

Aufgabenblatt 4

Aufgabe (1)

[3 Punkte]

Modifizieren und erweitern Sie die Klasse **Dokument** (s. Aufgabenblatt (2)) in folgender Weise:

1. Nennen Sie die Instanzvariable `@inhalt` in `@text` um.
2. Fügen Sie die Instanzvariablen `@token_list` und `@bekannte_abkuerzungen` hinzu und stellen Sie sicher, dass bei der Erzeugung einer Instanz der Dokument-Klasse `@text` mit den Dokument (Typ: String) und `@token_list` mit der Liste der Token des Dokuments (Typ: Array) initialisiert werden.
3. Weisen Sie bei der Initialisierung der Variablen `@bekannte_abkuerzungen` eine Liste mit 10 hochfrequenten Abkürzungen zu.

Aufgabe (2)

[4 Punkte]

Ergänzen Sie die Klasse **Dokument** durch die Methode **tokenisiere**, die ein Token (Typ: String) als Argument nimmt und als Wert einen Array liefert. Die Methode soll mit Ausnahme des Punktes alle Sonderzeichen vom Token entfernen. Der Array enthält das Ergebnis dieser Nachtokenisierung.

Beispiel

```
test = Dokument.new('Ein sehr kleines Dokument.')
p test.tokenisiere('Haus')
p test.tokenisiere('Haus.')
p test.tokenisiere('(Haus.)')
p test.tokenisiere('{(z.B.)}')
p test.tokenisiere('z.B.')
['Haus']
['Haus.']
['(', 'Haus.', ')']
['{', '(', 'z.B.', ')']
['z.B.']
```

Aufgabe (3)

[3 Punkte]

Entwickeln Sie eine Reihe sinnvoller UNIT-Tests (separate Datei) für ihr Programm.