

Übungsblatt 8: "Indexstrukturen und Hashing"

Abgabe bis Montag, 8. Juli 2019, entweder zu **Beginn** der Vorlesung, oder bis 14:00 Uhr
in den Briefkästen vor dem Sekretariat Informatik (Raum 1/308) im Mathematikum

Aufgabe 8-1 Erweiterbares Hashing

6 + 4 = 10 Punkte

Gegeben ist der folgende Limerick.

*JavaScript was made for the masses,
It lacked types, and modules, and classes.
But it became quite the giant,
Because it ran on server and client,
Until it crashed both, despite 100 test passes.*

Die Hash-Werte $h(w)$ der unterstrichenen Wörter w sind in der nachfolgenden Tabelle angegeben. Stopworte (deren Hashwerte nicht angegeben sind) können Sie im Folgenden ignorieren.

w	$h(w)$	w	$h(w)$
JavaScript	00001101	ran	10001010
made	00010101	test	10011101
lacked	00101111	client	10100111
modules	00110000	types	10110010
classes	01001111	server	11000001
giant	01011110	passes	11100101
masses	01101001	became	11111011
crashed	01111100		

1. Fügen Sie die Worte des Limerick in der Reihenfolge, in der sie auftauchen, in eine **erweiterbare Hash-Tabelle** mit einer Bucketgröße von 2 ein.

Verwenden Sie hierzu die Hash-Werte $h(w)$ aus der obigen Tabelle. Benutzen Sie zu Beginn zwei Bits des Hashwertes (beginnen Sie mit den zwei Bits auf der **linken Seite** und erweitern Sie nach rechts).

Wie sehen das Hash-Verzeichnis, das Mapping, und die Inhalte der Buckets aus, wenn die folgenden Zeitpunkte betrachtet werden?

- a) Am Ende der zweiten Zeile.
 - b) Am Ende der fünften Zeile.
2. Löschen Sie nun die Wörter der zweiten und fünften Zeile aus der in Teilaufgabe 1 erstellten Hash-Tabelle. Geben Sie das aktualisierte Hash-Verzeichnis, das Mapping, sowie die Inhalte der Buckets an. Was fällt Ihnen auf? Geben sie je einen Vor- und Nachteil einer möglichen Implementierung an, die dieses Phänomen in der Praxis berücksichtigt.

Aufgabe 8-2 Eigenschaften von B^+ -Bäumen

2 + 2 + 3 + 3 = 10 Punkte

In der Praxis sind bei der Implementierung von B^+ -Bäumen einige Parameter relevant, die sich auf die Größe des resultierenden Baumes auswirken. Dazu gehören der Speicherverbrauch z von Zeigern, die Speichergröße s von Seiten, der Speicherverbrauch k der Zugriffsattribute und die Größe d der im Baum gespeicherten Datensätze.

1. Leiten Sie eine Formel für die Ordnung y der Blattknoten in einem B^+ -Baum her, in Abhängigkeit von den oben aufgeführten Parametern. Nehmen Sie dazu an, dass Blätter als doppelt verlinkte Liste implementiert sind.
2. Leiten Sie eine Formel für die Ordnung x der inneren Knoten eines B^+ -Baumes her.
3. Geben Sie Abschätzungen für die minimale und maximale Höhe h eines B^+ -Baumes mit Ordnung (x, y) an, der n Datensätze enthält.
4. Facebook hat derzeit ca. 2,7 Milliarden Nutzer (Stand April 2019). Nehmen Sie an, dass die Nutzerdaten in einem B^+ -Baum verwaltet werden. Für jeden Nutzer ist der Vor- und Nachname, der Geburtstag, sowie einen Verweis auf weitere Daten (wie z.B. Profilbilder und Videos) gespeichert (durchschnittlich ca. 900 Byte). Nehmen sie weiter an, dass die Zeigergröße $z = 8$ Byte, die Länge des verwendeten Zugriffsattributs $k = 8$ Byte, und die Größe einer Seite $s = 4$ KByte betragen. Bestimmen Sie die Ordnung (x, y) des B^+ -Baumes und schätzen Sie dessen Höhe h ab.

Aufgabe 8-3 Vergleich von Indexstrukturen

1 + 1 + 1 + 1 = 4 Punkte

Die korrekte Wahl eines Index entscheidet oft über die Performanz und Flexibilität des Datenbanksystems, gerade bei systemkritischen Tabellen. Überlegen Sie sich daher im Folgenden, welchen Vorteil die jeweils zuerst genannten Strukturen gegenüber der zweiten Lösung haben, und wann diese relevant werden. Betrachten Sie dabei insbesondere die Aspekte der Suche, sowie insert-Häufigkeit oder Verteilung der Werte. Nennen Sie jeweils mindestens einen Vorteil.

1. B^+ -Bäume gegenüber B-Bäumen,
2. B-Bäume gegenüber B^+ -Bäumen,
3. B^+ -Bäume gegenüber Hashing,
4. Hashing gegenüber B^+ -Bäumen.

Informationen zur Abgabe. Die Aufgaben können in Gruppen bis zu **drei** Studierende bearbeitet und abgegeben werden. Bitte schreiben Sie die Namen aller Mitglieder ihrer Gruppe sowie die Nummer ihrer Übungsgruppe (1, 2 oder 3) auf das Frontblatt ihrer Abgabe! Zur Erinnerung, hier die Übungsgruppen:

- Gruppe 1: Donnerstag, 16:00 - 18:00 Uhr
Gruppe 2: Freitag, 11:00 - 13:00 Uhr
Gruppe 3: Freitag, 14:00 - 16:00 Uhr

Schreiben Sie klar und deutlich und verwenden Sie keinen Bleistift. **Tackern** Sie Ihre Lösungsblätter zusammen (keine Büroklammern, keine Origami-Kunstwerke). Sie können die Lösungen in dem ihrer Übungsgruppe entsprechenden Briefkasten vor dem Sekretariat Informatik (Raum 1/308) im Gebäude INF 205 (Mathematik) einwerfen oder zu Beginn der Vorlesung abgeben. Eine elektronische Abgabe ist prinzipiell nicht möglich.