, которая сопоставляется образцу;
- определение групп символов, соответствующих отдельным частям образца.

Если регулярное выражение используется для замены текста, то результатом работы будет новая текстовая строка, представляющая из себя исходный текст, из которого удалены найденные подстроки (сопоставленные образцу), а вместо них подставлены строки замены (возможно, модифицированные запомненными при разборе группами символов из исходного текста). Частным случаем модификации текста является удаление всех вхождений найденного образца — для чего строка замены указывается пустой.

1.2 Регулярные выражения в теории формальных языков

Регулярные выражения состоят из констант и операторов, которые определяют множества строк и множества операций на них соответственно. Определены следующие константы:

- (пустое множество) $\emptyset$.
- (пустая строка) ε обозначает строку, не содержащую ни одного символа; эквивалентно "".
- (символьный литерал) "a где a — символ используемого алфавита.
- (множество) из символов, либо из других множеств. и следующие операции:
- (сцепление, конкатенация) RS обозначает множество $\alpha\beta|\alpha \in R \& \beta \in S$. Например, {"boy "girl"}{"friend "cott"} = {"boyfriend "girlfriend "boycott "girlcott"}.
- (дизъюнкция, чередование) $R|S$ обозначает объединение R и S.

  Например, {"ab "c"}|{"ab "d "ef"} = {"ab "c "d "ef"}.
- (замыкание Клини, звезда Клини) R* обозначает минимальное надмножество множества R, которое содержит ε и замкнуто относительно конкатенации. Это есть множество всех строк, полученных конкатенацией нуля или более строк из R.

  Например, {"Run "Forrest"}* = {ε, "Run "Forrest "RunRun "RunForrest "ForrestRun

  "ForrestForrest "RunRunRun

  "RunRunForrest "RunForrestRun }.
- Регулярные выражения, входящие в современные языки программирования (в частности, PCRE), имеют больше возможностей, чем то, что называется регулярными выражениями в теории формальных языков; в частности, в них есть нумерованные обратные ссылки. Это позволяет им разбирать строки, описываемые не только регулярными грамматиками, но и более сложными, в частности, контекстно-свободными грамматиками.

1.3 Вывод

Были рассмотренны регулярные выражения и их обоснование в теории формальных языков.

2 | Конструкторская часть

2.1 Требования к программе

Требования к вводу:

- на вход подается строка;
- длина входной строки не превышает максимально допустимую длину для типа String ЯП Java.

Требования к программе:

- Программа выводит входные строки и результат валидации по конечному автомату (valid/invalid);
- программа выводит результат валидации по регулярным выражениям (true/false).

2.2 Конечный автомат

В этом разделе будет рассмотрен конечный автомат, на котором основан алгоритм (Рис. ??)

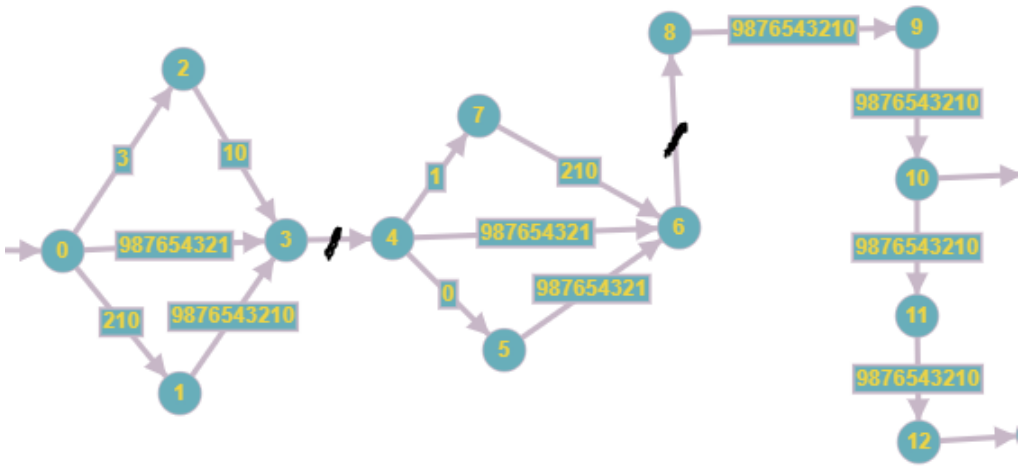

Рис. 2.1: Конечный автомат для формата даты

2.3 Вывод

В данном разделе был рассмотрен конечный автомат, на котором основан алгоритм и требования к работе программы.

3 | Технологическая часть

3.1 Выбор ЯП

Я выбрал в качестве языка программирования Java, потому как он данный язык имеет поддержку регулярных выражений и удобен при объектно-ориентированном подходе.

3.2 Листинг кода алгоритмов

В данном разделе будет представлен листинги кода валидации с помощью регулярных выражений (3.1), с помощью конечного автомата(3.2)

Листинг 3.1: Использование регулярных выражений

```
1
2 import java.util.regex.Pattern;
3
4 String[] Str = new String[5];
5 Str[0] = "31/12/01"; Str[1] = "1/3/1925"; Str[2] = "5/5,12"; Str[3] = "
    32/1/09"; Str[4] = "30.13.2012";
6
7 for (int i = 0; i < Str.length; i++){
8     System.out.println(Pattern.matches("(0?[1-9]|[12][0-9]|3[01])
        /(0?[1-9]|1[012])/((\\d{2})|(\\d{4}))",Str[i]));
9 }
```

Листинг 3.2: Использование конечного автомата

```
1
2
3 class Date {
4
5     byte day;
6     byte month;
7     short year;
8     char delimiter;
9
10    static Date stringToDate (String str) {
11        String temp = "";
12        int i = 0;
13
14        Date Ret = new Date();
15
16        //day
```

```
17          for ( ; Character.isDigit(str.charAt(i)) && i < str.length(); i
                ++){
18              temp += str.charAt(i);
19          }
20
21
22          //no delimiter
23          if (i == str.length()) return null;
24
25          //not value
26          for (int j = 0; j < temp.length(); j++){
27              if (!Character.isDigit(temp.charAt(j))) return null;
28          }
29
30
31          //length
32          if (1 > temp.length() || temp.length() > 2) return null;
33
34          //saving
35          Ret.day = Byte.parseByte(temp);
36          Ret.delimiter = str.charAt(i);
37          if (Ret.day > 31 || Ret.day < 1) return null;
38          temp = "";
39
40          i++;
41
42          //month
43          for ( ; Character.isDigit(str.charAt(i)) && i < str.length(); i
                ++){
44              temp += str.charAt(i);
45          }
46
47
48          //no second delimiter
49          if (i == str.length() || str.charAt(i) != Ret.delimiter) return
                null;
50
51          //not value
52          for (int j = 0; j < temp.length(); j++){
53              if (!Character.isDigit(temp.charAt(j))) return null;
54          }
55
56          //length
57          if (1 > temp.length() || temp.length() > 2) return null;
58
59          //saving
60          Ret.month = Byte.parseByte(temp);
61          if (Ret.month > 12 || Ret.month < 1) return null;
62          temp = "";
63
64          i++;
65          //year
66          for ( ; i < str.length(); i++){
```

```
67              temp += str.charAt(i);
68          }
69
70          //not value
71          for (int j = 0; j < temp.length(); j++){
72              if (!Character.isDigit(temp.charAt(j))) return null;
73          }
74
75          //length
76          if (2 != temp.length() && temp.length() != 4) return null;
77
78          //saving
79          Ret.year = Short.parseShort(temp);
80
81
82          return Ret;
83
84      }
85  }
```

3.3 Вывод

В данном разделе была рассмотрена структура ПО и листинги кода программы.

Заключение

В ходе лабораторной работы были изучены возможности регулярных выражений. Были реализованы алгоритмы валидации строки при помощи регулярных выражений и конечного автомата.

Литература

[1] Фридл, Дж. Регулярные выражения = Mastering Regular Expressions. — СПб.: «Питер», 2001. — 352 с. — (Библиотека программиста).

[2] Смит, Билл. Методы и алгоритмы вычислений на строках (regexp) = Computing Patterns in Strings. — М.: «Вильямс», 2006. — 496 с.

[3] Форта, Бен. Освой самостоятельно регулярные выражения. 10 минут на урок = Sams Teach Yourself Regular Expressions in 10 Minutes. — М.: «Вильямс», 2005. — 184 с.