

## Script generated by TTT

Title: Täubig: GAD (13.05.2014)

Date: Tue May 13 13:53:07 CEST 2014

Duration: 132:02 min

Pages: 55

## Übersicht

- 5 Hashing
  - Hashtabellen
  - Hashing with Chaining
  - Universelles Hashing
  - Hashing with Linear Probing
  - Anpassung der Tabellengröße
  - Perfektes Hashing
  - Diskussion / Alternativen

## Dynamisches Wörterbuch

Problem: Hashtabelle ist **zu groß** oder **zu klein**  
(sollte nur um konstanten Faktor von Anzahl der Elemente abweichen)

Lösung: Reallokation

- Wähle geeignete Tabellengröße
- Wähle neue Hashfunktion
- Übertrage Elemente auf die neue Tabelle

## Dynamisches Wörterbuch

Problem: Tabellengröße  $m$  sollte **prim** sein  
(für eine gute Verteilung der Schlüssel)

Lösung:

- Für jedes  $k$  gibt es eine Primzahl in  $[k^3, (k+1)^3]$
- Jede Zahl  $z \leq (k+1)^3$ , die nicht prim ist, muss einen Teiler  $t \leq \sqrt{(k+1)^3} = (k+1)^{3/2}$  haben.
- Für eine gewünschte ungefähre Tabellengröße  $m'$  (evt. nicht prim) bestimme  $k$  so, dass  $k^3 \leq m' \leq (k+1)^3$

## Dynamisches Wörterbuch

Problem: Tabellengröße  $m$  sollte **prim** sein (für eine gute Verteilung der Schlüssel)

Lösung:

- Für jedes  $k$  gibt es eine Primzahl in  $[k^3, (k+1)^3]$
- Jede Zahl  $z \leq (k+1)^3$ , die nicht prim ist, muss einen Teiler  $t \leq \sqrt{(k+1)^3} = (k+1)^{3/2}$  haben.
- Für eine gewünschte ungefähre Tabellengröße  $m'$  (evt. nicht prim) bestimme  $k$  so, dass  $k^3 \leq m' \leq (k+1)^3$
- Größe des Intervalls:  
 $(k+1)^3 - k^3 + 1 = (k^3 + 3k^2 + 3k + 1) - k^3 + 1 = 3k^2 + 3k + 2$
- Für jede Zahl  $j = 2, \dots, (k+1)^{3/2}$ :  
 streiche die Vielfachen von  $j$  in  $[k^3, (k+1)^3]$
- Für jedes  $j$  kostet das Zeit  $((k+1)^3 - k^3 + 1) / j \in O(k^2/j)$

## Dynamisches Wörterbuch

- Hilfsmittel: Wachstum der harmonischen Reihe

$$\ln n \leq H_n = \sum_{i=1}^n \frac{1}{i} \leq 1 + \ln n$$

- insgesamt:

$$\begin{aligned} \sum_{2 \leq j \leq (k+1)^{3/2}} O\left(\frac{k^2}{j}\right) &\leq k^2 \sum_{2 \leq j \leq (k+1)^{3/2}} O\left(\frac{1}{j}\right) \\ &\in k^2 \cdot O(\ln((k+1)^{3/2})) \\ &\in O(k^2 \ln k) \\ &\in \underline{o(m)} \end{aligned}$$

⇒ Kosten zu vernachlässigen im Vergleich zur Initialisierung der Tabelle der Größe  $m$  (denn  $m$  ist kubisch in  $k$ )

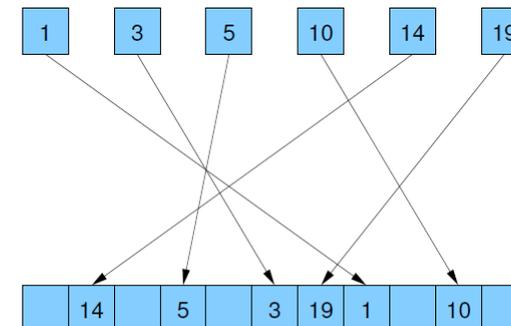
## Übersicht

### 5 Hashing

- Hashtabellen
- Hashing with Chaining
- Universelles Hashing
- Hashing with Linear Probing
- Anpassung der Tabellengröße
- Perfektes Hashing
- Diskussion / Alternativen

## Perfektes Hashing für statisches Wörterbuch

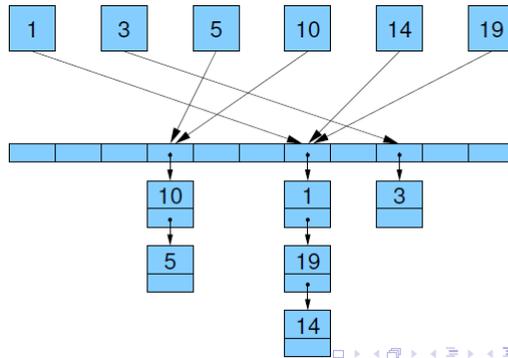
- bisher: konstante *erwartete* Laufzeit  
 falls  $m$  im Vergleich zu  $n$  genügend groß gewählt wird  
 (nicht ausreichend für Real Time Scenario)
- Ziel: konstante Laufzeit im **worst case** für find()  
 durch perfekte Hashtabelle ohne Kollisionen
- Annahme: statische Menge  $S$  von  $n$  Elementen



## Statisches Wörterbuch

- $S$ : feste Menge von  $n$  Elementen mit Schlüssel  $k_1$  bis  $k_n$
- $H_m$ :  $c$ -universelle Familie von Hashfunktionen auf  $\{0, \dots, m-1\}$  (Hinweis: 2-universelle Familien existieren für alle  $m$ )
- $C(h)$  für  $h \in H_m$ : Anzahl Kollisionen in  $S$  für  $h$ , d.h.

$$C(h) = |\{(x, y) : x, y \in S, x \neq y, h(x) = h(y)\}|$$



Beispiel:

$$C(h) = 2 + 6 = 8$$

## Erwartete Anzahl von Kollisionen

### Lemma

Für die Anzahl der Kollisionen gilt:  $\mathbb{E}[C(h)] \leq cn(n-1)/m$ .

## Erwartete Anzahl von Kollisionen

### Lemma

Für die Anzahl der Kollisionen gilt:  $\mathbb{E}[C(h)] \leq cn(n-1)/m$ .

### Beweis.

- Definiere  $n(n-1)$  Indikator-Zufallsvariablen  $X_{ij}(h)$ :  
Für  $i \neq j$  sei  $X_{ij}(h) = 1 \Leftrightarrow h(k_i) = h(k_j)$ .
- Dann ist  $C(h) = \sum_{i \neq j} X_{ij}(h)$

$$\mathbb{E}[C] = \mathbb{E}\left[\sum_{i \neq j} X_{ij}\right] = \sum_{i \neq j} \mathbb{E}[X_{ij}] = \sum_{i \neq j} \Pr[X_{ij} = 1] \leq n(n-1) \cdot c/m$$

⇒ Für **quadratische Tabellengröße** ist die erwartete Anzahl von Kollisionen (und damit die erwartete worst-case-Laufzeit für find) eine **Konstante**. □

## Markov-Ungleichung

### Satz (Markov-Ungleichung)

Für jede nichtnegative Zufallsvariable  $X$  und Konstante  $k > 0$  gilt:

$$\Pr[X \geq k] \leq \frac{\mathbb{E}[X]}{k} \quad \text{und für } \mathbb{E}[X] > 0: \Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$$

## Markov-Ungleichung

### Satz (Markov-Ungleichung)

Für jede nichtnegative Zufallsvariable  $X$  und Konstante  $k > 0$  gilt:

$$\Pr[X \geq k] \leq \frac{\mathbb{E}[X]}{k} \quad \text{und für } \mathbb{E}[X] > 0: \quad \Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$$

### Beweis.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{z \in \mathbb{R}} z \cdot \Pr[X = z] \\ &\geq \sum_{z \geq k \cdot \mathbb{E}[X]} z \cdot \Pr[X = z] \geq \sum_{z \geq k \cdot \mathbb{E}[X]} k \cdot \mathbb{E}[X] \cdot \Pr[X = z] \\ &\geq k \cdot \mathbb{E}[X] \cdot \Pr[X \geq k \cdot \mathbb{E}[X]] \end{aligned}$$

□

## Hashfunktionen mit wenig Kollisionen

### Lemma

Für mindestens *die Hälfte* der Funktionen  $h \in H_m$  gilt:

$$C(h) \leq 2cn(n-1)/m$$

## Hashfunktionen mit wenig Kollisionen

### Lemma

Für mindestens *die Hälfte* der Funktionen  $h \in H_m$  gilt:

$$C(h) \leq 2cn(n-1)/m$$

### Beweis.

- Aus Lemma  $\mathbb{E}[C(h)] \leq cn(n-1)/m$  und Markov-Ungleichung  $\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$  folgt für  $n \geq 2$ :

$$\Pr[C(h) \geq 2cn(n-1)/m] \leq \Pr[C(h) \geq 2\mathbb{E}[C(h)]] \leq \frac{1}{2}$$

- ⇒ Für höchstens die Hälfte der Funktionen ist  $C(h) \geq \frac{2cn(n-1)}{m}$
- ⇒ Für mindestens die Hälfte der Funktionen ist  $C(h) \leq \frac{2cn(n-1)}{m}$  ( $n \geq 1$ )

□

## Hashfunktionen mit wenig Kollisionen

### Lemma

Für mindestens *die Hälfte* der Funktionen  $h \in H_m$  gilt:

$$C(h) \leq 2cn(n-1)/m$$

### Beweis.

- Aus Lemma  $\mathbb{E}[C(h)] \leq cn(n-1)/m$  und Markov-Ungleichung  $\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$  folgt für  $n \geq 2$ :

$$\Pr[C(h) \geq 2cn(n-1)/m] \leq \Pr[C(h) \geq 2\mathbb{E}[C(h)]] \leq \frac{1}{2}$$

- ⇒ Für höchstens die Hälfte der Funktionen ist  $C(h) \geq \frac{2cn(n-1)}{m}$
- ⇒ Für mindestens die Hälfte der Funktionen ist  $C(h) \leq \frac{2cn(n-1)}{m}$  ( $n \geq 1$ )

□

## Hashfunktionen ohne Kollisionen

## Lemma

Wenn  $m \geq cn(n-1) + 1$ , dann bildet mindestens die Hälfte der Funktionen  $h \in H_m$  die Schlüssel **injektiv** in die Indexmenge der Hashtabelle ab.

## Beweis.

- Für mindestens die Hälfte der Funktionen  $h \in H_m$  gilt  $C(h) < 2$ .
  - Da  $C(h)$  immer eine gerade Zahl sein muss, folgt aus  $C(h) < 2$  direkt  $C(h) = 0$ .
- ⇒ keine Kollisionen (bzw. injektive Abbildung)

□

- Wähle zufällig  $h \in H_m$  mit  $m \geq cn(n-1) + 1$
  - Prüfe auf Kollision ⇒ behalten oder erneut wählen
- ⇒ Nach durchschnittlich 2 Versuchen erfolgreich

## Statisches Wörterbuch

Ziel: lineare Tabellengröße

Idee: **zweistufige** Abbildung der Schlüssel

- Wähle Hashfunktion  $h$  mit wenig Kollisionen ( $\approx 2$  Versuche)
- ⇒ 1. Stufe bildet Schlüssel auf Buckets von konstanter durchschnittlicher Größe ab
- Wähle Hashfunktionen  $h_\ell$  ohne Kollisionen ( $\approx 2$  Versuche pro  $h_\ell$ )
- ⇒ 2. Stufe benutzt quadratisch viel Platz für jedes Bucket, um alle Kollisionen aus der 1. Stufe aufzulösen

## Statisches Wörterbuch

- $B_\ell^h$ : Menge der Elemente in  $S$ , die  $h$  auf  $\ell$  abbildet,  $\ell \in \{0, \dots, m-1\}$  und  $h \in H_m$
- $b_\ell^h$ : Kardinalität von  $B_\ell^h$ , also  $b_\ell^h := |B_\ell^h|$
- Für jedes  $\ell$  führen die Schlüssel in  $B_\ell^h$  zu  $b_\ell^h(b_\ell^h - 1)$  Kollisionen
- Also ist die Gesamtzahl der Kollisionen

$$C(h) = \sum_{\ell=0}^{m-1} b_\ell^h(b_\ell^h - 1)$$

## Perfektes statisches Hashing: 1. Stufe

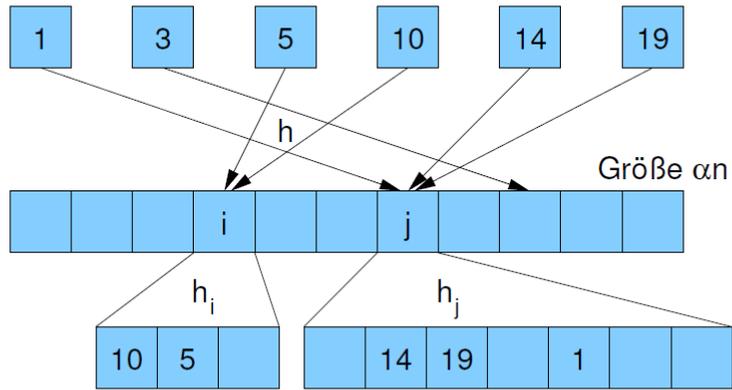
- 1. Stufe der Hashtabelle soll **linear** viel Speicher verwenden, also  $\lceil \alpha n \rceil$  Adressen, wobei wir Konstante  $\alpha$  später festlegen.
- Wähle Funktion  $h \in H_{\lceil \alpha n \rceil}$ , um  $S$  in Teilmengen  $B_\ell$  aufzuspalten. Wähle  $h$  dabei so lange zufällig aus  $H_{\lceil \alpha n \rceil}$  aus, bis gilt:

$$C(h) \leq \frac{2cn(n-1)}{\lceil \alpha n \rceil} \leq \frac{2cn(n-1)}{\alpha n} \leq \frac{2cn}{\alpha}$$

Da das für mindestens die Hälfte der Hashfunktionen in  $H_{\lceil \alpha n \rceil}$  gilt (vorletztes Lemma), erwarten wir dafür  $\leq 2$  Versuche.

- Für jedes  $\ell \in \{0, \dots, \lceil \alpha n \rceil - 1\}$  seien  $B_\ell$  die Elemente, die durch  $h$  auf Adresse  $\ell$  abgebildet werden und  $b_\ell = |B_\ell|$  deren Anzahl.

### Perfektes statisches Hashing

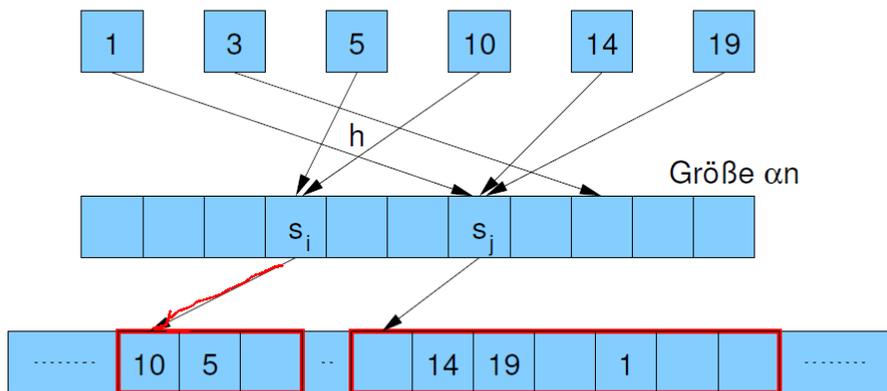


### Perfektes statisches Hashing: 2. Stufe

- Für jedes  $B_\ell$ :  
 Berechne  $m_\ell = c b_\ell (b_\ell - 1) + 1$ .  
 Wähle zufällig Funktion  $h_\ell \in H_{m_\ell}$ , bis  $h_\ell$  die Menge  $B_\ell$  injektiv in  $\{0, \dots, m_\ell - 1\}$  abbildet (also ohne Kollisionen).  
 Mindestens die Hälfte der Funktionen in  $H_{m_\ell}$  tut das.
- Hintereinanderreihung der einzelnen Tabellen ergibt eine Gesamtgröße der Tabelle von  $\sum_\ell m_\ell$
- Teiltabelle für  $B_\ell$  beginnt an Position  $s_\ell = m_0 + m_1 + \dots + m_{\ell-1}$  und endet an Position  $s_\ell + m_\ell - 1$
- Für gegebenen Schlüssel  $x$ , berechnen die Anweisungen  

$$\ell = h(x); \quad \text{return } s_\ell + h_\ell(x);$$
 dann eine injektive Funktion auf der Menge  $S$ .

### Perfektes statisches Hashing



### Perfektes statisches Hashing

- Die Funktion ist beschränkt durch:  

$$\sum_{\ell=0}^{\lceil \alpha n \rceil - 1} m_\ell = \sum_{\ell=0}^{\lceil \alpha n \rceil - 1} (c \cdot b_\ell (b_\ell - 1) + 1) \quad (\text{siehe Def. der } m_\ell\text{'s})$$

$$\leq c \cdot C(h) + \lceil \alpha n \rceil$$

$$\leq c \cdot \frac{2cn}{\alpha} + \alpha n + 1$$

$$\leq (2c^2/\alpha + \alpha)n + 1$$
- Zur Minimierung der Schranke betrachte die Ableitung  

$$f(\alpha) = (2c^2/\alpha + \alpha)n + 1$$

$$f'(\alpha) = (-2c^2/\alpha^2 + 1)n$$

$$\Rightarrow f'(\alpha) = 0 \text{ liefert } \alpha = \sqrt{2c}$$

$$\Rightarrow \text{Adressbereich: } 0 \dots 2\sqrt{2}cn$$

## Perfektes statisches Hashing

### Satz

Für eine beliebige Menge von  $n$  Schlüsseln kann eine perfekte Hashfunktion mit Zielmenge  $\{0, \dots, 2\sqrt{2}cn\}$  in linearer erwarteter Laufzeit konstruiert werden.

- Da wir wissen, dass wir für beliebige  $m$  eine 2-universelle Familie von Hashfunktionen finden können, kann man also z.B. eine Hashfunktion mit Adressmenge  $\{0, \dots, 4\sqrt{2}n\}$  in linearer erwarteter Laufzeit konstruieren. (Las Vegas-Algorithmus)
  - Unsere Minimierung hat nicht den Platz berücksichtigt, der benötigt wird, um die Werte  $s_\ell$ , sowie die ausgewählten Hashfunktionen  $h_\ell$  zu speichern.
- ⇒ Berechnung von  $\alpha$  sollte eigentlich angepasst werden (entsprechend Speicherplatz pro Element, pro  $m_\ell$  und pro  $h_\ell$ )

## Perfektes dynamisches Hashing

Kann man perfekte Hashfunktionen auch **dynamisch** konstruieren?

ja, z.B. mit **Cuckoo** Hashing

- 2 Hashfunktionen  $h_1$  und  $h_2$
- 2 Hashtabellen  $T_1$  und  $T_2$
- bei find und remove jeweils in beiden Tabellen nachschauen
- bei insert abwechselnd beide Tabellen betrachten, das zu speichernde Element an die Zielposition der aktuellen Tabelle schreiben und wenn dort schon ein anderes Element stand, dieses genauso in die andere Tabelle verschieben usw.
- evt. Anzahl Verschiebungen durch  $2 \log n$  beschränken, um Endlosschleife zu verhindern (ggf. kompletter Rehash mit neuen Funktionen  $h_1, h_2$ )

## Perfektes dynamisches Hashing

Kann man perfekte Hashfunktionen auch **dynamisch** konstruieren?

ja, z.B. mit **Cuckoo** Hashing

- 2 Hashfunktionen  $h_1$  und  $h_2$
- 2 Hashtabellen  $T_1$  und  $T_2$
- bei find und remove jeweils in beiden Tabellen nachschauen
- bei insert abwechselnd beide Tabellen betrachten, das zu speichernde Element an die Zielposition der aktuellen Tabelle schreiben und wenn dort schon ein anderes Element stand, dieses genauso in die andere Tabelle verschieben usw.
- evt. Anzahl Verschiebungen durch  $2 \log n$  beschränken, um Endlosschleife zu verhindern (ggf. kompletter Rehash mit neuen Funktionen  $h_1, h_2$ )

## Probleme beim linearen Sondieren

- Offene Hashverfahren allgemein:  
Erweiterte Hashfunktion  $h(k, i)$  gibt an, auf welche Adresse ein Schlüssel  $k$  abgebildet werden soll, wenn bereits  $i$  Versuche zu einer Kollision geführt haben
- Lineares Sondieren (Linear Probing):

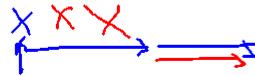
$$h(k, i) = (\underline{h(k)} + \underline{i}) \bmod m$$

## Probleme beim linearen Sondieren

- Offene Hashverfahren allgemein:

Erweiterte Hashfunktion  $h(k, i)$  gibt an, auf welche Adresse ein Schlüssel  $k$  abgebildet werden soll, wenn bereits  $i$  Versuche zu einer Kollision geführt haben

- Lineares Sondieren (Linear Probing):



$$h(k, i) = (h(k) + i) \bmod m$$

- Primäre Häufung** (primary clustering): tritt auf, wenn für Schlüssel  $k_1, k_2$  mit unterschiedlichen Hashwerten  $h(k_1) \neq h(k_2)$  ab einem bestimmten Punkt  $i_1$  bzw.  $i_2$  die gleiche Sondierfolge auftritt:

$$\exists i_1, i_2 \quad \forall j: \quad h(k_1, i_1 + j) = h(k_2, i_2 + j)$$

## Probleme beim quadratischen Sondieren

- Quadratisches Sondieren (Quadratic Probing):

$$h(k, i) = (h(k) + c_1 i + c_2 i^2) \bmod m \quad (c_2 \neq 0)$$

oder:  $h(k, i) = (h(k) - (-1)^i \lceil i/2 \rceil^2) \bmod m$

## Probleme beim quadratischen Sondieren

- Quadratisches Sondieren (Quadratic Probing):

$$h(k, i) = (h(k) + c_1 i + c_2 i^2) \bmod m \quad (c_2 \neq 0)$$

oder:  $h(k, i) = (h(k) - (-1)^i \lceil i/2 \rceil^2) \bmod m$

- $h(k, i)$  soll möglichst **surjektiv** auf die Adressmenge  $\{0, \dots, m-1\}$  abbilden, um freie Positionen auch **immer** zu finden. Bei

$$h(k, i) = (h(k) - (-1)^i \lceil i/2 \rceil^2) \bmod m$$

z.B. durch Wahl von  $m$  prim  $\wedge m \equiv 3 \pmod{4}$

## Probleme beim quadratischen Sondieren

- Quadratisches Sondieren (Quadratic Probing):

$$h(k, i) = (h(k) + c_1 i + c_2 i^2) \bmod m \quad (c_2 \neq 0)$$

oder:  $h(k, i) = (h(k) - (-1)^i \lceil i/2 \rceil^2) \bmod m$

- $h(k, i)$  soll möglichst **surjektiv** auf die Adressmenge  $\{0, \dots, m-1\}$  abbilden, um freie Positionen auch **immer** zu finden. Bei

$$h(k, i) = (h(k) - (-1)^i \lceil i/2 \rceil^2) \bmod m$$

z.B. durch Wahl von  $m$  prim  $\wedge m \equiv 3 \pmod{4}$

- Sekundäre Häufung** (secondary clustering):

tritt auf, wenn für Schlüssel  $k_1, k_2$  mit gleichem Hashwert

$h(k_1) = h(k_2)$  auch die nachfolgende Sondierfolge gleich ist:

$$\forall i: \quad h(k_1, i) = h(k_2, i)$$

## Double Hashing

- Auflösung der Kollisionen der Hashfunktion  $h$  durch eine zweite Hashfunktion  $h'$ :

$$h(k, i) = [h(k) + i \cdot h'(k)] \bmod m$$

wobei für alle  $k$  gelten soll, dass  $h'(k)$  teilerfremd zu  $m$  ist,

z.B.  $h'(k) = 1 + k \bmod m - 1$   
 oder  $h'(k) = 1 + k \bmod m - 2$

für Primzahl  $m$

- primäre und sekundäre Häufung werden weitgehend vermieden, aber nicht komplett ausgeschlossen

## Übersicht

- 6 Sortieren
  - Einfache Verfahren
  - MergeSort
  - Untere Schranke
  - QuickSort
  - Selektieren
  - Schnelleres Sortieren
  - Externes Sortieren

## Statisches Wörterbuch

Lösungsmöglichkeiten:

- Perfektes Hashing
  - Vorteil: Suche in konstanter Zeit
  - Nachteil: keine Ordnung auf Elementen, d.h. Bereichsanfragen (z.B. alle Namen, die mit 'A' anfangen) teuer
- Speicherung der Daten in sortiertem Feld
  - Vorteil: Bereichsanfragen möglich
  - Nachteil: Suche teurer (logarithmische Zeit)

## Sortierproblem

- gegeben: Ordnung  $\leq$  auf der Menge möglicher Schlüssel

- Eingabe: Sequenz  $s = \langle e_1, \dots, e_n \rangle$

Beispiel:



- Ausgabe: Permutation  $s' = \langle e'_1, \dots, e'_n \rangle$  von  $s$ , so dass  $\text{key}(e'_i) \leq \text{key}(e'_{i+1})$  für alle  $i \in \{1, \dots, n\}$

Beispiel:



## SelectionSort

Sortieren durch Auswählen

Wähle das kleinste Element aus der (verbleibenden) Eingabesequenz und verschiebe es an das Ende der Ausgabesequenz

### Beispiel

5	10	19	1	14	3
1	10	19	5	14	3
1	5	19	10	14	3
1	3	19	10	14	5
1	3	10	19	14	5
1	3	5	19	14	10
1	3	5	10	19	14
1	3	5	10	14	19
1	3	5	10	14	19



## SelectionSort

Sortieren durch Auswählen

```
selectionSort(Element[] a, int n) {
  for (int i = 0; i < n; i++)
    // verschiebe min{a[i], ..., a[n-1]} nach a[i]
    for (int j = i + 1; j < n; j++)
      if (a[i] > a[j])
        swap(a[i], a[j]);
}
```

Zeitaufwand:

- Minimumsuche in Feld der Größe  $i$ :  $\Theta(i)$
- Gesamtzeit:  $\sum_{i=1}^n \Theta(i) = \Theta(n^2)$
- Vorteil: einfach zu implementieren
- Nachteil: quadratische Laufzeit



## InsertionSort

Sortieren durch Einfügen

Nimm ein Element aus der Eingabesequenz und füge es an der richtigen Stelle in die Ausgabesequenz ein

### Beispiel

5	10	19	1	14	3
5	10	19	1	14	3
5	10	19	1	14	3
5	10	19	1	14	3
5	10	1	19	14	3
5	1	10	19	14	3
1	5	10	19	14	3
1	5	10	14	19	3
1	...	←	...	3	19
1	3	5	10	14	19



## InsertionSort

Sortieren durch Einfügen

```
insertionSort(Element[] a, int n) {
  for (int i = 1; i < n; i++)
    // verschiebe a_i an die richtige Stelle
    for (int j = i - 1; j >= 0; j--)
      if (a[j] > a[j + 1])
        swap(a[j], a[j + 1]);
}
```

Zeitaufwand:

- Einfügung des  $i$ -ten Elements an richtige Stelle:  $O(i)$
- Gesamtzeit:  $\sum_{i=1}^n O(i) = O(n^2)$
- Vorteil: einfach zu implementieren
- Nachteil: quadratische Laufzeit



## Einfache Verfahren

### SelectionSort

- mit besserer Minimumstrategie worst case Laufzeit  $O(n \log n)$  erreichbar  
(mehr dazu in einer späteren Vorlesung)

### InsertionSort

- mit besserer Einfügestrategie worst case Laufzeit  $O(n \log^2 n)$  erreichbar  
(→ ShellSort)

## MergeSort

### Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

### Beispiel

10	5	7	19	14	1	3
----	---	---	----	----	---	---

## MergeSort

### Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

### Beispiel

10	5	7	19	14	1	3	
10	5	7	19		14	1	3

## MergeSort

### Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

### Beispiel

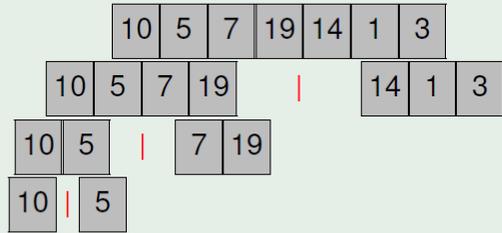
10	5	7	19	14	1	3	
10	5	7	19		14	1	3
10	5		7	19			

# MergeSort

Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

## Beispiel

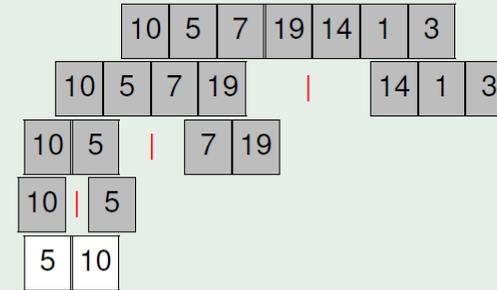


# MergeSort

Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

## Beispiel

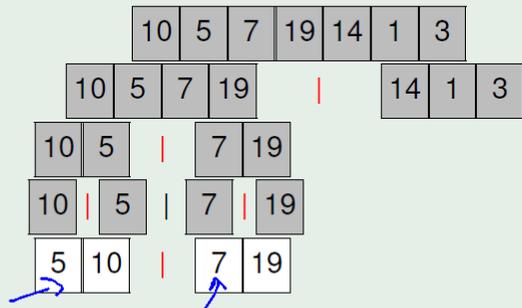


# MergeSort

Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

## Beispiel

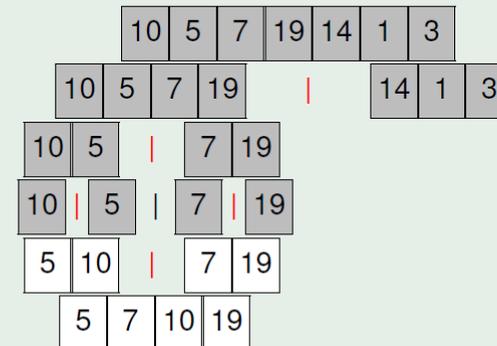


# MergeSort

Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

## Beispiel

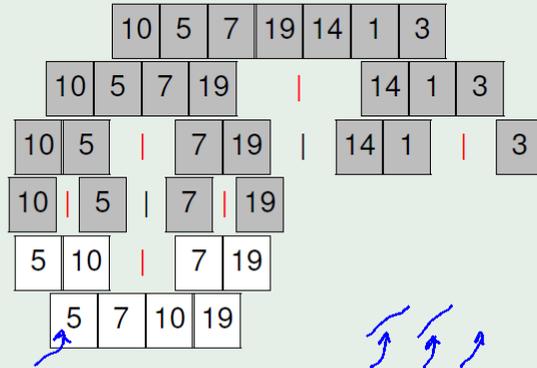


## MergeSort

Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

## Beispiel

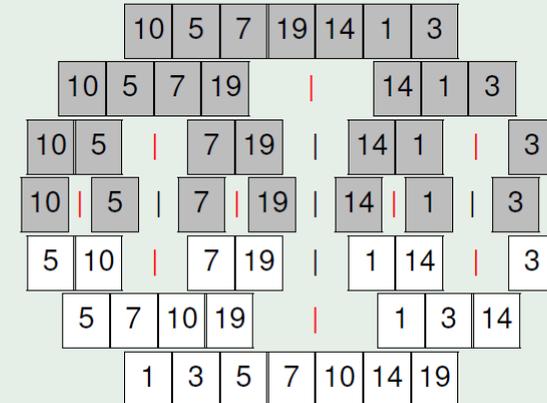


## MergeSort

Sortieren durch Verschmelzen

Zerlege die Eingabesequenz rekursiv in Teile, die separat sortiert und dann zur Ausgabesequenz verschmolzen werden

## Beispiel



## MergeSort

Sortieren durch Verschmelzen

```

mergeSort(Element[] a, int l, int r) {
    if (l == r) return;           // nur ein Element
    m = (l+r)/2;                 // Mitte
    mergeSort(a, l, m);         // linken Teil sortieren
    mergeSort(a, m+1, r);       // rechten Teil sortieren
    j = l; k = m+1;              // verschmelzen
    for i = 0 to r-l do
        if (j > m) { b[i] = a[k]; k++; } // linker Teil leer
        else
            if (k > r) { b[i] = a[j]; j++; } // rechter Teil leer
            else
                if (a[j] <= a[k]) { b[i] = a[j]; j++; }
                else { b[i] = a[k]; k++; }
    for i = 0 to r-l do a[l+i] = b[i]; // zurückkopieren
}

```

## MergeSort

Sortieren durch Verschmelzen

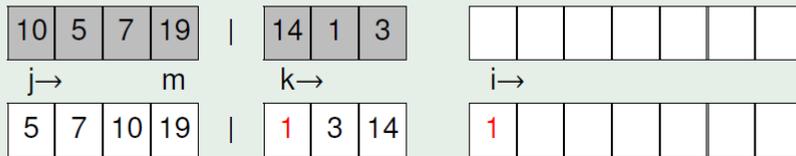
## Beispiel (Verschmelzen)



## MergeSort

Sortieren durch Verschmelzen

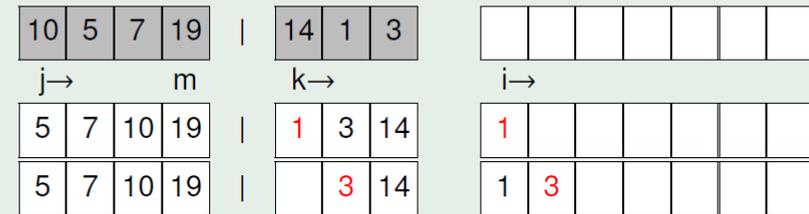
## Beispiel (Verschmelzen)



## MergeSort

Sortieren durch Verschmelzen

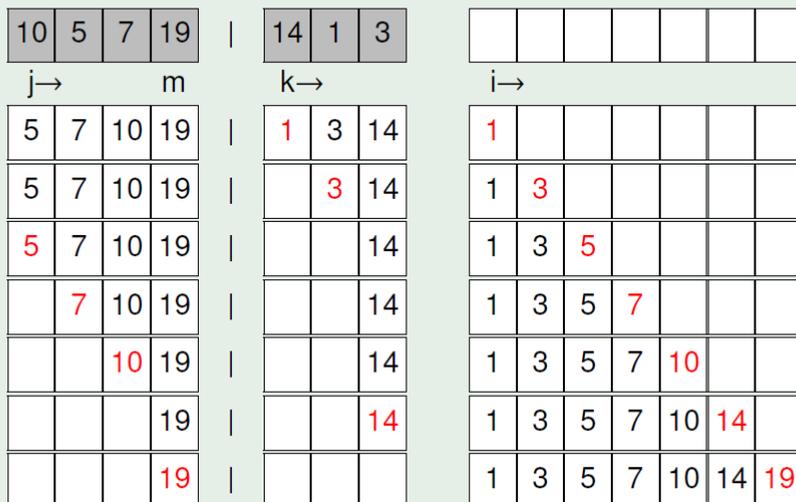
## Beispiel (Verschmelzen)



## MergeSort

Sortieren durch Verschmelzen

## Beispiel (Verschmelzen)



## MergeSort

Sortieren durch Verschmelzen

Zeitaufwand:

- $T(n)$ : Laufzeit bei Feldgröße  $n$
- $T(1) = \Theta(1)$   
 $T(n) = T(\lceil n/2 \rceil) + T(\lfloor n/2 \rfloor) + \Theta(n)$

⇒  $T(n) \in O(n \log n)$   
 (folgt aus dem sogenannten Master-Theorem)