

Hausaufgaben 11

03./04.06.2019

Abgabe der Lösung am 10.06.2019

Aufgabe 1

Textsuche mit Sonderzeichen

Implementieren Sie die Methode

```
public static ArrayList<Integer> textSearch(String text, String pattern)
```

die ein übergebenes Pattern in einem übergebenen Text sucht und eine ArrayList zurückgibt, in der die Startindizes aller Vorkommen des Patterns im Text gespeichert sind. Die Vorkommen dürfen sich dabei auch überschneiden. Beispiel: Das Pattern xx kommt im Text xyxxx an den Stellen 2 und 3 vor.

Einige Zeichen haben dabei eine Sonderbedeutung. Folgende Sonderbedeutungen sollen implementiert werden:

- Ein Punkt (.) steht für ein beliebiges Zeichen.
- Eckige Klammern kennzeichnen Zeichenklassen. [abc] steht beispielsweise für „a, b oder c“. Zwischen den eckigen Klammern darf eine beliebige Anzahl von Zeichen stehen, von denen eines zutreffen muss.
 - Findet sich zu einer öffnenden eckigen Klammer kein schließendes Gegenstück, wird eine `PatternSyntaxException` geworfen. Diese Klasse ist in der Java-Bibliothek schon vorhanden.
 - In PCRE (Perl Compatible Regular Expressions) gibt es in eckigen Klammern weitere Sonderbedeutungen, z.B. für Zeichenbereiche. Diese sind nicht Teil der Aufgabe.
- Der Backslash (\) „maskiert“ das folgende Zeichen. Das folgende Zeichen hat keine Sonderbedeutung mehr, sondern wird als „normales“ Zeichen interpretiert. Maskiert werden können Punkte, eckige Klammern und der Backslash selbst.

Hinweise:

- Die Sonderbedeutung sind ein Teil der PCRE, die in Java z.B. in `String.split` oder in den Klassen `Pattern` und `Matcher` verwendet werden. Diese Möglichkeiten sollen in Java aber nicht genutzt werden.
- Ein einzelnes Backslash muss in Java stets "\\\" geschrieben werden, da es sonst als Einleitung einer Escape-Sequenz interpretiert wird.
- Gehen Sie davon aus, dass die Patterns die folgenden Bedingungen erfüllen:
 - Das letzte Zeichen ist kein Backslash.
 - Innerhalb einer Zeichenklasse steht keine eckige Klammer und kein Backslash. Es werden auch keine Zeichen maskiert.
 - Der Punkt hat innerhalb einer Zeichenklasse keine Sonderbedeutung (ist in PCRE auch so), muss also nicht maskiert werden.
 - In einer Zeichenklasse steht mindestens ein Zeichen.
 - Maskierte normale Zeichen werden einfach als normale Zeichen gewertet.

Testfälle:

```
System.out.println(textSearch("abcabcdababdc.", "ab")); //0, 3, 7, 9
System.out.println(textSearch("abcabcdababdc.", "c.")); //2, 5, 12
System.out.println(textSearch("abcabcdababdc.", "c\\.")); //12
System.out.println(textSearch("abcabcdababdc.", "b[cd]")); //1,4,10
System.out.println(textSearch("abcabcdababdc.", "a...c")); //0,7
System.out.println(textSearch("a[aababa][ab]a", "a[ab]a")); //3,5
System.out.println(textSearch("a[aababa][ab]a", "a\\.\\[a]")); //7
```