

SEW3 UE06 2024/25

Reguläre Ausdrücke II

Suchen

1. Schreibe die Methode

```
Map<String, Double> tagStatistik(Path htmlDatei) { ... }
```

Diese Methode soll im Ergebnis zu jedem HTML-Tag aus der `htmlDatei` die relative Häufigkeit des Vorkommens (bezogen auf die Gesamtanzahl der Tags) liefern, d.h. eine Zahl zwischen 0 und 1

- Zwischen Groß- und Klein-schreibung soll nicht unterschieden werden.

2. Schreibe die Methode

```
Set <String> links(Path htmlDatei) { ... }
```

Diese Methode soll alle absoluten Links liefern, die in der `htmlDatei` vorkommen, d.h. Links, die mit `http://` oder `https://` beginnen.

3. Schreibe die Methode

```
Set<String> domains(Path htmlDatei) { ... }
```

Diese Methode soll die Domains aller absoluten Links liefern, die in der `htmlDatei` vorkommen. Die Domain umfasst die beiden letzten Bestandteile des Hostnamens, z.B. ist *rennweg.at* die Domain von *http://www.htl.rennweg.at/index.html*.

4. Schreibe die Methode

```
Set<String> klassen(Path cssDatei) { ... }
```

Diese Methode soll alle CSS-Klassen liefern, auf die in der `cssDatei` Bezug genommen wird, d.h. Namen, die mit einem Punkt beginnen.

Suchen und Ersetzen

5. Schreibe drei Methoden, die mit Hilfe von RegEx ihr String-Argument in der folgenden Weise umformen und das Ergebnis als String liefern:

- a) Mehrfache Leerzeichen sollen durch jeweils ein einziges Leerzeichen ersetzt werden.
- b) Jedes Wort soll einen großen Anfangsbuchstaben bekommen.
- c) Der Text soll in Zeilen aufgeteilt werden, die max. 40 Zeichen lang sind. Die Zeilentrennungen dürfen nur bei Leerzeichen erfolgen.