Computerlinguistische Programmiertechnik Beliankou • Beyer • Naumann WS 2016

Aufgabenblatt (2)

Aufgabe (1) [3 Punkte]

Die Funktion *slurp* kann verwendet werden, um den Inhalt einer Textdatei zu lesen. Sie akzeptiert einen Dateibezeichner als Argument und liefert als Wert einen String, der den Inhalt der Datei enthält.

Definieren Sie eine Variable text. Initialisieren Sie die Variable mit der leeren Zeichenkette und versehen Sie sie mit einem sinnvollen Doc-String. Definieren Sie außerdem die Funktion lese-datei, die den Inhalt einer Textdatei ausliest und als String in der Variable text speichert.

Aufgabe (2) [3 Punkte]

Definieren Sie die Funktion abkürzung, die ein Token (String) als Argument nimmt und prüft, ob es sich bei dem Token um eine Abkürzung handelt. Verwenden Sie folgende Kriterien:

- 1. Das Token ist Element der internen Liste hochfrequenter Abkürzungen.
- 2. Das Token enthält mehr als einen Punkt.
- 3. Das Token endet mit einem Punkt und enthält keinen Vokal.

Aufgabe (3) [4 Punkte]

Definieren Sie die Funktion text-2-tokenliste, die einen Text (String) als Argument nimmt und als Wert eine Liste aller Token dieses Textes liefert.:

Beispiel

- > (text-2-tokenliste "Ein erster Satz. Und noch ein Satz!") ⇒ ("Ein" "erster" "Satz" "." "Und" "noch" "ein" "Satz'*!")
- Aufgabe (4) [4 Punkte]

Definieren Sie die Funktion generiere-satzliste, die eine Liste von Token als Argument nimmt und als Wert eine Liste aller Sätze dieses Textes liefert.:

Beispiel

```
> (generiere-satzliste ("Ein" "erster" "Satz" "." "Und" "noch" "ein" "Satz""!")) \Rightarrow (("Ein" "erster" "Satz" ".") ("Und" "noch" "ein" "Satz"*!"))
```