

Übungsblatt 10

Abgabe: Aufgabe 1 am 26.1.2009

Aufgabe 2 am 19.1.2009

Version1.2

Aufgabe 1 Theorie (70 %)

Zeigt die Korrektheit der Implementierung der `append` - Methode für die in `MyList.java` gegebenen Ringlisten. Folgt dabei der semantischen Modellierung von Listen aus der Vorlesung vom 12. Januar, der Vorlesung vom 14. Januar und dem Skript `listen.pdf`, insbesondere dem dort angegebenen Beispielbeweis für `delete`.

Aufgabe 2 und Praxis (30 %)

Man sagt, der Wortschatz, der in der Tagesschau verwendet wird, würde nur etwa 500 Wörter umfassen. Es ist also Zeit, die Wortgewalt eurer Aufgabenlösungen zu testen. Wie sich das für angehende Informatiker gehört, lässt man diese natürlich automatisch ermitteln. Daher ist ein Programm zu erstellen, das einen Text einliest und eine Statistik über die Häufigkeit der darin enthaltenen Wörter ausgibt. Ein Wort ist eine Folge von Zeichen, für die `Character.isLetter(char)` gilt.

Aufgabe 2.1 Statistik

Die Statistik soll (natürlich) in einer Ringliste gespeichert werden. Die Liste enthält Einträge, die zum einen das jeweilige Wort enthalten, zum anderen die Anzahl, in der dieses Wort im Text vorkommt, der Datentyp `Rlist` muss also geeignet ergänzt werden. Um die Ausgabe zu vereinfachen, soll die Liste nach der Häufigkeit der Wörter sortiert sein, wobei die häufigsten Wörter immer am Anfang stehen.

Jedes Mal, wenn ein neues Wort in die Liste eingetragen werden soll, wird zuerst ermittelt, ob dieses bereits in der Liste vorkommt. Ist dies der Fall, wird sein Eintrag aus der Liste entfernt, seine Häufigkeit um eins erhöht und dann der Eintrag entsprechend der neuen Häufigkeit wieder eingefügt. Kam ein Wort bisher noch nicht vor, wird ein neuer Eintrag erzeugt und ebenfalls in die Liste einsortiert. Dadurch ist die Liste stets nach der Häufigkeit sortiert und häufige Wörter stehen am Anfang, wo sie schnell gefunden werden.

Aufgabe 2.2 Wörter zählen

Schreibt nun eine Methode, die eine Textdatei einliest, in Worte zerlegt und diese in die Statistik einträgt. Lasst die Statistik am Ende ausgeben, aber beschränkt Euch in der Abgabe auf die 20 häufigsten Wörter.

Testet mit dem \LaTeX -Quelltext eurer Aufgabenlösung.

Hinweise: Die Klasse `MyReader.java` enthält ein Codefragment, mit dem die Zeichen einer Datei eingelesen und weiterverarbeitet werden können.