

Hashing

Es ist die Wirksamkeit verschiedener Schlüsseltransformationen zu untersuchen. Dazu erstelle man ein Programmsystem, mit welchem

1. verschiedene Transformationsfunktionen
2. verschiedene Kollisionsbehandlungen

realisiert werden.

Die Schlüssel seien als ASCII-String der Länge n vorgegeben.

Als Transformationsfunktion wähle man:

1. Aufsummierung der Elemente des Strings
2. Aufsummierung mit Gewichtungsbewertung der Stringelemente
3. Aufsummierung der ersten m Elemente des Strings

Für die Kollisionsbehandlung sind die folgenden Verfahren zu wählen:

1. gestreute Verkettung (Benutzung einer Überlaufliste)
2. lineares Austesten (offene Adressierung)
 - a. mit linearer Sondierung $adr = (adr + 1) \% Tabl$
 - b. mit linearer Sondierung $adr = (adr + n) \% Tabl$, $n \dots$ Kollisionszähler
 - c. mit quadratischer Sondierung $adr = (adr + n^2) \% Tabl$
3. doppeltes Hashing

Dies ergibt 15 Kombinationen, die auszuwerten sind.

Zur Auswertung trage man die Anzahl benötigter Zugriffe über dem Füllgrad auf:

$$\text{Füllgrad}(\%) = f(\text{Anzahl Zugriffe})$$

Zur Erinnerung: Auslastung = Anzahl belegter Plätze : Anzahl freier Plätze
 Füllgrad = Anzahl belegter Plätze : Gesamtanzahl der Plätze

Die Schlüssel sind mit einem Zufallszahlengenerator zu erzeugen. Jeder Kombination Transformationsfunktion-Kollisionsfunktion müssen aber die **gleichen** zufälligen Strings zugrunde liegen!

Berechnen Sie **10 Punkte** für jedes Diagramm. Diese Punkte erreichen Sie, wenn Sie für eine gegebene Anzahl der Schlüssel die Größe der Bucket-Feldes variieren (die Feldlänge konstant zu halten und die Anzahl der Schlüssel zu variieren ist die schlechtere Möglichkeit, da dabei den einzelnen Punkten nicht die gleichen Datenbestände zugrunde liegen)

Beachten Sie die **richtige Kollisionszählung** bei der **gestreuten Verkettung**:

$$\text{Anzahl Listenelemente} = \text{Anzahl Zugriffe} = \text{Anzahl Kollisionen} + 1$$

Stellen Sie die Funktionen f grafisch mit **Excel** oder **OpenOffice Spreadsheet** dar. Erstellen Sie 3 Diagramme (für jede Hash-Funktion ein), jeweils mit 5 Kurven (für jede Kollisionsfunktion eine). Lassen Sie Ihr Programm eine **Textdatei** mit entsprechenden Daten erstellen, die sie dann in Excel oder OpenOffice Spreadsheet importieren.