

Java

Регулярные выражения, Отладка

Регулярные выражения

- Шаблон для поиска в строке
- Присутствует во многих языках
- Сложно и писать и читать

Виды регулярных выражений

- POSIX (BRE)
- POSIX (ERE)
- PCRE
- Unicode

Базовые выражения

- Word
- `a|б` - или
- `(..)` - группа
- `^` - начало строки
- `$` - конец строки
- `\b \B` - граница слова или граница не слова
- `\r \n \t` - как в Java
- `\` - escape символ

Количество

- $a?$ - ноль или один раз
- a^* - от 0 до бесконечности
- a^+ - от 1 до бесконечности
- $a\{3\}$ - три раза
- $a\{3,\}$ - от трёх и более раз
- $a\{3, 6\}$ - от трёх до шести

Токены символов

- [abсaбв21] - один из перечисленных символов
- [^abc] - один из символов, кроме перечисленных
- [a-zA-Za-яA-ЯeЁ] - один из промежутка символов
- . - любой символ

Встроенные группы символов

- `\s` - whitespace (пробел, таб, перенос строки)
- `\S` - не whitespace
- `\d \D` - Любая цифра / не цифра
- `\w \W` - “символ слова”

Группы

- (..) - группа
- (?<name> ..) - именованная группа
- (?= ..) - Positive Lookahead, проверяет что вложенное выражение присутствует после основной части
- (?! ..) - Negative lookahead
- (?<= ..) - Positive Lookbehind
- (?<!= ..) - Negative Lookbehind

Флаги

- g - глобальный поиск (несколько результатов)
- m - многострочный, символы \$ и ^ считают построчно
- i - игнорировать регистр
- x - игнорировать whitespace в выражении
- s - воспринимать текст как одну строку
- u - поддержка символов юникода

Регулярные выражения в Java

- Хранится в объекте Pattern
- `Pattern.compile(literal, flags)` - создание выражения
- `Pattern.matches(regex, text)` - Однократная проверка
- `pattern.split(text)` - разбить по выражению

Matcher

- `pattern.matcher(text)` - обработка и получение результатов
- Методы:
 - `boolean find()` - поиск
 - `int start(), int end()` - индекс начала и конца совпадения
 - `String replaceFirst(replacement) / replaceAll(replacement)` - замена
 - `group(group) / groupCount()` - получение групп совпадений
 - `matches()` - совпадает ли всё выражение
 - `reset()` - сброс в начальное состояние

Литература

- Песочница - <https://regex101.com/>
- Задачки - <https://regexcrossword.com/>

Отладка

- Точки останова
- Условия остановки
- Удалённая отладка: Java Debug Wire Protocol
- java
-agentlib:jdwp=transport=dt_socket,server=y,suspend=n,address=8000
OurApplication