

Übung 7 – AVL-Bäume und B-Bäume

1. Testen Sie Ihre AVL-Baum Implementierungen mit dem Shakespeare Beispiel aus Übung 3. Um die Worte in Shakespeares Werk zu zählen fügen Sie diese in einen Baum ein. Als Schlüssel wird das Wort selber verwendet, als Element können Sie *null* abspeichern. Es muss sichergestellt sein, dass pro Schlüssel nur ein Element abgespeichert wird.

Als weitere Optimierung können Sie versuchen, (regelmässige) Pluralformen zu eliminieren. Dies kann z.B. erreicht werden indem die Wortmenge nach dem Einfügen aller Worte einmal inorder (sortiert) traversiert wird und jeweils aufeinanderfolgende Worte miteinander verglichen werden. Falls für die beiden Worte *prev* und *curr* die Bedingung

```
curr.length() == prev.length() + 1
&& curr.startsWith(prev) && curr.charAt(curr.length()-1)=='S'
```

gilt so kann eine Form gelöscht werden (wobei so auch Worte gelöscht werden die keine Pluralformen sind, wie z.B. IT und ITS).

Geben Sie neben der Anzahl unterschiedlicher Worte auch die Höhe des Binärbaumes aus.

2. In einen anfänglich leeren B-Baum der Ordnung $n = 2$ werden folgende Elemente eingefügt:

13, 97, 45, 23, 103, 84, 100, 111, 44, 12, 33, 66, 99, 11, 1, 96, 101, 5

a) Bauen Sie den Baum auf und zeichnen Sie das Resultat.

b) Löschen Sie aus diesem Baum die beiden Elemente 100 und 1.

Die Resultate dieser Aufgabe können mit dem B-Baum Applet überprüft werden.

3. Bestimmen Sie die minimale Anzahl Elemente in einem B-Baum der Ordnung n und Höhe h .

Wird diese Formel nach h aufgelöst so erhalten Sie eine obere Schranke für die Höhe eines B-Baumes der Ordnung n mit N Elementen.

4. Gegeben seien 20'000 Datensätze mit je
 - einem Schlüssel von 10 Bytes und
 - zusätzlich Daten von 200 Bytes.

Der Zugriff auf diese Daten soll mit einem B-Baum realisiert werden, d.h. die Knoten sollen sowohl die Referenzen auf die Seiten mit den Daten als auch auf die Nachfolgeseiten enthalten. Die Referenz auf eine Seite ist 2 Bytes gross.

a) Bestimmen Sie die Ordnung des B-Baumes wenn die Seiten 2048 Bytes gross sind und diese möglichst optimal genutzt werden.

b) Wie hoch ist die maximal benötigte Anzahl von Seitenzugriffen für das Lesen eines Datensatzes (Beachten Sie dass die Daten ebenfalls mit einem Seitenzugriff geladen werden müssen).