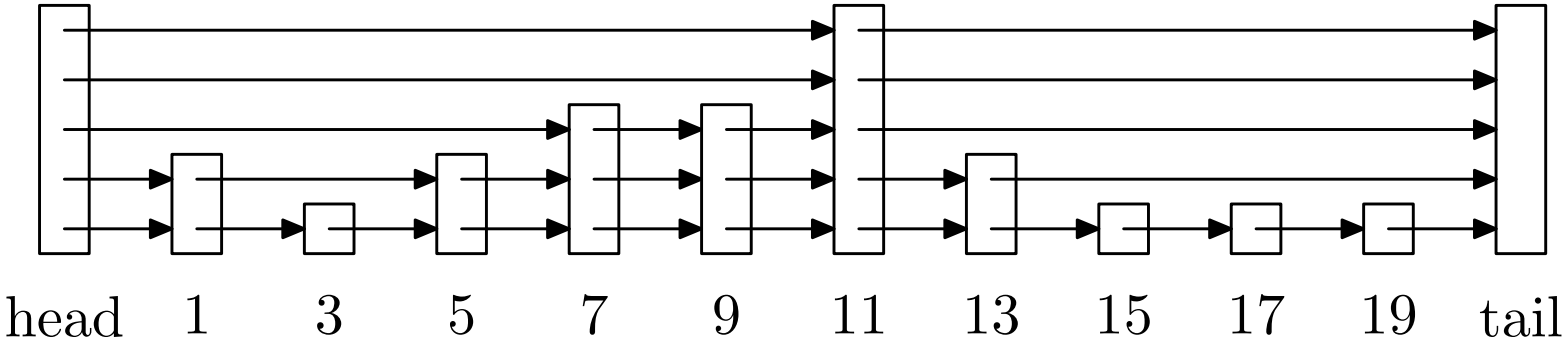
Skip-Lists – Analyse

Es sei h die Höhe, n die Anzahl der Schlüssel und p die Wahrscheinlichkeit der Skip-List.



Die erfolgreiche Suche führt entlang eines Wegs, der oben in head beginnt und schlimmstens unten im gesuchten Knoten endet.

Sehen wir uns den Weg rückwärts an!
Wieweit gehen wir im Durchschnitt, bis der Weg nach oben führt?
$\rightarrow 1/(1-p)$ viele Schritte!
$\rightarrow$ Insgesamt $h/(1-p) = O(h)$ Schritte im Erwartungswert.

Skip-Lists – Analyse

Erfolgreiche Suche: $O(h)$

Wie hoch ist eine Skip-Liste mit n Elementen?
Es sei h_i die Höhe des $i$ten Knotens.

$$\Pr[h_i \geq t] = (1-p)^{t-1}$$

$$\Pr[h_i \geq t \text{ für ein } i] \leq n(1-p)^{t-1}$$

Setze $t = -2\log_{1-p}(n) + 1 = 2\log_{1/(1-p)}(n) + 1 = O(\log n)$.

$$\rightarrow \Pr[h = O(\log n)] \geq 1 - n(1-p)^{\log_{1-p}(1/n^2)} = 1 - \frac{1}{n}$$

Also gilt $E(h) = O(\log n)(1-1/n) + O(n) \cdot 1/n = O(\log n)$.

Skip-Lists – Analyse

Theorem

Skip-Listen unterstützen die Operationen Einfügen, Löschen und Suchen in erwarteter Zeit $O(\log n)$.
Der Speicherverbrauch ist im Erwartungswert $O(n)$.

Beweis.

Löschen benötigt asymptotisch so viel Zeit wie Suchen, also $O(\log n)$.
Einfügen ebenfalls, außer die Höhe nimmt zu. Die Zeit dafür ist aber im Mittel nur $O(1)$, obwohl sie (mit kleiner Wahrscheinlichkeit) unbeschränkt groß werden kann.
Jeder Knoten benötigt im Durchschnitt $O(1)$ Platz, insgesamt ergibt das $O(n)$. □

Skip-Lists – Fragen

Wie lange benötigt eine erfolglose Suche nach einem Element, das größer ist als alle Schlüssel in der Skip-List?

Welchen Fehler darf man bei der Implementierung nicht machen, um dies zu vermeiden:

Die Liste ist fast leer, doch das Einfügen geht sehr, sehr langsam.

Mengen

Der abstrakte Datentyp Menge sollte folgende Operationen unterstützen:

- $x \in M$?
- $M \rightarrow M \cup \{x\}$
- $M \rightarrow M \setminus \{x\}$
- $M = \emptyset$?
- wähle irgendein $x \in M$

Möglicherweise auch

- $M_1 \rightarrow M_2 \cup M_3$
- $M_1 \rightarrow M_2 \cap M_3$
- $M_1 \rightarrow M_2 \setminus M_3$
- ...

Mengen können durch assoziative Arrays implementiert werden:

Java

```
public class Set<K> {
  private final Map<K,?> h;
  public Set() {h = new Hashtable<K, Integer>(); }
  public Set(Map<K,?> m) {h = m; }
  public void insert(K k) {h.insert(k, null); }
  public void delete(K k) {h.delete(k); }
  public void union(Set<K> U) {
    SimpleIterator<K> it;
    for(it = U.iterator(); it.more(); it.step())
      insert(it.key()); }
  public boolean iselement(K k) {return h.iselement(k); }
  public SimpleIterator<K> iterator() {
   return h.simpleiterator(); }
  public Array<K> array() {return h.array(); }
}
```

Bitarrays

Wenn das Universum U klein ist, können wir Mengen durch Bitarrays implementieren.

Laufzeiten:

- Suchen, Einfügen, Löschen: $O(1)$
- Vereinigung, Schnitt: $O(|U|)$
- Auswahl: $O(|U|)$ (oder $O(1)$ mit Zusatzzeigern)

Insertion Sort

Wir sortieren ein unsortiertes Array, indem wir wiederholt Elemente in ein bereits sortiertes Teilarray einfügen.

Eingabe:

12	73	36	71	79	67	3	17	32	31	14	14	33	4	74	23

Ausgabe:

3	4	12	14	14	17	23	31	32	33	36	67	71	73	74	79

Insertion Sort

12	73	36	71	79	67	3	17	32	31	14	14	33	4	74	23

Algorithmus

procedure *insertionsort*(n) :
  for $i = 2, ..., n$ **do**
    $j := i$;
    while $j \geq 2$ **and** $a[j-1] < a[j]$ **do**
      vertausche $a[j-1]$ *und* $a[j]$;
      $j := j - 1$
    od
  od

Inversionen

Die Laufzeit von Insertion Sort ist $O(n^2)$.

Definition

Sei $\pi \in S_n$ eine Permutation. Die Menge der Inversionen von π ist

$$I(\pi) = \{ (i,j) \in \{1,\ldots,n\}^2 \mid i < j \text{ und } \pi(i) > \pi(j) \}.$$

Beim Sortieren durch Einfügen werden Schlüssel nur durch Vertauschungen benachbarter Elemente bewegt.
So eine Vertauschung verringert die Anzahl der Inversionen höchstens um eins.
Wenn die Eingabe genau falschherum sortiert ist, hat Insertion Sort die Laufzeit $\Theta(n^2)$.
Was ist die durchschnittliche Laufzeit?

Inversionen

Theorem

Eine zufällig gewählte Permutation $\pi \in S_n$ hat im Erwartungswert $n(n-1)/4$ Inversionen.

Beweis.

Es gibt $n(n-1)/2$ viele Paare (i,j) mit $1 \leq i < j \leq n$.
Wegen

$$\Pr[\pi(i) > \pi(j)] = \frac{1}{2} \text{ falls } i \neq j$$

gilt

$$E(|I(\pi)|) = \sum_{i<j} \Pr[\pi(i) > \pi(j)] = \frac{n(n-1)}{4}.$$

□

Inversionen

Theorem

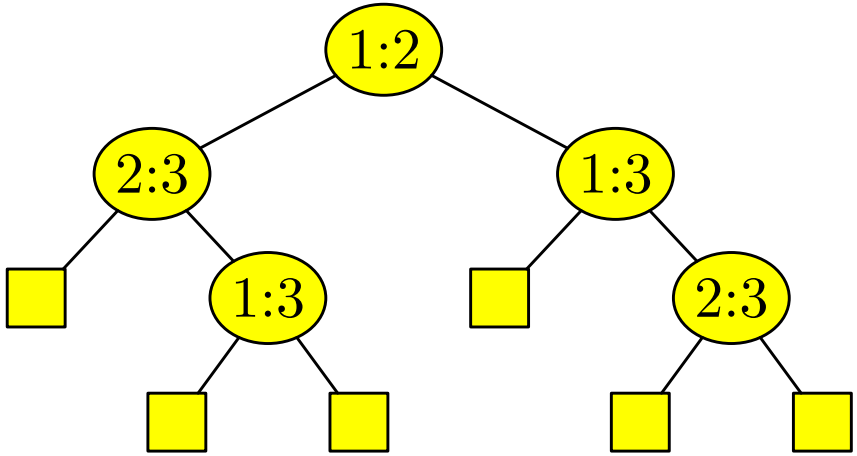
Jedes Sortierverfahren, das Schlüssel nur durch Vertauschungen benachbarter Elemente bewegt, benötigt im Durchschnitt $\Omega(n^2)$ Zeit.

Folgerung:

Wenn wir schneller sein wollen, müssen Schlüssel über weite Strecken bewegt werden.

Vergleichsbäume

Insertion Sort mit $n = 3$:



Jeder vergleichsbasierte Sortieralgorithmus hat einen Vergleichsbaum.

Die Wurzel ist der erste Vergleich.

Links folgen die Vergleiche beim Ergebnis kleiner.
Rechts folgen die Vergleiche beim Ergebnis größer.

Vergleichsbäume

Lemma

Der Vergleichsbaum eines vergleichsbasierten Sortieralgorithmus hat mindestens $n!$ Blätter.

Beweis.

Wenn zwei Permutationen zum gleichen Blatt führen, wird eine von ihnen falsch sortiert.

Es gibt aber $n!$ viele Permutationen. □

Theorem

Jeder vergleichsbasierte Algorithmus benötigt für das Sortieren einer zufällig permutierten Eingabe im Erwartungswert mindestens

$$\log(n!) = n \log n - n \log e - \frac{1}{2} \log n + O(1)$$

viele Vergleiche.

Beweis.

Sei T ein entsprechender Vergleichsbaum. Die mittlere Pfadlänge zu einem Blatt ist am kleinsten, wenn der Baum balanziert ist. In diesem Fall ist die Höhe $\log(n!)$. Stirling-Formel:

$$n! = \frac{1}{\sqrt{2\pi n}} \left(\frac{n}{e}\right)^n (1 + O(n^{-1})).$$

□

Mergesort

Wir sortieren ein unsortiertes Array durch einen Divide-and-Conquer–Algorithmus:

Eingabe:

12	73	36	71	79	67	3	17	32	31	14	14	33	4	74	23

Ausgabe:

3	4	12	14	14	17	23	31	32	33	36	67	71	73	74	79

Frage: Wie zerteilen wir die Eingabe in zwei unabhängige Teilprobleme?

Mergesort

Eine einfache Strategie: Teilen in der Mitte.

12	73	36	71	79	67	3	17

32	31	14	14	33	4	74	23

1. Teile das Array in der Mitte.
2. Sortiere beide Hälften.
3. Mische beide in eine sortierte Folge.

3	12	17	36	67	71	73	79

4	14	14	23	31	32	33	74

3	4	12	14	14	17	23	31	32	33	36	67	71	73	74	79

Mergesort

Mischen ist der schwierige Teil.

3	12	17	36	67	71	73	79

4	14	14	23	31	32	33	74

3	4	12	14	14	17	23	31	32	33	36	67	71	73	74	79

Algorithmus, der $a[l], \ldots, a[m-1]$ mit $a[m], \ldots, a[r]$ mischt:

Algorithmus

$i := l; j := m; k := l;$
while $k \leq r$ **do**
  if $a[i] \leq a[j]$ **and** $i\langle m$ **or** $j\rangle r$
  then $b[k] := a[i]; k := k+1; i := i+1$
  else $b[k] := a[j]; k := k+1; j := j+1$ **fi**
od;
for $i = l, \ldots, r$ **do** $a[k] := b[k]$ **od**

Mergesort

Algorithmus

procedure $mergesort(l, r)$:
  if $l \geq r$ **then return fi**;
  $m := \lceil (r + l)/2 \rceil$;
  $mergesort(l, m - 1)$;
  $mergesort(m, r)$;
  $i := l; j := m; k := l$;
  while $k \leq r$ **do**
    if $a[i] \leq a[j]$ **and** $i \langle m$ **or** $j \rangle r$
    then $b[k] := a[i]; k := k + 1; i := i + 1$
    else $b[k] := a[j]; k := k + 1; j := j + 1$ **fi**
  od;
  for $i = l, \ldots, r$ **do** $a[k] := b[k]$ **od**

Analyse von Mergesort

Das Mischen dauert $\Theta(n)$.

Sei $T(n)$ die Laufzeit. Wir erhalten die Gleichung

$$T(n) = \Theta(n) + T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil).$$

Falls n eine Zweierpotenz ist, gilt

$$T(n) \leq cn + 2T(n/2).$$

Wiederholtes Einsetzen liefert

$$T(n) \leq c\left(n + 2\frac{n}{2} + 4\frac{n}{4} + \cdots\right) = O(n \log n).$$