

24

регулярные выражения
Основы

Основы

Регулярное выражение — это строка, задающая шаблон поиска подстрок в тексте. Одному шаблону может соответствовать много разных строчек.

Используем только, если паттерн очень сложный и никакой другой способ не подойдет!

ТОЧНО НЕ ИСПОЛЬЗУЕМ, когда:

- есть запрещенные комбинации
- надо сравнивать посимвольно
- надо найти определенное количество символов

re — Операции с регулярными выражениями

`finditer(паттерн, строка)` - все вхождения подстрок, подходящие под регулярку

Регулярное выражения удобнее записывать в неэкранированные строки
Например, `r"[a-z]+\. [a-z]+"`

re — Операции с регулярными выражениями

Квантификаторы

* ноль или больше

+ один или больше

{3} ровно три

{4,8} от четырех до восьми

A^* Пустота или буква A

A^+ буква A (от 1 до ∞)

$A^{\{3\}}$ AAAA

$A^{\{4,8\}}$ AA, AAA, AAAAA

Символьные классы

[ABC] один из символов A, B или C

$[ABC]^{\{4\}}$ BCAA, CABCB

[A-Z] один символ от A до Z

$[\text{A-Z}]^{\{3\}}$ 123, 234, 948, 007

[^CDE] не C, не D, не E

. - любой произвольный символ

, A BA, DA, fA, AA

ВАЖНО: если служебные символы проверяются как обычные, то перед ними поставить \ (например \., \+ \{ и т д)

re — Операции с регулярными выражениями

Скобочные группы

() - группировка символов (запоминающие скобки)

(?=(....)) - поиск всех возможных совпадений (даже пересечения!)

Перечисления

AB|CD|DE - AB или CD или DE

1

В текстовом файле k7a-1.txt находится цепочка из символов латинского алфавита А, В, С, D, E. Найдите длину самой длинной подцепочки, состоящей из символов А, В или С (в произвольном порядке).

2

Текстовый файл 24-196.txt содержит строку из заглавных латинских букв X, Y и Z, всего не более чем из 106 символов. Определите максимальное количество идущих подряд пар символов ZX или ZY.

3

Текстовый файл 24-212.txt содержит строку из набора А, В, С, D, О, всего не более чем из 106 символов. Определите максимальное количество идущих подряд пар символов вида «согласная + гласная»

4

Текстовый файл 24-204.txt содержит строку из заглавных латинских букв А, В и С, всего не более чем из 106 символов. Найдите максимальное количество подряд идущих пар символов АА или СС. Искомая подстрока может включать только пары АА, только пары СС или содержать одновременно как пары АА, так и пары СС.

5

Текстовый файл 24-215.txt содержит строку из символов А, В, С и цифр 1, 2, 3, всего не более чем 106 символов. Определите максимальное количество идущих подряд троек символов вида «цифра + буква + цифра».

6

Текстовый файл 24-239.txt состоит не более чем из 106 символов и содержит только заглавные буквы латинского алфавита X, Y и Z. Найдите максимальную длину подстроки, которая состоит из сочетаний XY, YZ, YZZ, записанных в произвольном порядке. В ответе укажите наибольшую длину подходящей подстроки. Например, в строке ZXZXZZXYYZYZZYYY самая длинная подходящая подстрока – XYZYZZZ имеет длину 7.

7

Текстовый файл 24-298.txt состоит не более чем из 106 символов и содержит только символы, обозначающие знаки «-», «*», и цифры 0, 7, 8, 9. Определите в прилагаемом файле максимальное количество идущих подряд символов, которые образуют математически правильную последовательность, в которую входят знаки «-» или «*» и натуральные числа без незначащих нулей.

Текстовый файл 24-310.txt состоит не более чем из 106 символов и содержит только цифры шестнадцатеричной системы счисления, а также знаки «+» и «*» (сложения и умножения). Определите максимальное количество символов в непрерывной последовательности, которая является корректным арифметическим выражением с целыми неотрицательными числами, записанными в троичной системе счисления. В этом выражении никакие два знака арифметических операций не стоят рядом, в записи чисел отсутствуют незначащие (ведущие) нули и число 0 не имеет знака.

24

регулярные выражения
(задачи уровня “могила”)

0

Текстовый файл 24-299.txt состоит не более чем из 10^6 символов и содержит только десятичные цифры, а также знаки «+» и «*» (сложения и умножения). Определите максимальное количество символов в непрерывной последовательности, являющейся корректным арифметическим выражением с целыми неотрицательными числами (без знака), значение которого равно нулю. В этом выражении никакие два знака арифметических операций не стоят рядом, порядок действий определяется по правилам математики. В записи чисел отсутствуют незначащие (ведущие) нули. В ответе укажите количество символов в найденном выражении.

*

Текстовый файл 24-305.txt состоит не более чем из 106 символов и содержит только цифры шестнадцатеричной системы счисления, а также знаки «+» и «*» (сложения и умножения). Определите максимальное количество символов в непрерывной последовательности, которая начинается символами AF, за которыми следует правильное арифметическое выражение с целыми неотрицательными числами (без знака), записанными в десятичной системе счисления. В этом выражении никакие два знака арифметических операций не стоят рядом. В записи чисел отсутствуют незначащие (ведущие) нули. В ответе укажите количество символов в найденном выражении.

1

Текстовый файл 19969 состоит не более чем из 10^6 символов и содержит только строчные латинские буквы, а также знаки «@» и «.». Определите максимальное количество символов в непрерывной последовательности вида [буквы]@[буквы].[буквы] (например yasdamege@nasto.ballov). В ответе укажите количество символов.

В волшебном королевстве, где звезды искрятся на ночном небе, жил-был Числовой Заклинатель. Он умел создавать удивительные заклинания из чисел и знаков, используя только цифры 0, 2, 3, 4, 5 и знаки операций «+» (сложение) и «*» (умножение). Эти числа и знаки были ключами к его магии!

Однажды, наш Числовой Заклинатель нашел древний свиток, на котором были написаны только эти символы. Он решил выяснить, сколько символов в непрерывной последовательности на свитке могут образовать корректное арифметическое выражение с целыми неотрицательными числами.

3

Текстовый файл 19968 состоит не более чем из 106 символов и содержит только цифры шестнадцатеричной системы счисления, а также знаки «+» и «*» (сложения и умножения). Определите максимальное количество символов в непрерывной последовательности, которая является корректным арифметическим выражением с целыми неотрицательными числами, записанными в шестеричной системе счисления. В этом выражении никакие два знака арифметических операций не стоят рядом, в записи чисел отсутствуют незначащие (ведущие) нули и число 0 не имеет знака.

4

Текстовый файл 17685 состоит из десятичных цифр, знаков «+» и «*» (сложения и умножения). Определите максимальное количество символов в непрерывной последовательности, являющейся корректным арифметическим выражением с целыми неотрицательными числами (без знака), значение которого равно нулю. В этом выражении никакие два знака арифметических операций не стоят рядом, порядок действий определяется по правилам математики. В записи чисел отсутствуют незначащие (ведущие) нули.

В ответе укажите количество символов.

5

Текстовый файл 19970 состоит из десятичных цифр, знаков «+» и «*» (сложения и умножения). Определите максимальное количество символов в непрерывной последовательности, являющейся корректным арифметическим выражением с целыми неотрицательными числами, кратными пяти. В этом выражении никакие два знака арифметических операций не стоят рядом, порядок действий определяется по правилам математики. В записи чисел отсутствуют незначащие (ведущие) нули.

6

Текстовый файл 19149 состоит из символов ')', '(', знака '+' и цифр 1, 2, 3, 4. Определите максимальное количество символов в непрерывной последовательности, начинающихся с символа '(' и кончающихся символом ')', при этом последовательность является математически корректной и её значение кратно 2.

В этом выражении никакие два знака арифметических операций не стоят рядом, в начале и в конце математической последовательности не может стоять знак '+'.

В ответе укажите количество символов.