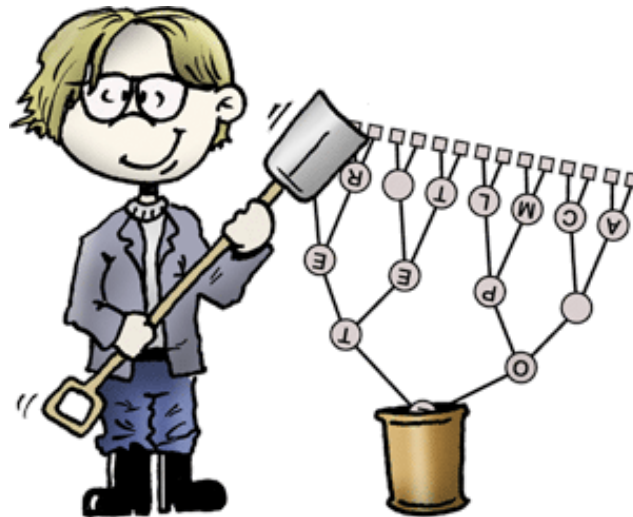


Algorithmen und Datenstrukturen



Vorlesung im Rahmen des dualen Studiums
„Angewandte Mathematik und Informatik“
(FH Aachen) / MATSE-Ausbildung

Prof. Dr. Hans Joachim Pflug

(Ende März – Anfang April)

Grundbegriffe; auf Feldern basierende Datenstrukturen

Termin	Inhalt
1	Grundbegriffe, sequentielle und binäre Suche
2	ADTs, dynamische Felder, zirkuläre dynamische Felder
3	Mengen, assoziative Felder, Hashtabellen

... Java-Klausur

(Anfang April bis Ende April)

Verkettete Listen und Bäume

Termin	Inhalt
4	Verkettete Listen
5	Binärbäume, binäre Suchbäume
6	Balancierte Bäume, AVL-Bäume
7	B-Bäume
8	Varianten von B-Bäumen, Baumdurchlauf (1)
9	Heaps, Heap-Sort

(Ende April bis Mitte Mai)

Bäume und Graphen

Termin	Inhalt
10	Graphen, Breitensuche
11	Tiefensuche, Baumdurchlauf
12	Travelling Salesman
13	Backtracking 1
14	Backtracking 2
15	Dijkstra-Algorithmus

(Mitte Mai bis Anfang Juni)

Textsuche

Termin	Inhalt
16	Floyd-Warshall-Algorithmus
17	Textsuche
18	Reguläre Ausdrücke, PCRE
19	Simulation nach Kleene

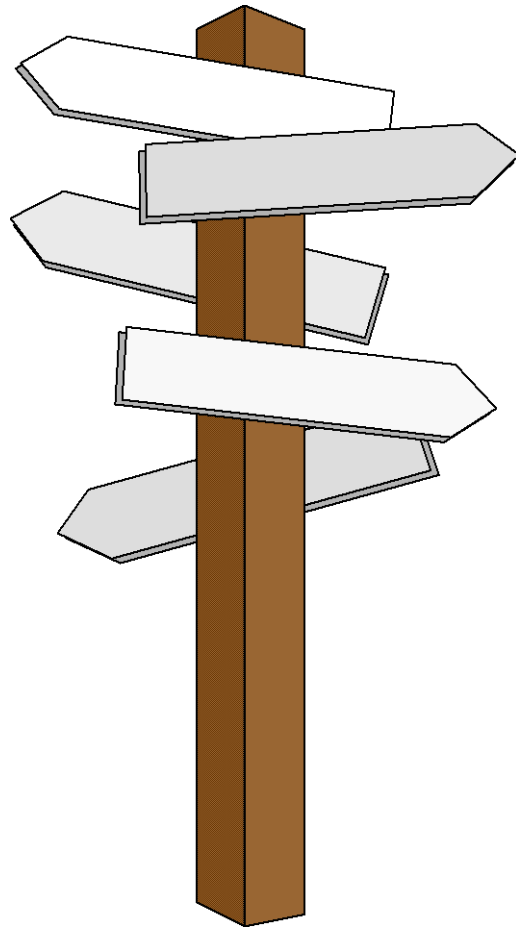
(Anfang Juni bis Ende Juni)

Sortieralgorithmen

Termin	Inhalt
20	Einfache Sortierverfahren
21	Quick-Sort
22	Merge-Sort und Radix-Sort
23	Klausurvorbereitung
24	Klausurvorbereitung

- **Th. C. Cormen et al. :**
„Algorithmen – Eine Einführung“
1339 Seiten
Oldenbourg, 4. Auflage 2013
- **A. V. Aho, J. D. Ullman:**
„Foundations of Computer Science“
786 Seiten
Computer Science Press, 1995
<http://infolab.stanford.edu/~ullman/focs.html>
- **Owsnicki-Klewe, B.:**
„Algorithmen und Datenstrukturen“
224 Seiten
Wißner, 5. Auflage 2009





Vorlesung 1

Grundbegriffe, sequentielle und binäre Suche

Abu Ja'far Mohammed ibn Musa *al-Khowarizmi*:



- Persischer Mathematiker, lebte um 820 in **Bagdad**
- Sein Name wurde in den lateinischen Übersetzungen des Mittelalters zu *Algorismi*.
- Der Name stammt eigentlich von seiner Geburtsstadt *Khowarizm*, dem heutigen *Chiwa* in Usbekistan. Die Stadt enthält zahlreiche Kulturdenkmäler und wurde 1990 komplett zum Weltkulturerbe erklärt.

- Bedienungsanleitung für Münztelefone
- Kochrezept
- Lösungsanleitung für den Zauberwürfel
- ...

Der Algorithmusbegriff

- Ein **Algorithmus** (in der EDV) ist
 - eine **Lösungsvorschrift** für eine Problemklasse
 - eine Problemklasse ist z.B.: „löse quadratische Gleichung“.
 - das konkrete Problem wird durch Eingabeparameter spezifiziert.
 - geeignet für Implementierung als Computerprogramm
 - **endliche** Folge von **elementaren, ausführbaren Instruktionen** (Verarbeitungsschritten)

- **Beispiele für Instruktionen:**

```
x = x + z;
```

```
if (a > 0) b = 1;
```

Terminierung:

Ein Algorithmus heißt *terminierend*, wenn er (bei jeder erlaubten Eingabe) nach endlichen vielen Schritten abbricht.

Determiniertheit:

Festgelegtes **Ergebnis**: Bei vorgegebener Eingabe wird ein eindeutiges Ergebnis geliefert.

Determinismus:

Festgelegter **Ablauf**: Eindeutige Vorgabe der Abfolge der auszuführenden Schritte. (\Rightarrow Determiniertheit)

Die meisten hier betrachteten Algorithmen sind deterministisch und terminierend. Sie definieren eine *Ein-/Ausgabefunktion*:

$$f : \text{Eingabewerte} \rightarrow \text{Ausgabewerte}$$

- Verbale Umschreibung (Handlungsanweisung)
 - graphisch: Struktogramme, Flussdiagramme, ...
 - Pseudo-Code
 - Höhere Programmiersprache
 - ...
-
- Churchsche These \Rightarrow Alle Darstellungsformen von Algorithmen sind äquivalent.

- Unterschiedlichste Darstellungen für denselben Algorithmus; auch Programme sind Darstellungen ...
 - Außerdem: Unterschiedliche Algorithmen für das gleiche Problem
- ⇒ Ziel: Algorithmen miteinander **vergleichen (Komplexitätsanalyse)**
- Kriterien: **Laufzeit** und **Speicherplatzbedarf**
 - Komplexität wird ausgedrückt in Abhängigkeit von Menge und Größe der bearbeiteten Daten
- Komplexität von Algorithmus (bedingt) auf konkrete Implementierung übertragbar

- **Komplexität:**
Zur Berechnung erforderliche Aufwand an Betriebsmitteln (Speicherplatz, Rechenzeit, benötigte Geräte, usw.)
- **Komplexität eines Algorithmus:** erforderlicher Aufwand bei Realisierung des Algorithmus (innerhalb des Berechnungsmodells)
- **Komplexität einer Funktion:** Komplexität des bestmöglichen Algorithmus von allen Algorithmen, die die Funktion berechnen
- Sei A ein Algorithmus, der die Funktion f berechnet
 - Komplexität von $A \geq$ Komplexität von f
 - Komplexität von $f \leq$ Komplexität von A
- Fallen untere und obere Schranke zusammen, so hat man einen **optimalen Algorithmus** für das gestellte Problem.

Möglichkeiten für das „Messen“ der Effizienz eines Algorithmus:

1. Algorithmus auf **realem Rechner** implementieren;
Zeitverbrauch messen
Nachteil: Zu viele Einflussgrößen abgesehen von Algorithmus selbst
2. Zählen der **Rechenschritte**, die bei Durchführung für eine Eingabe gemacht werden \Rightarrow Frage: was ist ein Rechenschritt?
 \Rightarrow formales Berechnungsmodell, z.B. Turingmaschine oder RAM
 \Rightarrow Algorithmus in **künstlicher Programmiersprache** „programmieren“;
Operationen zählen und gewichten
Nachteil: Aufwand; Frage der Übertragbarkeit
3. **Operationen auf sehr hohem Niveau zählen** (z.B. Anzahl der Vergleiche beim Suchen in Liste oder Anzahl der zu vertauschenden Elementpaare beim Sortieren)
Nachteil: Grobe Abschätzung

Zeitkomplexität: Einführung

- Wir benutzen als einfach zu handhabende Messgröße den *Zeitverbrauch*.
- Gegeben ist ein Problem der Problemgröße n
Beispiel: Sortieren von n Werten.
- Es gibt mehrere mögliche Algorithmen für das Problem.
- Welcher Algorithmus der schnellste ist, hängt von der Problemgröße n ab.
- Für sehr große n wird aber immer derjenige Algorithmus der schnellste sein, dessen Laufzeit am wenigsten mit n ansteigt.

Zusammenfassung einiger mathematischer Zusammenhänge, die in der Vorlesung gebraucht werden

- Gaußsche Summenformel

$$\sum_{k=1}^n k = 1 + 2 + 3 + 4 + \dots + n = \frac{n(n+1)}{2}$$

- Geometrische Reihe

$$\sum_{k=0}^{\infty} q^k = 1 + q^2 + q^3 + \dots = \frac{1}{1-q}; \quad |q| < 1$$

Insbesondere:

$$q = \frac{1}{2} \Rightarrow \sum_{k=0}^{\infty} q^k = 2 \quad \text{und} \quad q = \frac{2}{3} \Rightarrow \sum_{k=0}^{\infty} q^k = 3$$

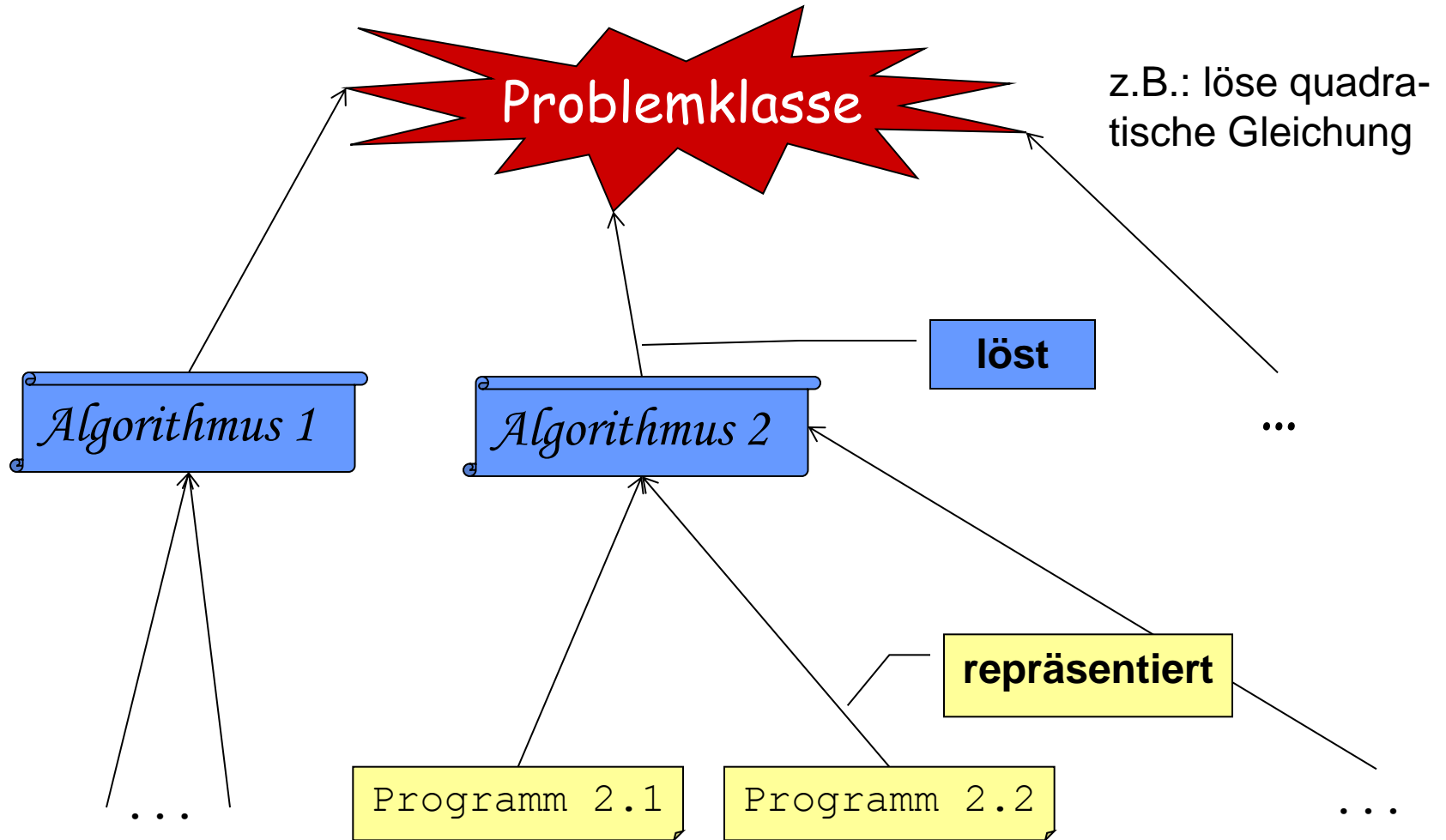
Zusammenfassung einiger mathematischer Zusammenhänge, die in der Vorlesung gebraucht werden

Logarithmus

$$a = b^x \Leftrightarrow x = \log_b a = \frac{\ln a}{\ln b}$$

Speziell:

$$a = 2^x \Leftrightarrow x = \log_2 a = \frac{\ln a}{\ln 2}$$



Problemklasse: Suchen im sortierten Feld

Algorithmus 1: Sequentielle Suche

Algorithmus 2: Binäre Suche

Beide Algorithmen können unterschiedlich (gut oder weniger gut) implementiert werden.

Problemklasse: Test einer natürlichen Zahl, ob sie eine Primzahl ist

Algorithmus: Versuche alle Teiler zwischen 2 und \sqrt{n} .

- Man unterscheidet den Zeitbedarf im
 - im schlechtesten Fall (**worst case**),
 - in einem „durchschnittlichen“ Fall (**average case**) und
 - im besten Fall (**best case**).

- Der durchschnittliche Fall ist dann wichtig, wenn der beste oder der schlechteste Fall selten auftretende Sonderfälle sind.

- Man sucht das Feld von vorne bis hinten durch.
- Im besten Fall (**best case**) steht das gesuchte Element vorne.
- Im schlechtesten Fall (**worst case**) steht das gesuchte Element hinten oder ist gar nicht vorhanden.
- Den „durchschnittlichen“ Fall (**average case**) definiert man hier als die durchschnittliche Suchdauer für ein vorhandenes Element.
- Der durchschnittliche Fall ist dann wichtig, wenn der beste oder der schlechteste Fall selten auftretende Sonderfälle sind.

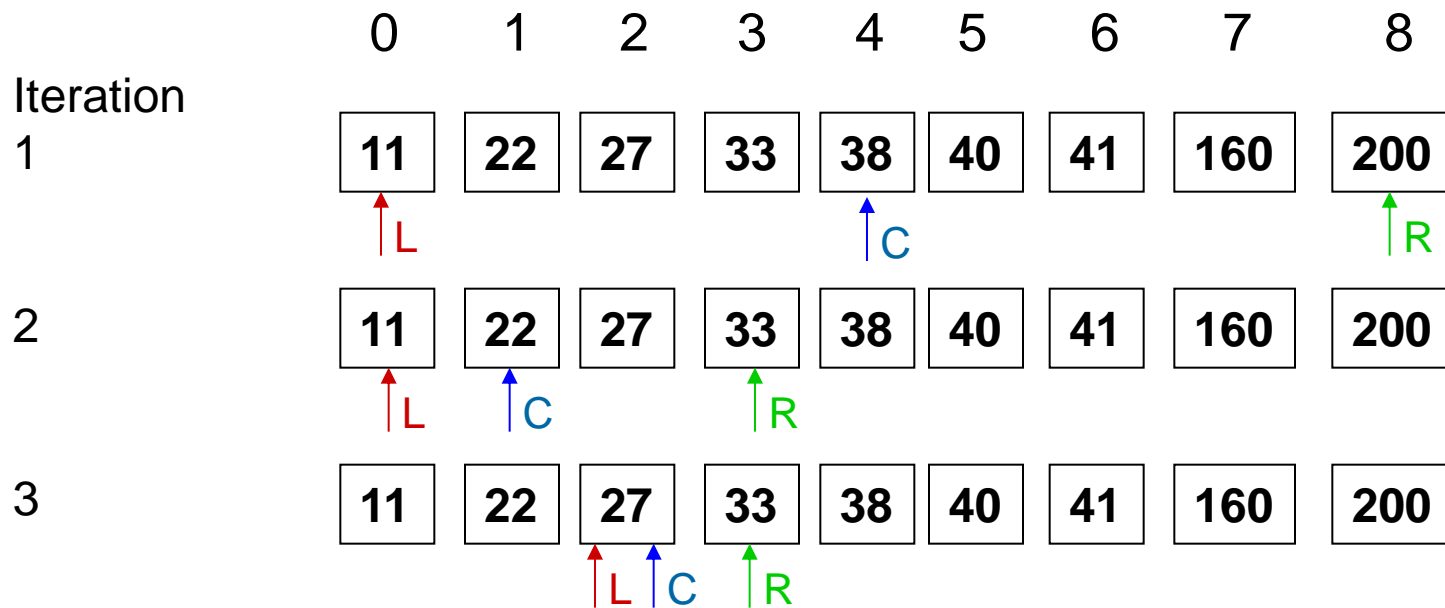
Voraussetzung:

- Liste ist sortiert

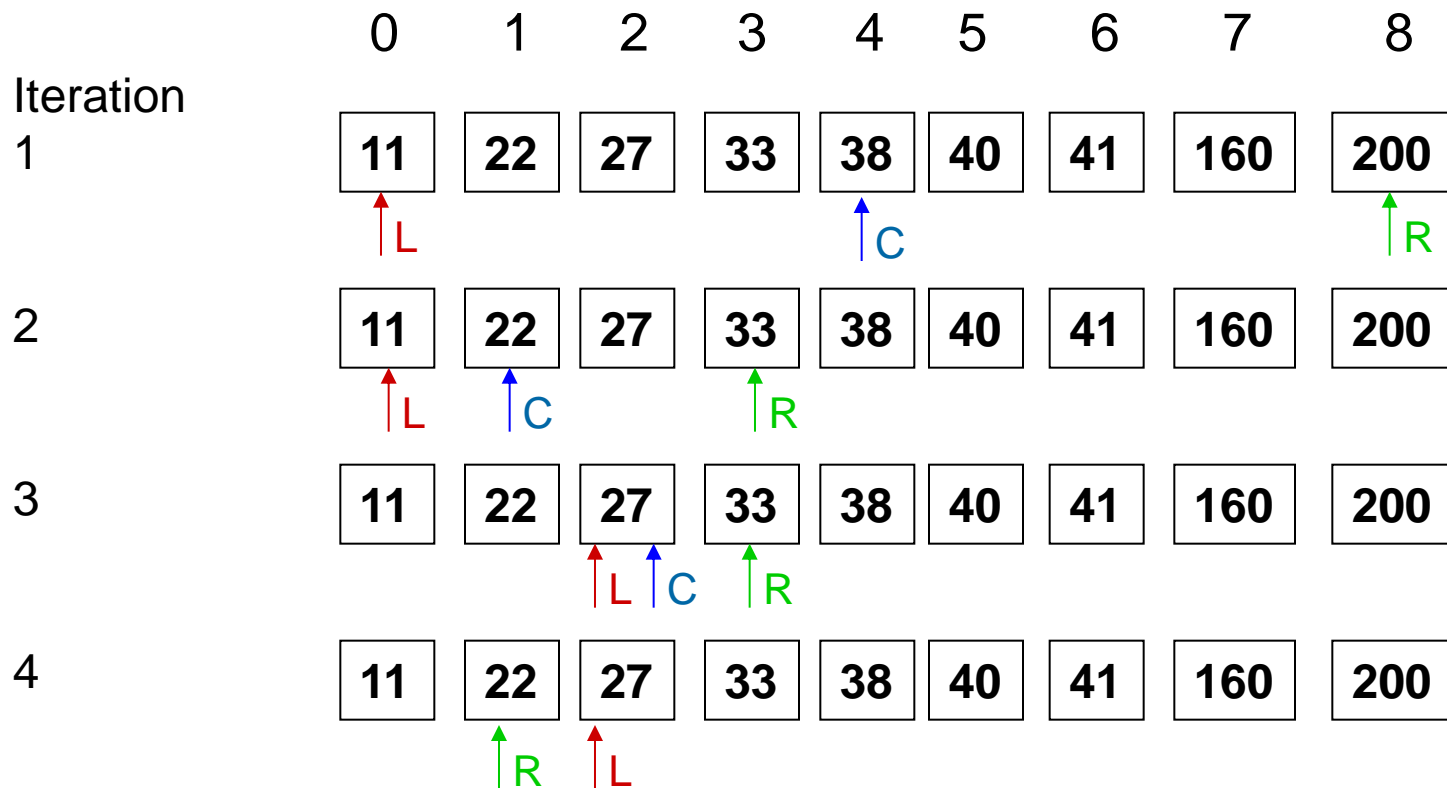
Prinzip:

- Greife zuerst auf das mittlere Element zu
- Prüfe, ob der Schlüssel des gesuchten Elements
 - kleiner als der Schlüssel des betrachteten Elements ist
⇒ suche genauso in linker Teilliste weiter
 - gleich dem Schlüssel des betrachteten Elements ist
⇒ fertig
 - größer als der Schlüssel des betrachteten Elements ist
⇒ suche genauso in rechter Teilliste weiter

- Suchen der 27



- Suchen der 26



```
java.util.Arrays.binarySearch(type[] a, type key);
```

type: (Es gibt Varianten für int, long, float, double, ..., Object)

Searches the specified array for the specified value using the binary search algorithm. The array **must** be sorted (as by the sort method, above) prior to making this call. If it is not sorted, the results are undefined. If the array contains multiple elements with the specified value, there is no guarantee which one will be found.

Parameters:

a - the array to be searched.

key - the value to be searched for.

Gibt es auch für Collections:

bis Java 1.4:

```
java.util.Collections.binarySearch(List list, Object key);
```

ab Java 1.5 (Verwendung von Generic Types)

```
java.util.Collections.binarySearch(List <? extends  
Comparable<? super T>> list, T key);
```

In Java 1.5 stellt die Methode von vornherein sicher, dass *key* mit den Elementen aus *list* vergleichbar ist.

Fortgeschrittene Verwendung von Generic Types, an dieser Stelle keine Erläuterung.

Verwendung wie bei Java 1.4 möglich.

Landau-Symbole

- Uns interessiert, wie ein Algorithmus für große n von n abhängt. Um das mathematisch genau fassen zu können, benutzt man die Landau-Symbole.



Edmund Georg Hermann Landau:

* 14. Februar 1877, † 19. Februar 1938

Deutscher Mathematiker, der sich um die analytische Zahlentheorie verdient gemacht hat

Ausgangspunkt: Man möchte wissen, wie sich die Laufzeitfunktion $T(n)$ für große n verhält.

- Daraus kann man Rückschlüsse ziehen, wie groß n bei akzeptabler Laufzeit sein darf.

In den meisten Fällen entspricht das dem asymptotischen Verhalten

$$\lim_{n \rightarrow \infty} t(n)$$

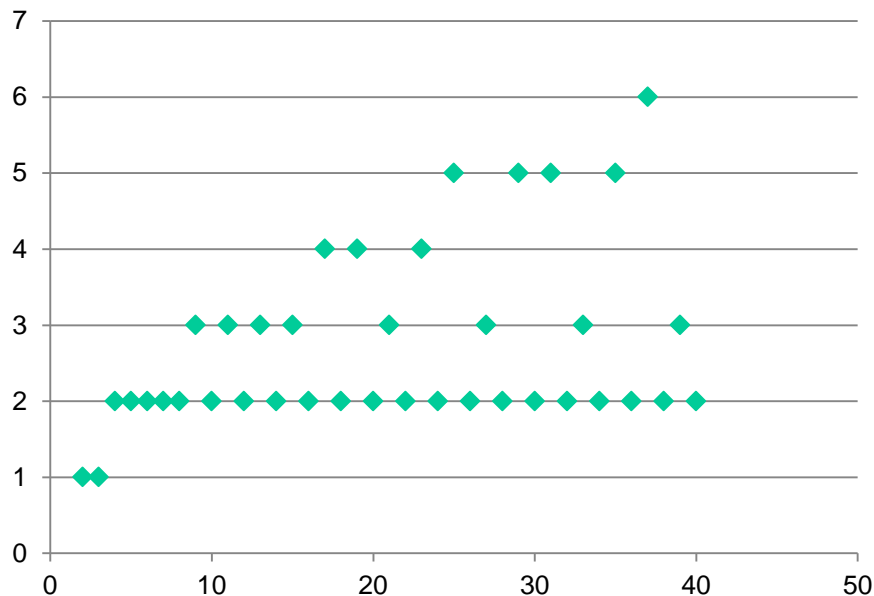
Für die sequentielle Suche gilt: Best Case: $O(1)$, Average/Worst Case: $O(n)$

Das folgende Beispiel zeigt aber, dass diese Definition nicht allgemein genug ist.

- Ein Programm soll herausfinden, ob eine Zahl n eine Primzahl ist.
- Es wird kontrolliert, ob n Teiler zwischen 2 und \sqrt{n} besitzt.
- Warum reicht es, die Zahlen bis \sqrt{n} zu kontrollieren?
 - Falls n keine Primzahl ist, besitzt n ein Teilerpaar $n=a \cdot b$.
 - Wenn $a \geq \sqrt{n}$ ist, muss $b \leq \sqrt{n}$ sein.

```
public static boolean istPrim(int z) {  
    for (int i=2; i<=Math.sqrt(z); i++) {  
        if (z%i==0) {  
            return false;  
        }  
    }  
    return true;  
}
```

istPrim



◆ istPrim

- Die Laufzeit des Algorithmus schwankt stark:
- Für Primzahlen ist sie hoch.
- Für gerade Zahlen ist sie niedrig.
- Betrachtet man nur die Primzahlen, ist die Laufzeit $T(n)$ eine Stufenfunktion.
- Man kann kein asymptotisches Verhalten angeben.
- Geeignete Notation nötig.

Schleifenobergrenze in Abhängigkeit von der zu prüfenden Zahl.

In der Regel ist man nicht an exakter Anzahl v. Operationen interessiert, sondern an **Komplexitätsklassen**:

„Wie verändert sich der Rechenaufwand, wenn man die Eingabedaten um einen bestimmten Faktor vergrößert? Wie ist der qualitative Verlauf der Laufzeitfunktion?“

Für eine Funktion $g: \mathbb{N} \rightarrow \mathbb{R}^{>0}$ sind folgende Funktionenmengen definiert:

$$O(\mathbf{g}) = \{ \mathbf{f}: \mathbb{N} \rightarrow \mathbb{R}^{>0} \mid \exists n_0 \in \mathbb{N}^{>0}, \exists c \in \mathbb{R}^{>0} \text{ mit } \mathbf{f}(n) \leq c \cdot \mathbf{g}(n) \quad \forall n \geq n_0 \}$$

$$\Omega(\mathbf{g}) = \{ \mathbf{f}: \mathbb{N} \rightarrow \mathbb{R}^{>0} \mid \exists n_0 \in \mathbb{N}^{>0}, \exists c \in \mathbb{R}^{>0} \text{ mit } \mathbf{f}(n) \geq c \cdot \mathbf{g}(n) \quad \forall n \geq n_0 \}$$

Bezeichnungen für O (sprich: „groß O “; eigentlich „groß Omikron“):

- „**Asymptotische obere Schranke**“;
- „**Ordnung** der Funktion“;
- „**Komplexitätsklasse**“

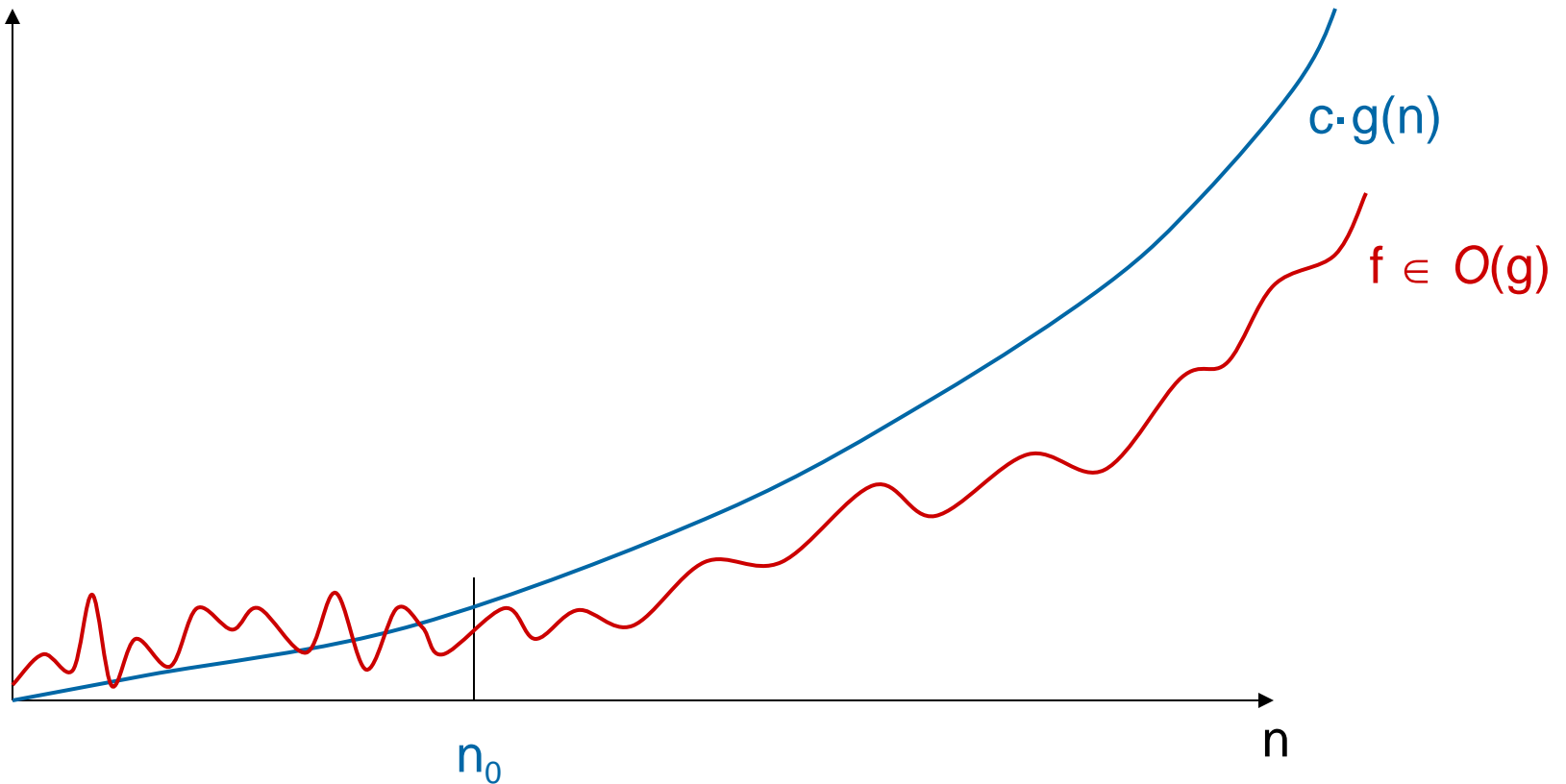
- $O(g)$ ist die Menge aller Funktionen, die asymptotisch höchstens so schnell wachsen wie $c \cdot g$.
- z.B. ist

$$2n^2 + 5n + 13 \in O(n^2)$$

$$2n^2 + 5n + 13 \notin O(n)$$

$$5n + 13 \in O(n^2)$$

$$5n + 13 \in O(n)$$



- Bei Summen setzt sich der am schnellsten ansteigende Term durch.

Beispiel:

$$f(n) = 2n^2 + 7n + 10$$

steigt am schnellsten an.

$$\Rightarrow f(n) \in O(n^2)$$

Beweis: $2n^2 + 7n + 10 \leq 3n^2$

für $n_0 = 9$

Man könnte auch schreiben

$$f(n) = 2n^2 + 7n + 10 \in O(2n^2 + 7n + 10)$$

Es interessiert aber nur der Vergleich mit einfachen Funktionen, wie $O(n)$, $O(n^2)$,...

- Faustregel: Kleinstmögliche Komplexitätsklasse (in O-Notation) ergibt sich aus $T_A(n)$ durch:
 - Extraktion des „dominanten“ (größten) Terms und
 - Weglassen des Koeffizientenz.B. $T(n)=60n^2 + 4n \Rightarrow T \in O(n^2)$;
 $T(n)=\text{ld}(n) + 1 \Rightarrow T \in O(\text{ld}(n))=O(\log(n))$, wegen

$$\log_a(n) = \frac{\log_b(n)}{\log_b(a)}$$

- Ein Algorithmus habe den maximalen Zeitbedarf $T_A(n) = 3n^2$. Also ist
 - $T_A(n) \in O(n^2)$
 - $T_A(n) \in O(n^3)$
 - usw.
- Interessant ist natürlich nur das minimale $O(g)$, hier also $O(n^2)$.
- Das Wörtchen *minimal* wird aber oft weggelassen.

- $T_A(n) = n(n-1)/2$
 $T_A \in O(n^2)$, denn:

- $T_A(n) = n^2 + 2n$
 $T_A \in O(n^2)$, denn:

Also:

- Algorithmus mit $T_A(n) = n(n-1)/2$ ist zwar besser als einer mit $n^2 + 2n(!)$;
- aber sie sind beide in der gleichen Komplexitätsklasse (etwa gleich gut).

- $T_A(n) = n(n-1)/2$

$T_A \in O(n^2)$, denn:

$$n(n-1)/2 = 1/2 (n^2 - n) \leq n^2 - n \leq n^2 \quad (n > 0)$$

\Rightarrow mit $n_0=1$, $c=1$ erfüllt $T_A(n)$ die Definition der O -Notation

- $T_A(n) = n^2 + 2n$

$T_A \in O(n^2)$, denn:

$$n^2 + 2n \leq n^2 + 2n^2 = 3n^2$$

\Rightarrow mit $n_0=1$, $c=3$ ist die Definition erfüllt

Also:

- Algorithmus mit $T_A(n) = n(n-1)/2$ ist zwar besser als einer mit $n^2 + 2n(!)$;
- aber sie sind beide in der gleichen Komplexitätsklasse (etwa gleich gut).

- **Schleife**

```
for (int i=1; i<=n; i++) {  
    a[i] = 0;  
}
```

$O()$

- **Nacheinanderausführen**

```
for (int i=1; i<=n; i++)  
    a[i] = 0;  
for (int i=1; i<=n; i++)  
    for (int j=1; j<=n; j++)  
        a[i] = a[j] + i + j;
```

$O()$

- **Geschachtelte Schleife**

```
for (int i=1; i<=n; i++)  
    for (int j=1; j<=n; j++)  
        k++;
```

$O()$

- **Bedingte Anweisung**

```
if (x > 100) {  
    y = x;  
} else {  
    for (int i=1; i<=n; i++)  
        if (a[i] > y)  
            y = a[i];  
}
```

$O()$

- **Innere Schleifenbedingung hängt von Lauf ab**

```
for (int i=1; i<=n; i++)  
    for (int j=1; j<= i ; j++)  
        k++;
```

T(n)=

⇒ O()

- **Aussprung**

```
for (int i=1; i<=n; i++)  
    if (i > 2)  
        return;
```

T(n)=

⇒ O()

- **Schleifenvariable in jedem Lauf halbiert**

```
int i=n;  
while (i>=1) {  
    x ++;  
    i /= 2;    // i wird immer halbiert  
}
```

T(n)=

⇒ O()

O(n) Funktionsaufrufe

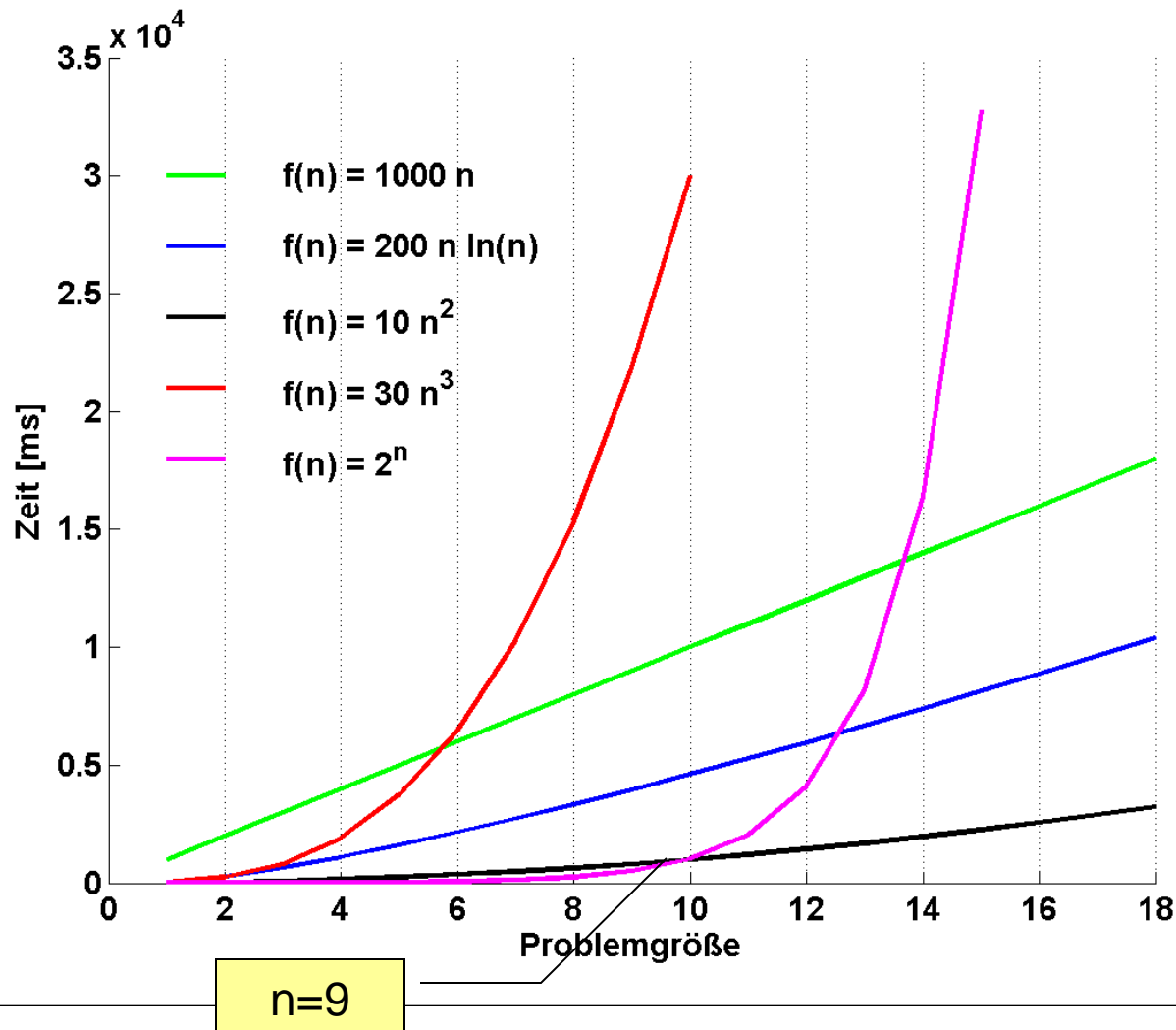
```
static long fakultaet(int n) {  
    if (n>0) {  
        return (n * fakultaet(n-1));  
    } else {  
        return 1;  
    }  
}
```

- Bei jedem Suchschritt halbiert sich die Größe des noch zu durchsuchenden Restfelds
⇒ im Worst Case $\lfloor \lg(n) \rfloor + 1$ Suchschritte
- Mit k Suchschritten sind $2^k - 1$ Elemente erreichbar.
- Nach $\lfloor \lg(n) \rfloor$ Schritten ist etwa die Hälfte aller Elemente erreichbar
⇒ Worst Case nicht wesentlich aufwändiger als Average Case

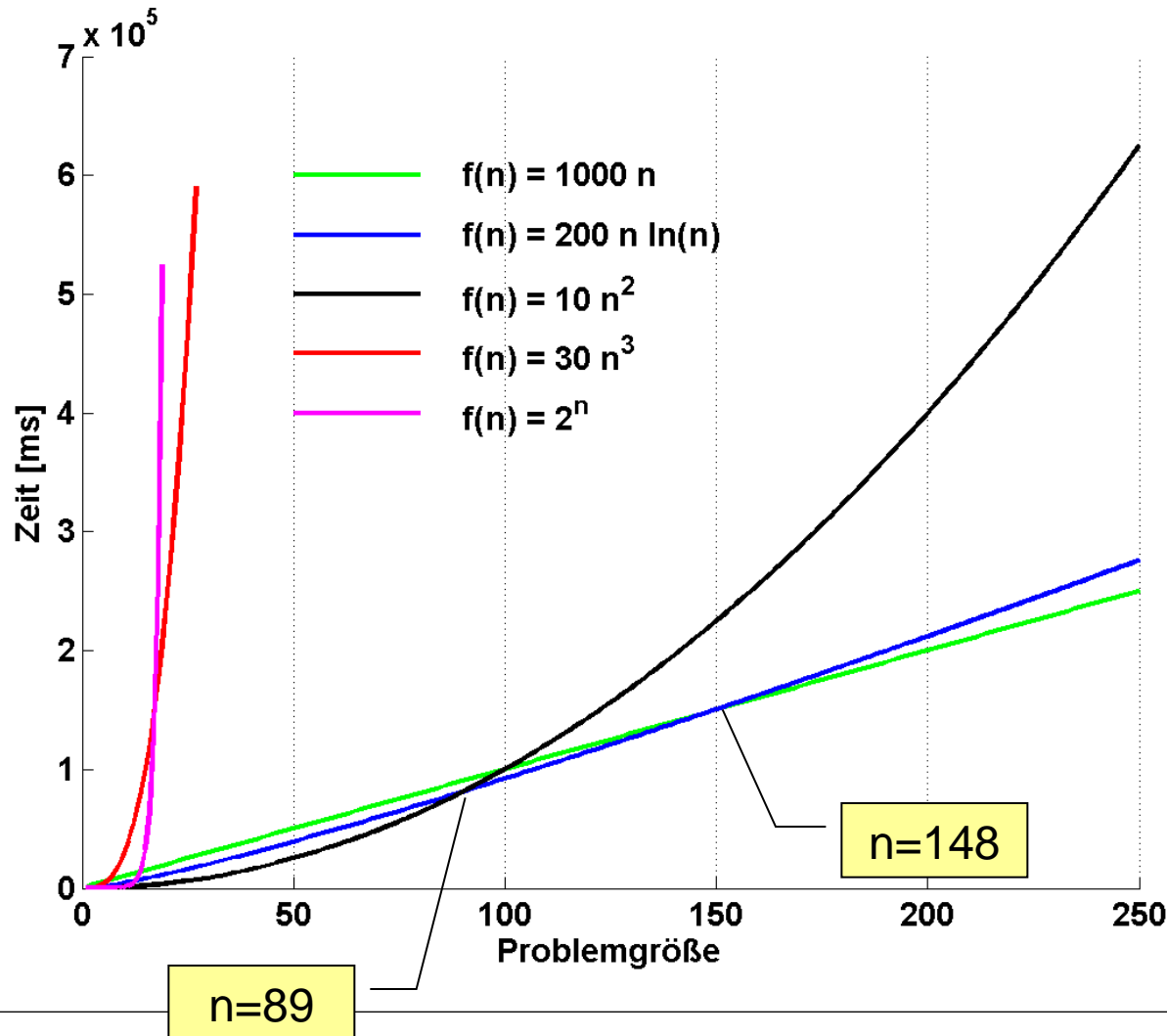
- Prüfen:
 - Element muss in der Liste gesucht werden.
 - Best Case: $O(1)$
 - Worst Case: $O(\log n)$
 - Average Case: $O(\log n)$

- Günstig, falls oft gesucht wird
(aber selten eingefügt oder gelöscht)

Zeitkomplexität (bis n=18)



Zeitkomplexität (bis n=250)



Faktor in Problemgröße bei 10 mal schnellerem Rechner

10
8
3.162
2.163
1.191

Klasse	Bezeichnung	Beispiel
1	konstant	elementarer Befehl
$\log n$	logarithmisch	Binäre Suche (2er-Logarithmus)
n	linear	Minimum einer Folge
$n \log n$	überlinear	effiziente Sortierverfahren
n^2	quadratisch	einfache Sortierverfahren
n^3	kubisch	Matrizeninversion
n^k	polynomiell	Lineare Programmierung
2^{cn}	exponentiell	Erschöpfende Suche, Backtracking
$n!$	Fakultät	Permutationen, Handlungsreisender

Laufzeit: $t = n^2 + 100 n$ (ms)

n = 5: Laufzeit 0.525 sec

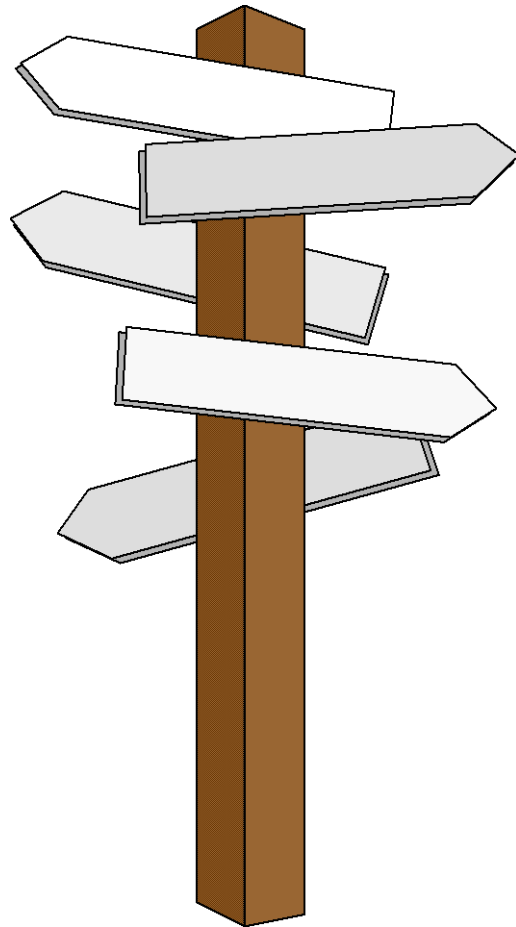
n = 10: Laufzeit 1.100 sec

n = 33: Laufzeit 4.389 sec

⇒ **IRRIGE** Annahme: lineares Wachstum: $t = 110 n$ (ms)

⇒ n = 200: Laufzeit: 22 sec

In Wirklichkeit sind es 420 sec = 7 min



Vorlesung 2

Abstrakte Datentypen
Dynamische Felder
Zirkuläre dynamische Felder

Datentypen

Atomare Datentypen

(elementare, skalare Datentypen;
Java: primitive Datentypen)

boolean char byte
short int long float
double

Datenstrukturen

(strukturierte, zusammen-
gesetzte Datentypen;
Java: Objekte)

Vektor = Strukturierter Datentyp (Datenstruktur) mit

- Datenelementen,
- Strukturregeln
- speziellen Operationen

$$\vec{a} = \begin{pmatrix} 2.1 \\ 0.1 \\ 6.3 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 5.3 \\ 2.2 \\ 3.2 \end{pmatrix}$$

$$\vec{c} = \vec{a} + \vec{b} = \begin{pmatrix} 2.1 \\ 0.1 \\ 6.3 \end{pmatrix} + \begin{pmatrix} 5.3 \\ 2.2 \\ 3.2 \end{pmatrix} = \begin{pmatrix} 2.1 + 5.3 \\ 0.1 + 2.2 \\ 6.3 + 3.2 \end{pmatrix} = \begin{pmatrix} 7.4 \\ 2.3 \\ 9.5 \end{pmatrix}$$

In einer *homogenen Datenstruktur* haben alle Komponenten den gleichen Datentyp.

In einer *heterogenen Datenstruktur* haben die Komponenten unterschiedliche Datentypen.

- 2 Prinzipien:
 - Kapselung: Zu einem ADT gehört eine Schnittstelle. Zugriffe auf die Datenstruktur erfolgen ausschließlich über die Schnittstelle (z.B.: `addElement(...)`, `deleteElement(...)`)
 - Geheimnisprinzip: Die interne Realisierung eines ADT-Moduls bei der Umsetzung bleibt verborgen.

- Viele wichtige abstrakte Datentypen werden in Java durch *Interfaces* beschrieben.
- Es gibt ein oder mehrere Implementierungen dieser Interfaces mit unterschiedlichen dahinter stehenden Konzepten.
 - **Package `java.util`**
- Ähnliches gilt für andere Bibliotheken, z.B. .NET:
 - **Namespace `System.Collections.Generic`**

- Abstrakte Datentypen können unterschiedlich realisiert werden.
- Z.B. der ADT „Liste“ mit der Datenstruktur „dynamisches Feld“ oder „verkettete Liste“.
- Daher gibt es in Java zum Interface „List“ die Implementationen „ArrayList“ und „LinkedList“.

Grundlegende ADTs in dieser Vorlesung

ADT	Java-Interface	Java-Klassen (Bsp.)
Feld	-	(Felder), HashMap
Liste	List	ArrayList, LinkedList, TreeSet*
Menge	Set	HashSet, TreeSet
Prioritätswarteschlange	-	PriorityQueue

* Bestandteil der Bibliothek „Apache Commons“.

Weitere ADTs in dieser Vorlesung

ADT	Kurzbeschreibung	Grund-ADT	Java-Klassen
Stack	Stapel	List	
Queue	Warteschlange	List	
Deque	„Double ended queue“	List	Deque (Interface), ArrayDeque
Map	Assoziatives Feld „Telefonbuch“	Set	Map (Interface), HashMap, TreeMap

Weitere ADTs in dieser Vorlesung

ADT	Kurzbeschreibung	Grund-ADT	Java-Klassen (Bsp.)
BidiMap	Map in 2 Richtungen	Map	BidiMap*, BiMap** (Interf.) HashBiMap**, TreeBidiMap*
Multiset, Bag	Menge, in der doppelte Element erlaubt sind.	Map	Bag*, Multiset** (Interf.) TreeBag*, TreeMultiset**, HashMultiset**

* Bestandteil der Bibliothek „Apache Commons“.

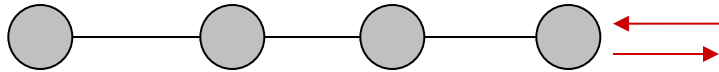
** Bestandteil der Bibliothek „Google Guava“.

- Der ADT **Array(Feld)** hat folgende spezielle Eigenschaften:
 - **Feste Anzahl von Datenobjekten**
 - **Auf jedes Objekt kann direkt (lesend oder schreibend) zugegriffen werden**
 - **In Java: Normales Feld**

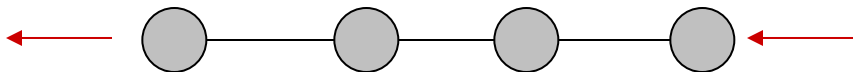


- Für die ADT **Liste(List)** kommt gegenüber einem Feld hinzu:
 - Eine Liste kann wachsen und schrumpfen.
- In Java
 - ADT in Java: Interface **List**
 - 1. Wahl für Implementierung: **ArrayList**
 - Dynamisches Feld
 - Schwächen: Einfügen/Löschen vorne, Mitte, worst case auch hinten
 - Alternativen: **LinkedList** (doppelt verkettete Liste), evtl. **HashMap** (Hashtabelle)

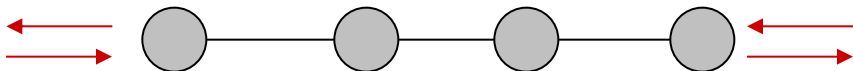
- Spezieller sind Stack, Queue und Deque
 - **Stack** (Stapel, Keller): Daten können an einem Ende hinzugefügt oder entnommen werden.



- **Queue** (Warteschlange): Daten können an einem Ende hinzugefügt und am anderen Ende entnommen werden.



- **Deque** (Double ended queue): Daten können an beiden Enden hinzugefügt und entnommen werden.



- In Java
- Stack:
 - 1. Wahl für Implementierung: **ArrayList** oder **LinkedList**
- Queue und Deque
 - Gleiche Anforderungen an Algorithmus
 - ADT in Java: Interface **Deque**
 - 1. Wahl für Implementierung: Klasse **ArrayDeque**
 - Zirkuläres dynamisches Feld
 - oder **LinkedList**

- Englische Namen: Associative Array, Dictionary, Map
- Wichtigste Operationen
 - **Schlüssel-Wert-Paar einfügen.**
 - **Wert zum Schlüssel finden.**
- In Java
 - **ADT: Java-Interface **Map**.**
 - **1. Wahl bei Implementierung: Klasse **HashMap**.**
 - Datenstruktur Hashtabelle
 - Einfügen/Löschen $O(1)$!
 - **Alternativen: Klasse **TreeMap****
 - Datenstruktur Rot-Schwarz-Baum
 - Einfügen/Löschen $O(\log n)$
 - Daten liegen (im Gegensatz zur Hashtabelle) **sortiert** vor.

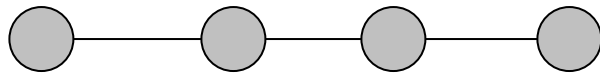
- Eine Menge (*Set*) ist eine Sammlung von Elementen des gleichen Datentyps.
 - Innerhalb der Menge sind die Elemente ungeordnet.
 - Jedes Element kann nur einmal in der Menge vorkommen.
- Wichtigste Operationen: Einfügen, Löschen, Testen (ob vorhanden).
- In Java:
 - ADT: Java-Interface **Set**
 - 1. Wahl bei der Implementierung: Klasse **HashSet**
 - Datenstruktur Hashtabelle
 - Alternative: Klasse **TreeSet**
 - Datenstruktur Rot-Schwarz-Baum, Daten sind **sortiert**.

- Auch **Vorrangwarteschlange** oder **Priority Queue** genannt.
- Eine Warteschlange, deren Elemente einen Schlüssel (Priorität) besitzen.
- Wichtige Operationen bei Prioritätswarteschlangen:
 - Element in Schlange einfügen
 - Element mit der höchsten Priorität entnehmen.
- In Java:
 - **1. Wahl bei Implementierung: Klasse **PriorityQueue**.**
 - Datenstruktur: Dynamisches zirkuläres Feld
 - **Alternative: Klasse **TreeSet**.**

Java	C#
Interface <code>List</code>	Interface <code>ICollection</code>
<code>ArrayList</code>	<code>List</code>
<code>LinkedList</code>	<code>LinkedList</code>
<code>ArrayDeque</code>	<code>Queue</code> , <code>Stack</code>
Interface <code>Map</code>	Interface <code>IDictionary</code>
<code>HashMap</code>	<code>Dictionary</code>
<code>TreeMap</code>	<code>SortedDictionary</code>
Interface <code>Set</code>	Interface <code>ISet</code>
<code>HashSet</code>	<code>HashSet</code>
<code>TreeSet</code>	-

Datenstrukturen zum ADT Liste

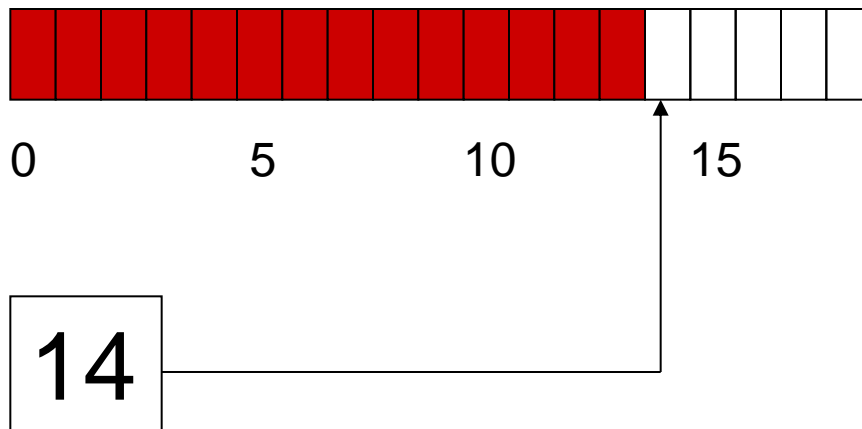
- Gut geeignet: Lineare homogene Datenstrukturen
 - mehrere gleichartigen **Datenobjekte**
 - **1:1 – Relation**



- Die wichtigsten linearen homogenen Datenstrukturen sind:
 - Das **Feld**, eventuell **dynamisch** und/oder **zirkulär** (heute)
 - Die **verkettete Liste** (Vorlesung 5)

- Lineare homogene Datenstrukturen eignen außer für den ADT **Liste** noch für
- den ADT **Feld**,
- sowie die Spezialfälle der Liste **Stack**, **Queue** und **Deque**.

- Ein dynamisches Feld besteht aus:
 - Einem (normalen) Feld, das nicht vollständig gefüllt ist.
 - Einem Zeiger, der anzeigt, welches das erste unbesetzte Element ist.



Dynamisches Feld:

Solange das neue Element noch in das Feld passt:

```
list[n] = newEntry;  
n++;
```

⇒ Einfügen eines Elements: $O(1)$

- Wenn das neue Element nicht mehr in das Feld passt:
 - **Ein Feld mit einer größeren Anzahl von Elementen wird angelegt.**
 - **Alle bisherigen Elemente werden umkopiert.**
 - **Das neue Element wird angehängt.**

⇒ $O(???)$

Einfügen hinten (2)

- Nehmen wir als Beispiel: 1000 Elemente sollen in die Liste eingefügt werden.
- Java nimmt als Anfangsgröße für das Feld 10
 - **Dies kann (und sollte auch) verändert werden, falls die maximale Größe der Liste vorher bekannt ist.**
- Immer wenn das Feld zu klein wird, wird ein neues Feld mit doppelt so vielen Elementen erzeugt und alle Elemente umkopiert:
 - **Dies ist eine Vereinfachung zur besseren Rechnung. In Wirklichkeit beträgt der Faktor:**
 - **Sun-Java: 1,5**
 - **Gnu-Java: 2**
 - **C# (Mono): 2**
 - **Python: 1,125**

Element Nr.	Neue Feldgröße	Umzukopierende Elemente
11	20	10
21	40	20
41	80	40
81	160	80
161	320	160
321	640	320
641	1280	640

Summe: $s = 1270$

- Wird eine Liste aus n Elemente aufgebaut, ist die Anzahl der kopierten Elemente immer $\leq 2n$
- \Rightarrow Aufbau von n Elementen: $O(n)$

Aufbau einer Liste von n Elementen durch Einfügen neuer Elemente von vorne:

Bei jedem Einfügen müssen alle Elemente einen Platz nach hinten kopiert werden

Das ergibt:

- $1+2+3+\dots+(n-1)=\frac{n \cdot (n-1)}{2}$ Kopiervorgänge
- $\Rightarrow O(n^2)$

Auslesen

- Dynamisches Feld: Elemente können generell direkt ausgelesen werden.

⇒ $O(1)$

	Dyn. Feld
Java-Klasse	Array List
Auslesen Anfang	$O(1)$
Auslesen Mitte	$O(1)$
Auslesen Ende	$O(1)$
Einfügen/Löschen Anfang	$O(n)$
Einfügen/Löschen Mitte	$O(n)$
Einfügen/Löschen Ende	avg. $O(1)$ wst. $O(n)$

- Ein dynamisches Feld ist für einen Stack gut geeignet.
 - Einfügen Ende + Auslesen Ende $O(1)$
- Ein dynamisches Feld ist für eine Queue schlecht geeignet.
 - Entweder Einfügen oder Auslesen $O(n)$
- Daher werden für Queues **zirkuläre dynamische Felder** benutzt:
 - Manchmal auch doppelt verkettete Listen (später)

14	15	0	1	2
13				3
12				4
11				5
10	9	8	7	6

Umlauf im Uhrzeigersinn:

```
i = (i+1) % data.length;
```

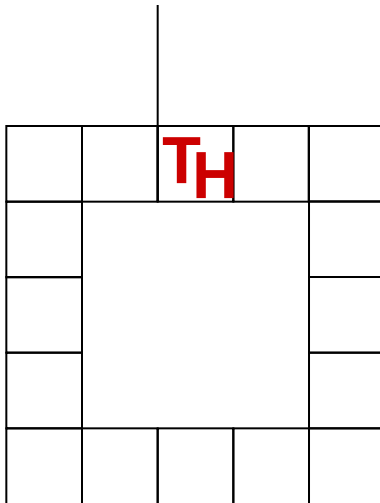
Umlauf gegen den Uhrzeigersinn:

```
i = (i-1+data.length) % data.length;
```

Queue mit Hilfe eines Feldes (1)

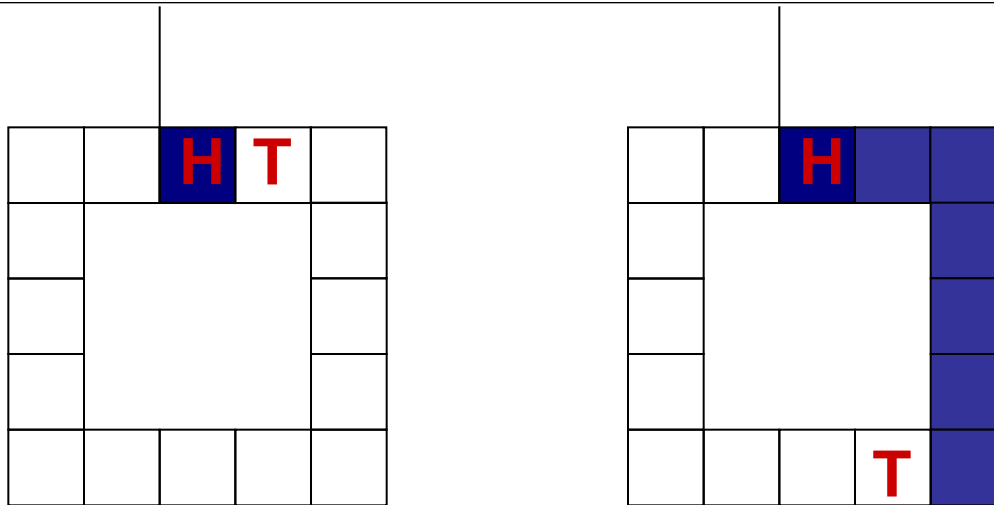
```
public class ArrayQueue {  
    private int head;  
    private int tail;  
    private Object [] data;  
  
    ...  
}
```

Queue: Methode isEmpty()



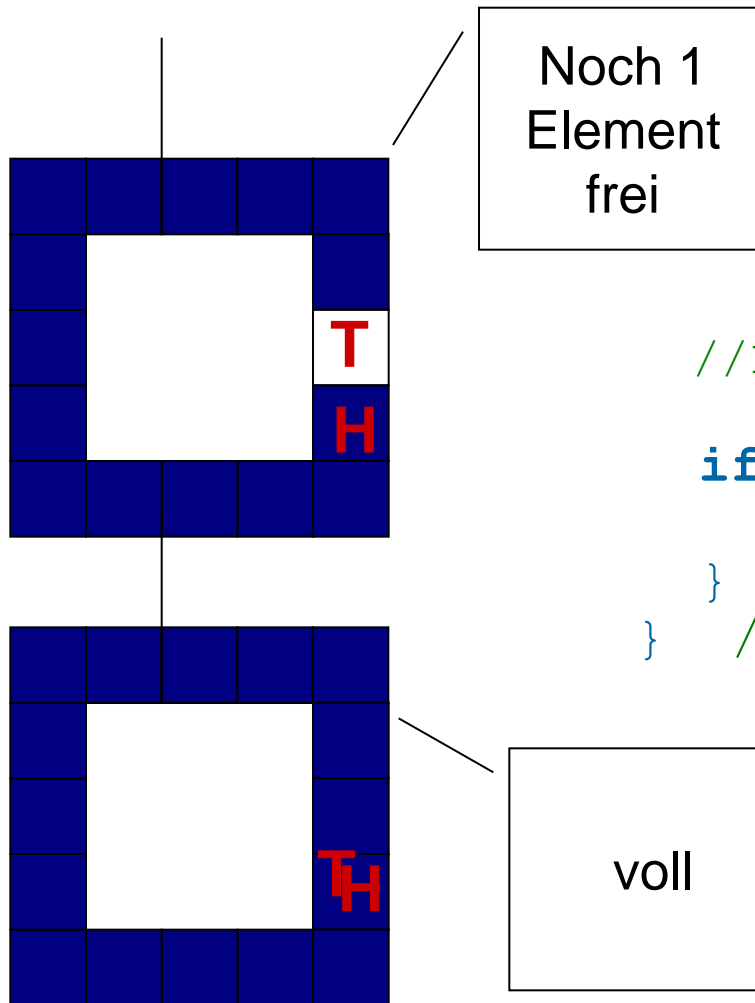
```
private void clear() {  
    head = 0;  
    tail = 0;  
}  
  
public boolean isEmpty() {  
    return head == tail;  
}
```

Queue: Methode put(..)



```
public void put(Object o) {  
    data[tail] = o;  
    tail = (tail+1) % data.length;  
  
    //Fortsetzung naechste Folie
```

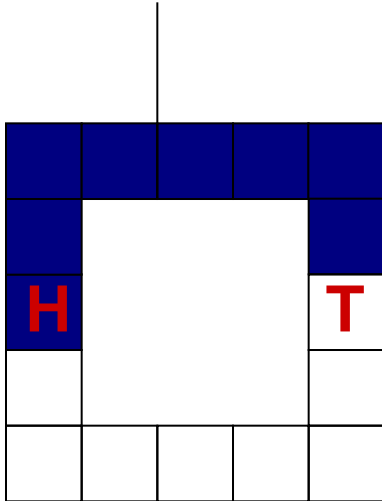
Queue: Methode put() (2)



//Fortsetzung von voriger Folie

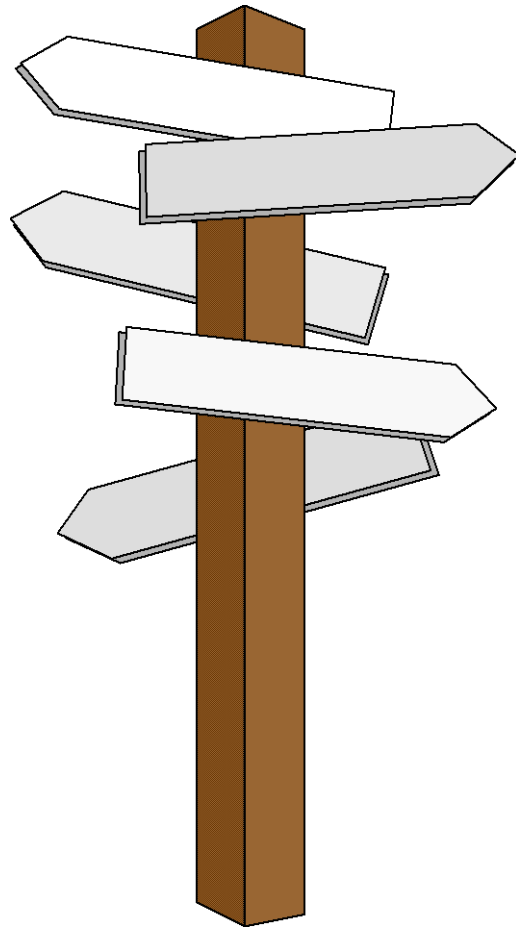
```
if (head==tail) {  
    resize();  
}  
} //Ende der Methode
```

Queue: Methode get()



```
public Object get() throws QueueException {  
    if (isEmpty()) {  
        throw new QueueException  
            ("Queue ist leer!");  
    }  
    Object ret = data[head];  
    head = (head+1) % data.length;  
  
    return ret;  
}
```

- Wiederholung: Eine Deque ist eine lineare Datenstruktur, bei der die Daten an beiden Enden eingefügt oder entfernt werden können.
- Die geeigneten Datenstrukturen entsprechen denen der Queue.
 - **Es wird dazu übergegangen, statt einer Queue gleich eine Deque zu implementieren.**
 - **So auch in Java:**
 - Klasse **ArrayDeque** (zirkuläres dynamisches Feld) für Queues
 - Zugehöriges Interface: **Deque**



Vorlesung 3

Hashtabellen

Eine Menge (*Set*) ist eine Sammlung von Elementen des gleichen Datentyps.

Innerhalb der Menge sind die Elemente ungeordnet.

Jedes Element kann nur einmal in der Menge vorkommen.

Operationen:

- **Elementare Operationen: Einfügen, Löschen, Testen (ob vorhanden).**
- **Darauf aufbauende Operationen: Typische Operationen aus der Mengenlehre: Schnittmenge, Restmenge, ...**

- Die Datenstruktur Set ist nur in wenigen Programmiersprachen direkt in der Sprache verfügbar (z.B. PASCAL).
- In Java ist *Set* ein Interface, das (unter anderem) folgende Klassen implementieren:
 - **TreeSet** (keine .NET-Entsprechung): Basiert auf der Datenstruktur **Rot-Schwarz-Baum**, implementiert Erweiterung **SortedMap**.
 - **HashSet** (.NET: **HashSet**): Basiert auf der Datenstruktur **Hash-Tabelle**.

- Englische Namen: Associative Array, Dictionary, Map
- Sonderform des Feldes
 - **Verwendet keinen numerischen Index zur Adressierung eines Elements (wie $a[1]$).**
 - **Verwendet zur Adressierung einen Schlüssel (z.B. in der Form $a["Meier"]$)**
- Operationen
 - **Feldeintrag einfügen**
 - **Feldeintrag auslesen**
 - **Feldeintrag löschen**

- Assoziative Felder eignen sich dazu, Datenelemente in einer großen Datenmenge aufzufinden.
- Jedes Datenelement wird durch einen eindeutigen Schlüssel identifiziert.

`daten[suchschluessel]=datenelement`

- Beispiel: Telefonbuch
 - **Datenelement: (Name, Adresse, Telefonnummer)**
 - **Suchschlüssel: Name**

- ADT entspricht dem Interface `java.util.Map`
- 2 Implementationen
 - **TreeMap** (.NET: **SortedDictionary**): Basiert auf der Datenstruktur **Rot-Schwarz-Baum**, implementiert Erweiterung **SortedMap**.
 - **HashMap** (.NET: **Dictionary**): Basiert auf der Datenstruktur **Hash-Tabelle**.
- Assoziative Felder gibt es auch in vielen anderen Sprachen (gewöhnlich als Hash-Tabelle):
 - **C++ (Map), Python (Dictionary), Ruby (Hash), Perl, PHP, Visual Basic, ...**

- Am Beispiel: Telefonbuch als ass. Feld
- Groovy (in C# fast identisch):

```
Map tel = new HashMap();  
tel["Mustermann"]=123456;  
System.out.println(tel["Mustermann"]);
```

- Java

```
Map tel = new HashMap();  
tel.put("Mustermann", 123456);  
System.out.println(tel.get("Mustermann"));
```

- Beispiel in Java (Klasse `HashMap`).

...

```
HashMap<String, Integer> map =  
    new HashMap<String, Integer>();  
map.put("Januar", 1);  
map.put("Februar", 2);  
map.put("Maerz", 3);  
System.out.println(map.get("Februar"));   ⇒ 2
```

- Man kann auch prüfen, ob ein String ein Monat ist

```
boolean s = map.containsKey("Januar"); ⇒ true
```


- Ein assoziatives Feld kann immer auch als Menge verwendet werden:
 - **Das Datenelement ist dann ein Dummy-Element.**

- Eine Menge kann leicht zu einem assoziatives Feld umgewandelt werden.
- ```
public class Element {
 Object key;
 Object data;

 //Zwei Elemente sind gleich, falls
 //die Schluessel gleich sind
 public boolean equals(Element b) {
 return (this.key.equals(b.key));
 }
}
```

- Diskussion der Datenstrukturen ist für beide ADT gleich.
- Im folgenden werden die Datenstrukturen immer am Beispiel der Menge vorgestellt.
  - **Die Erklärungen sind dann etwas einfacher.**
- Die Hauptanwendungen sind aber assoziative Felder.
  
- In Java:
  - **HashSet**  $\longleftrightarrow$  **HashMap**
  - **TreeSet**  $\longleftrightarrow$  **TreeMap**

- Assoziative Felder können sehr unterschiedlich verwirklicht werden:
- Einfache Lösung: Verwendung einer ArrayList:
  - Einfügen: `add(element)`
  - Löschen: `remove(element)`
  - Prüfen: `contains(element)`
  - Auslesen: `get(indexOf(element))`
- Aufwendige Lösung: Verwendung einer Datenbank
  - **MySQL, Oracle, Access, ...**

Wichtige Fragen:

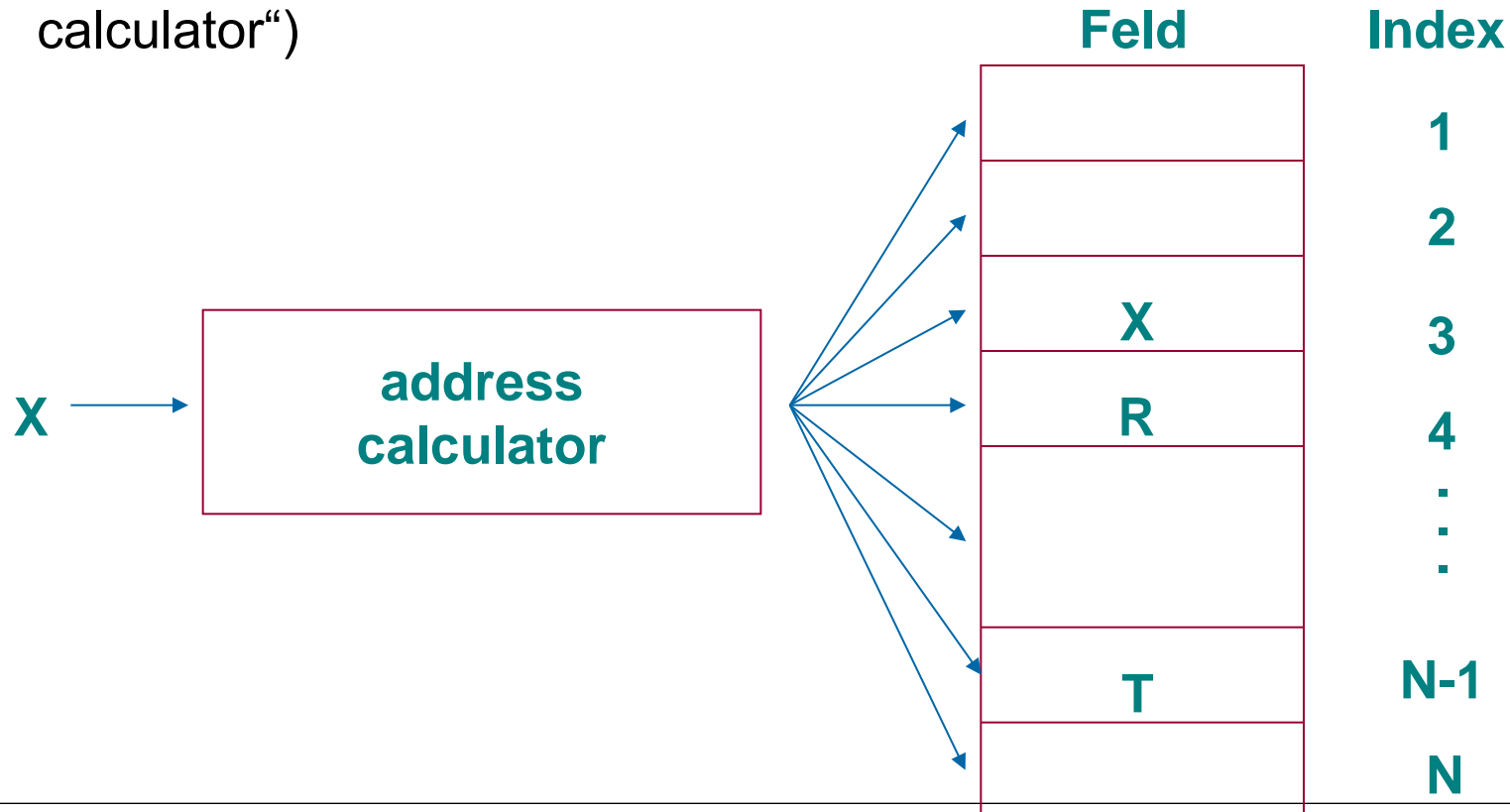
- Welche Datenstrukturen werden an welcher Stelle verwendet?
- Welche Datenstrukturen sind wann geeignet?
  
- Folgende Datenstrukturen werden untersucht:
  - **Die schnellste Datenstruktur (Hashtabelle) wird nachfolgend betrachtet.**
  - **Weitere Datenstrukturen (balancierte Bäume) betrachten wir später.**

(Hashtables / Streuwerttabellen)

- Einführung: Hashfunktion und Kollision
- Hashing in Teillisten
- Offene Adressierung (Sondieren)

# Grundprinzip einer Hash-Tabelle

- Elemente werden in einem großen Feld gespeichert.
- In welchem Feldelement ein bestimmter Eintrag gespeichert wird, berechnet eine **Hashfunktion**  $h(X)$  aus dem Schlüssel  $X$  („address calculator“)



$S$  sei eine Schlüsselmenge und  $I$  eine **Adressmenge** im weitesten Sinn.

Dann heißt  $h : S \rightarrow I$  **Hash - Funktion**.

Die Bildmenge  $h(S) \subseteq I$  bezeichnet die Menge der **Hash-Indizes**.

$|I| =$  Größe  $N$  der  
Hash-Tabelle

## Bemerkung:

Die Schlüsselmenge ist im allgemeinen *sehr viel* größer als die Adressmenge.

Deshalb wird eine Hash-Funktion **surjektiv** ( $h(S) = I$ ),  
aber **nicht injektiv** sein ( $h(s)=h(t) \nRightarrow s=t$ ) !

Alle Adressen  
werden  
erreicht.



- Hash-Funktionen hängen ab von
  - Datentypen der Schlüssel
  - Anwendung
- Integer: **Divisionsrest-Methode** („Divisions-Hash“):
$$h(x) = x \bmod N$$
  - Bevorzugtes Verfahren, wenn die Schlüsselverteilung nicht bekannt ist.
  - Etwaige Regelmäßigkeiten in der Schlüsselverteilung sollten sich nicht in der Adressverteilung auswirken.
  - Daher sollte  $N$  eine Primzahl sein.
  - Andere Methoden: Faltung, Mid-Square, ...

- Andere Datentypen: Rückführung auf Integer.
  - **Alle Datentypen:** Verwenden der Speicheradresse.
  - **Strings:** ASCII/Unicode-Werte addieren (evtl. von einigen Buchstaben, evtl. gewichtet)

Beispiel für Hash-Funktion:  $h(s) = s \bmod 10$

| Index | Eintrag |
|-------|---------|
| 0     |         |
| 1     |         |
| 2     | 42      |
| 3     |         |
| 4     |         |
| 5     |         |
| 6     |         |
| 7     |         |
| 8     |         |
| 9     | 119     |

Einfügen von 69  
führt zu **Kollision**

h3(s):

```
int index = ord(s.charAt(0));
index += ord(s.charAt(1));
index += ord(s.charAt(2));
index = index % 17;
```

0:

1:

2:

3: Mai, September

4:

5: Januar

6: Juli

7: März

8: Juni

9: August, Oktober

10: Februar

11:

12:

13:

14: November

15: April, Dezember

16:

- Collections Framework unterstützt Hashtabellen, z.B. mit Klasse `HashMap`
- `java.lang.Object` definiert Methode  
`int hashCode()`
  - berechnet Hash-Wert für Objekt (u.U. aus Speicheradresse)
  - `java.lang.Integer` und `java.lang.String` liefern eigene Implementierungen
  - kann in selbstgeschriebenen Klassen überschrieben werden
- `hashCode()` dient als Basis für Hashfunktionen ...

## Definition:

Sei  $S$  Schlüsselmenge,  $h$  Hash-Funktion.

Ist für  $s_1 \neq s_2$  (mit  $s_i \in S$ )

$$h(s_1) = h(s_2),$$

so spricht man von einer **Kollision**.

- Wahrscheinlichkeit von Kollisionen ist abhängig von Hashfunktion  
⇒ Hash-Funktionen sollten möglichst gut **streuen!**
- Außerdem: Hash-Funktion muss effizient berechenbar sein.
- Wahl einer guten Hash-Funktion schwierig (→ Wikipedia)

Annahme:

ideale Hash-Funktion, d.h. **gleichmäßige Verteilung** über die Hash-Tabelle

„**Geburtstagproblem**“:

Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 von  $n$  Leuten auf einer Party am gleichen Tag Geburtstag haben?

Analogie zum Geburtstagsproblem:

- $m = 365$  Tage = Größe Hash-Tabelle (bisher  $N$  genannt)
- $n$  Personen = Zahl Elemente

$p(i;m)$  := Wahrscheinlichkeit, dass  $i$ -ter Schlüssel auf freien Platz abgebildet wird ( $i=1, \dots, n$ ), wie alle Schlüssel vorher:

$$p(1;m) = \frac{m-0}{m} = \left(1 - \frac{0}{m}\right) \quad (\text{alle Plätze frei})$$

$$p(2;m) = \frac{m-1}{m} = \left(1 - \frac{1}{m}\right) \quad (\text{ein Platz belegt, } m-1 \text{ Plätze frei})$$

⋮

$$p(i;m) = \frac{m-i+1}{m} = \left(1 - \frac{i-1}{m}\right) \quad (i-1 \text{ Plätze belegt, } m-i+1 \text{ Plätze frei})$$

Wahrscheinlichkeit für „keine Kollision“:

$$P(\text{NoKol} \mid n, m) = \prod_{i=1}^n p(i;m) = \prod_{i=0}^{n-1} \left(1 - \frac{i}{m}\right)$$

Wahrscheinlichkeit für „mindestens eine Kollision“:

$$P(\text{Kol} \mid n, m) = 1 - P(\text{NoKol} \mid n, m)$$



## Tabelle zum Geburtstagsproblem (m=365)

| n   | $Pr(\text{Kol} n,m)$ |
|-----|----------------------|
| 10  | 0,117                |
| 20  | 0,411                |
| ... | ...                  |
| 22  | 0,476                |
| 23  | 0,507                |
| 24  | 0,538                |
| ... | ...                  |
| 30  | 0,706                |
| 40  | 0,891                |
| 50  | 0,970                |

Schon bei 23 Gästen ist die Wahrscheinlichkeit für „Geburtstagspärcchen“ > 0,5!  
(„Geburtstagsparadoxon“)

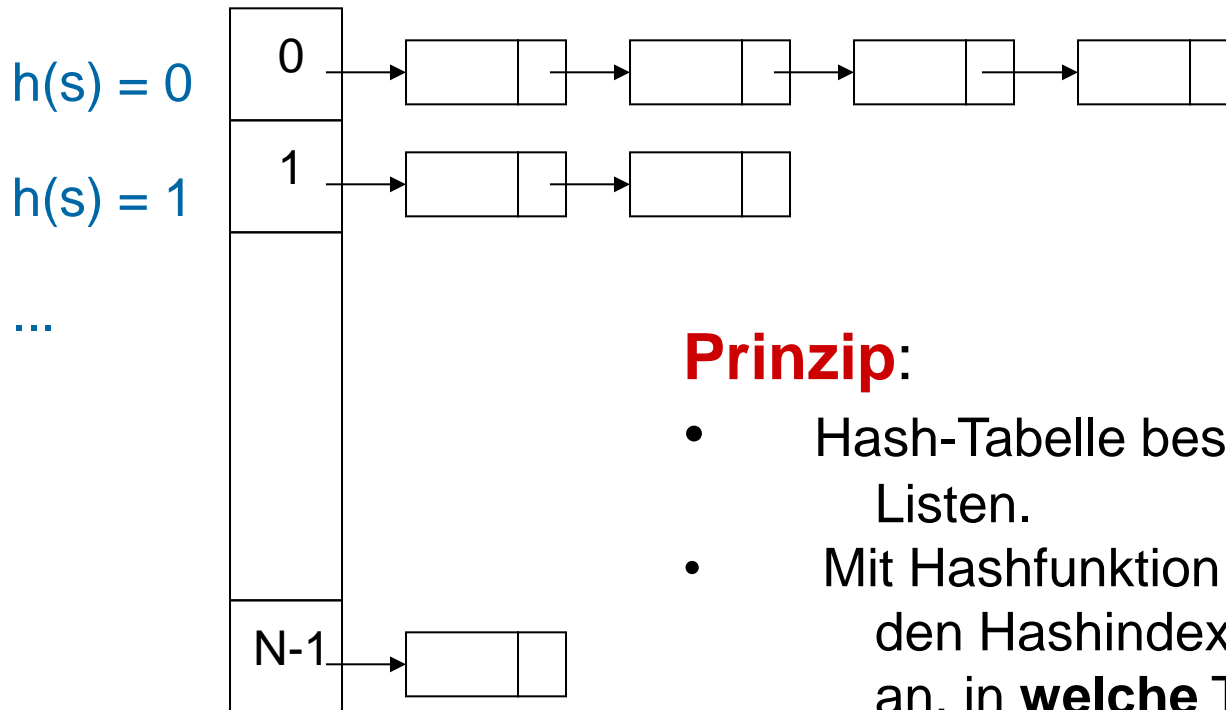
Bei 50 Gästen ist „Kollision“ fast sicher

**Situation:** zwei Einträge werden auf die gleiche Feldadresse abgebildet,  
d.h.  $h(x_1) = h(x_2)$

**Strategien** bei Kollisionsbehandlung (u.a.):

- a) „**Verkettung** der Überläufer“, „Hash in **Teillisten**“:  
Liste für alle Elemente, die die gleiche Position belegen.
  
- b) „**Sondieren**“, „Hashing mit **offener Adressierung**“:  
Suchen einer alternativen Position innerhalb des Feldes. Wir betrachten:
  - 1. Lineares Sondieren
  - 2. Doppeltes Hashing
  - 3. Quadratisches Sondieren

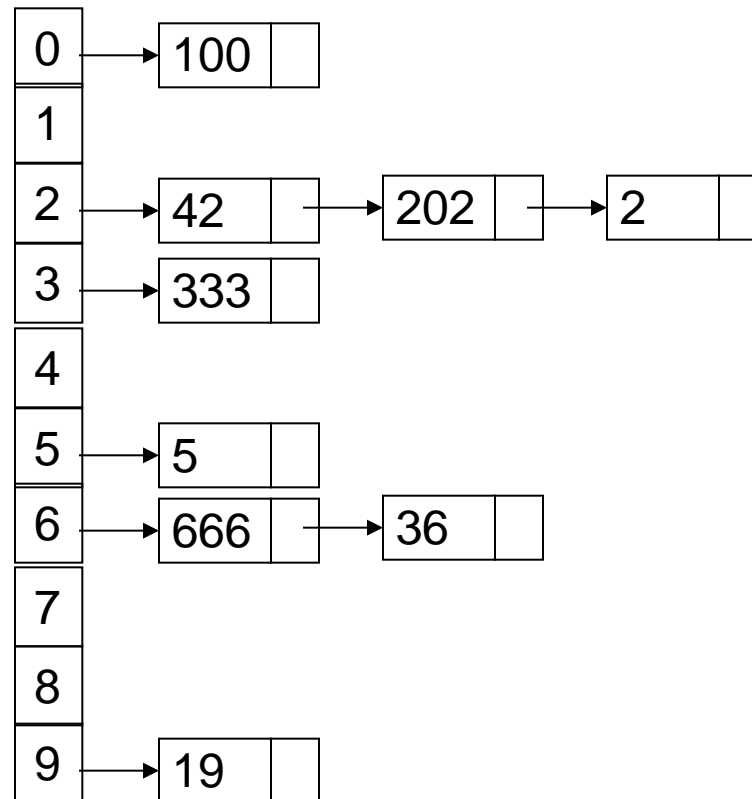
# Hash in Teillisten



## Prinzip:

- Hash-Tabelle besteht aus  $N$  linearen Listen.
- Mit Hashfunktion  $h$  aus Schlüssel  $s$  den Hashindex  $h(s)$  berechnen (gibt an, in **welche** Teilliste der Datensatz gehört).
- Dann **innerhalb** der Teillisten sequentiell speichern.

$$h(i) = i \bmod 10$$



Die **Schrittzahl**  $S(s)$ , die nötig ist, um den Datensatz mit Schlüssel  $s$  zu speichern bzw. wiederzufinden, setzt sich bei Hashing in Teillisten zusammen aus

- der Berechnung der Hash-Funktion und
- dem Aufwand für die Suche/Speicherung innerhalb der Teilliste.

Der **Füllgrad** einer Hash-Tabelle ist der Quotient

$$\alpha = n/N,$$

mit

- $N$  := Größe der Hash-Tabelle (# Adressen) und
- $n$  := # gespeicherte Datensätze (normalerweise ist  $N \geq n$ ).

$N = \#$  Teillisten;  $n = \#$  gespeicherte Datensätze

Füllgrad der Hashtabelle:  $\alpha = n/N$

- Bei idealer Speicherung entfallen  $\alpha$  Elemente auf jede Teilliste.

- Schrittzahl im Mittel:

Hash-Index berechnen.

– erfolgreiche Suche:  $c_1 + c_2 \cdot \alpha/2$

– erfolglose Suche:  $c_1 + c_2 \cdot \alpha$

Lineare Suche in Teilliste.

⇒ Suchaufwand:  $O(\alpha) = O(n/N)$

⇒ wird der Füllgrad zu groß, sollte die Hashtabelle vergrößert werden (dynamisches Hashing).

- Um zu viele Kollisionen zu vermeiden, muss die Hash-Tabelle ab einem gewissen Füllgrad **vergrößert** werden  
→ **dynamisches Hashing**
- Folge einer Vergrößerung: Die gesamte Hashtabelle muss neu aufgebaut werden.
  - **Sowohl beim Hash in Teillisten als auch beim Hash mit offener Adressierung.**
  - **Es sollte stets gelten:  $\alpha < 1/2$  (nach Sedgewick).**

- **Speichern:**
  - Hashindex mit Hashfunktion aus Schlüssel des Datensatzes berechnen.  
Falls Speicherplatz frei: dort speichern
  - Bei Kollision: **Ersatzadresse** berechnen und Speicherversuch wiederholen  
Falls berechneter Speicherplatz erneut belegt: Neue Ersatzadresse berechnen; solange bis
    - freier Platz gefunden oder
    - verfügbarer Speicher ganz durchlaufen (⚡)
- **Suchen:** analog zum Ablauf beim Speichern
- **Löschen** ist aufwändig!



## Definition:

Wird die Ersatzadresse bei jeder Kollision durch Erhöhen der alten Adresse um 1 berechnet, so spricht man von **linearem Sondieren** („linear probing“).

Die *i-te Ersatzadresse* für einen Schlüssel  $s$  mit Hash-Index  $h(s)$  wird also wie folgt berechnet:

$$h_i(s) = ( h(s) + i ) \bmod N$$

( $h_0(s) = h(s)$ : Hashindex der Hashfunktion selbst)

- **Aufgabenstellung:**
  - Eine Firma mit zur Zeit 60 Mitarbeitern vergibt Personalnummern. Da Nummern von ehemaligen Mitarbeitern nicht neu vergeben werden, hat man maximal vierstellige Nummern vorgesehen. (**Schlüsselmenge:**  $S = \{1 \dots 9999\}$ ).
  - 100 Speicherplätze für Datensätze (von 0 an nummeriert) zur Speicherung der Personaldaten (**Hash-Indizes:**  $I = \{0, \dots, 99\}$ ).
- **Hash-Funktion:**  $h : S \rightarrow I$  mit  $h(s) = s \bmod 100$
- **Bei Kollision:**
  - berechneten Hashindex solange um 1 (modulo 100) erhöhen, bis freier Platz (Speichern) bzw. gesuchter Eintrag (Suchen) gefunden wird.

## Beispiel zum Sondieren (2)

**Einfügen:**

| Schlüssel | Name    | h(s) |                                                          |
|-----------|---------|------|----------------------------------------------------------|
| 1233      | Müller  | 33   |                                                          |
| 2034      | Meier   | 34   |                                                          |
| 9539      | Schulze | 39   |                                                          |
| 3433      | Schmitz | 33   | 1.Kollision, Ersatzadr. 34<br>2.Kollision, Ersatzadr. 35 |

**Wiederfinden:**

| Schlüssel | Name    | h(s) |                                                                                        |
|-----------|---------|------|----------------------------------------------------------------------------------------|
| 2034      | Meier   | 34   | Datensatz gefunden                                                                     |
| 3433      | Schmitz | 33   | nicht gefunden; Ersatzadr. 34<br>nicht gefunden; Ersatzadr. 35<br>Datensatz gefunden   |
| 7334      | Huber   | 34   | nicht gefunden; Ersatzadr. 35<br>nicht gefunden; Ersatzadr. 36<br>leer: nicht gefunden |

Nachteil des Divisions-Hash mit linearem Sondieren:

- es bilden sich Ketten belegter Plätze. Formal:

$$h_0(s) \neq h_0(t) \wedge h_i(s) = h_k(t) \Rightarrow h_{i+1}(s) = h_{k+1}(t)$$

⇒ erhöht Wahrscheinlichkeit für weitere Kollisionen in diesem Bereich.

- Diesen Effekt nennt man **primäre Häufung** („primary clustering“).



Alle Hashindizes, die auf die Häufung treffen, landen auf dem ersten Element nach der Häufung und tragen zu ihrer Vergrößerung bei.

- Ähnlich: **Sekundäre Häufung** („secondary clustering“)  
(hängt von Hashfunktion ab):

$$h_0(s) = h_0(t) \Rightarrow h_i(s) = h_i(t)$$

- Wenn primäre Häufungen durch bessere Verfahren vermieden werden, können immer noch sekundäre Häufungen entstehen, wenn viele (ursprüngliche) Hash-Adressen gleich sind.

- Ähnlich wie lineares Sondieren.
- Der Schlüssel wird nicht um 1 erhöht, sondern der Inkrement wird mit einer **zweiten Hashfunktion** berechnet.
- Beseitigt praktisch die Probleme der primären und sekundären Häufung.
- Nicht alle Felder werden durchprobiert. Im ungünstigen Fall kann ein neues Element nicht eingefügt werden, auch wenn noch Felder frei sind.
- Bei der Verwendung von Speicheradressen als Hash-Index können die unteren Bits als Hash-Index und die mittleren Bits als Inkrement benutzt werden.

- Für Mengen bzw. assoziative Felder sind Hash-Tabellen anderen Datenstrukturen überlegen (außer in Spezialfällen).
- Die Unterschiede zwischen den Hash-Verfahren sind vergleichsweise gering.
- Gewöhnlich wird gewählt:
  - **Hashing in Teillisten oder doppeltes Hashing**
  - **generell in Kombination mit dynamischem Hashing.**

| Sprache        | Klasse             | Variante            | Max. Füllgrad |
|----------------|--------------------|---------------------|---------------|
| Sun-Java       | HashMap, Hashtable | Hash mit Teillisten | 3/4           |
| Gnu-Java (gcj) | HashMap, Hashtable | Hash mit Teillisten | 3/4           |
| C# / Mono      | Dictionary         | Doppeltes Hashing   | 3/4           |
| Python         | Dictionary         | Mehrfaches Hashing  | 2/3           |
| Ruby           | Hash               | Hash mit Teillisten | 5             |

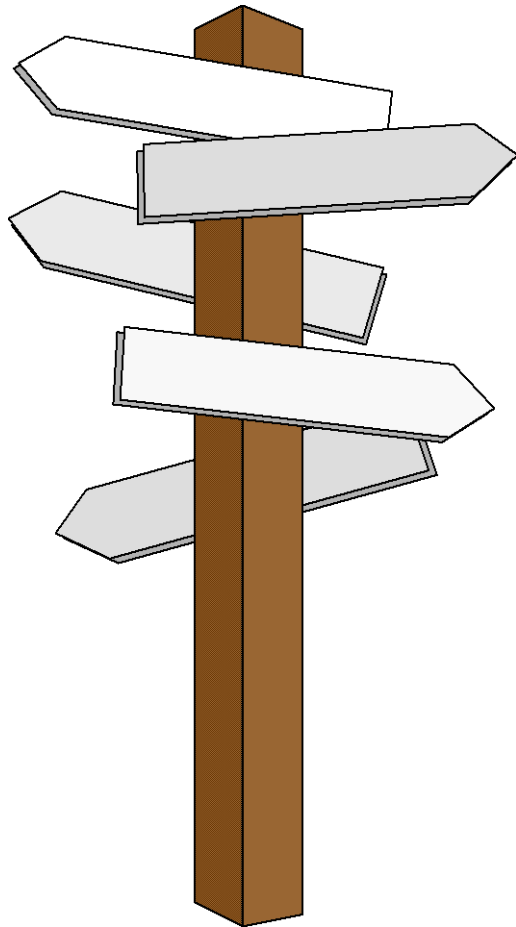


- Beispielhaft für dynamische Hashtabellen mit Teillisten:
- Suchen:
  - **Durchschnittlich werden  $\alpha$  Elemente durchsucht. Bei dynamischen Tabellen gilt:  $\alpha < c$  ( $c$  fest).**  **$O(1)$**
- Einfügen:
  - **Suchen ( $O(1)$ )**
  - **Einfügen ( $O(1)$ )**
  - **Eventuell Neuaufbau der Tabelle. Dies ist durchschnittlich  $O(1)$  aus den gleichen Gründen wie bei der dynamischen Liste**  **$O(1)$**
- Löschen
  - **Suchen ( $O(1)$ )**
  - **Löschen ( $O(1)$ )**  **$O(1)$**

- Bei ungünstiger Verteilung der Hashwerte (alle Elemente haben denselben Hashwert) ist der Suchaufwand  $O(n)$ .
- Muss die Hashtabelle vergrößert werden, müssen alle Hashwerte neu berechnet werden.
  - **Für einzelne Elemente sehr lange Funktionszeiten möglich (worst case  $O(n)$ ).**

## Komplexität von Mengen

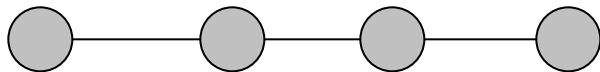
|          | Unsortierte Liste | Sortierte Liste | Baum (später) | (Dyn.) Hashtabelle    |
|----------|-------------------|-----------------|---------------|-----------------------|
| Suchen   | $O(n)$            | $O(\log n)$     | $O(\log n)$   | $O(1)$                |
| Einfügen | $O(n)$            | $O(n)$          | $O(\log n)$   | $O(1)$<br>wst: $O(n)$ |
| Löschen  | $O(n)$            | $O(n)$          | $O(\log n)$   | $O(1)$                |



## Vorlesung 4

## Verkettete Listen

- Linearen homogenen Datenstrukturen ist gemeinsam:
  - mehrere gleichartigen **Datenobjekte**
  - **1:1 - Relation**

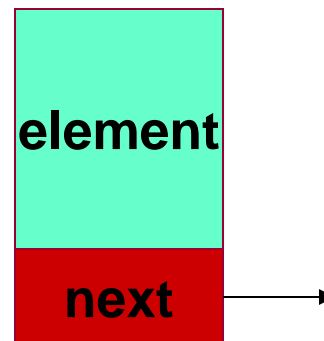


- Sie sind besonders geeignet für die ADT Liste, Stack, Queue und Deque.

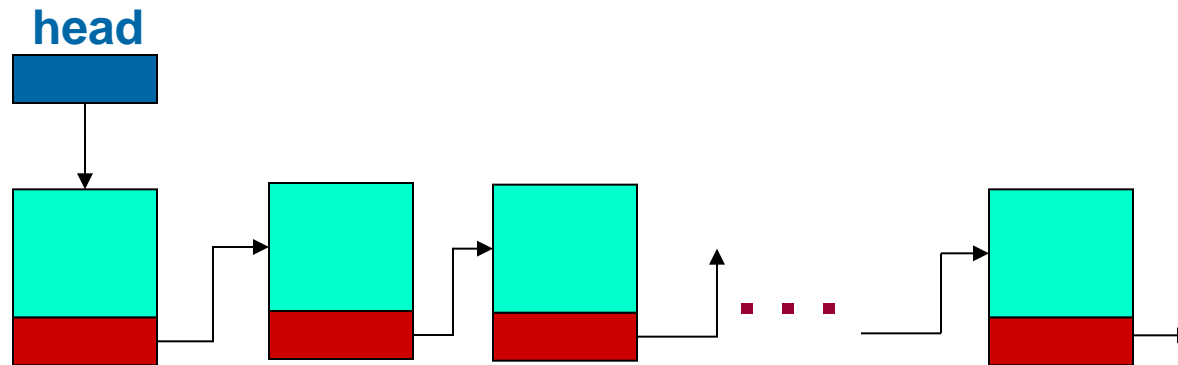
- Die wichtigsten linearen homogenen Datenstrukturen sind:
- Die **verkettete Liste** (heute) und das
- **Feld** (eventuell **dynamisch** und/oder **zirkulär**)
  - **Felder wurden in Vorlesung 2 behandelt.**

Eine **Einfach verkettete Liste** besteht aus **Knoten**.

```
class Node {
 Object element; // Datenkomponente
 Node next; // Verweis auf Restliste
}
```



- Durch Hintereinanderhängen von Knoten entsteht eine Liste



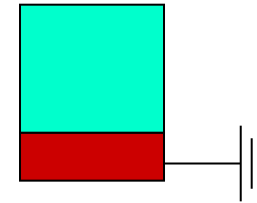
- Die Listenklasse muss sich nur den Kopf merken
- Jeder Knoten zeigt auf seinen Nachfolger

```
class MyList {
 Node head;

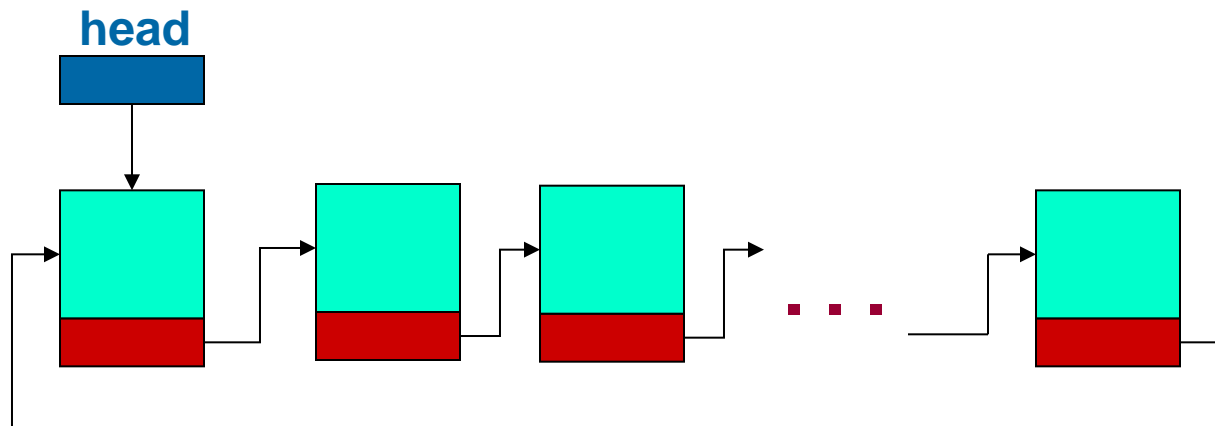
}
```

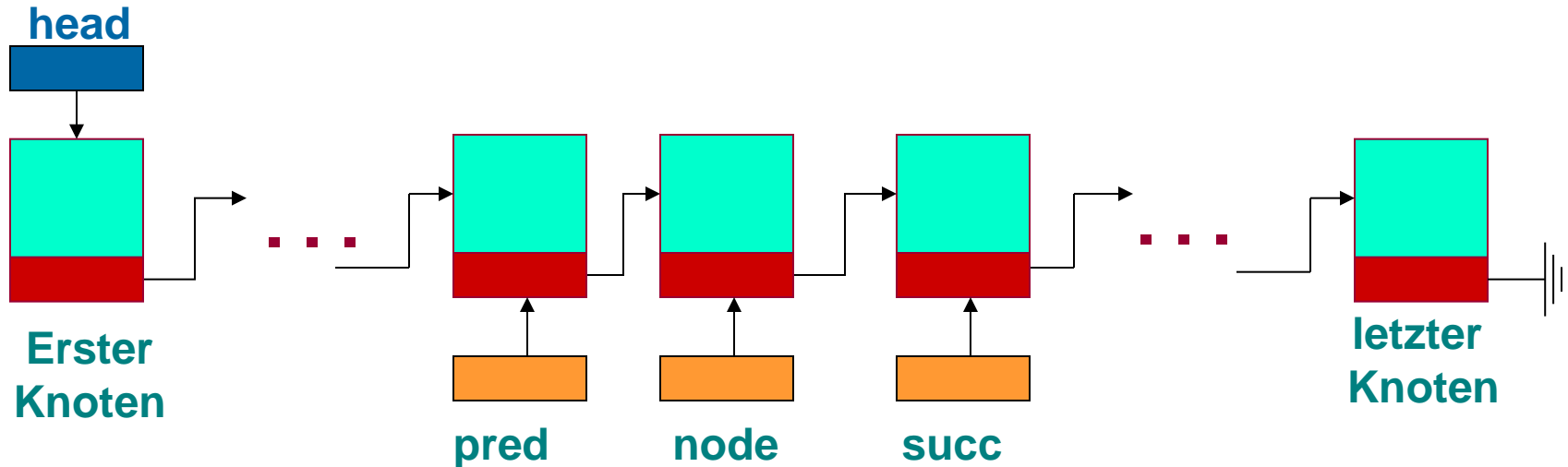


- Leerer Verweis am Listenende
  - Für letztes Element gilt:  
`node.next = null;`



- Alternative : Rückverweis auf Listenanfang ( $\Rightarrow$  Zirkuläre Liste)





Entsprechung für Durchlauf von Arrays:

```
for (int i=0; i < a.length; i++)
 ...
```

bei Listen:

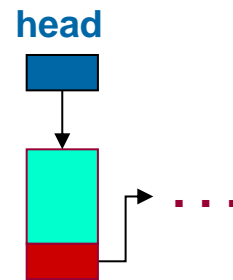
```
for (Node node=head; node != null; node = node.next)
 ...
```

- *void insert (int pos, Datenelement e)*
  - Füge Element an Position *pos* ein
- *void delete (int pos)*
  - Lösche den Knoten an Position *pos*

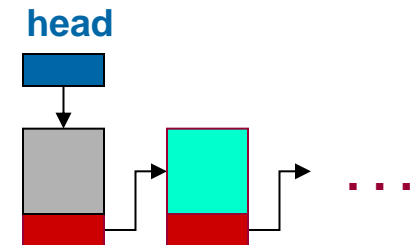
➔ Übung am Computer

# Varianten für Listenanfang

Listenanfang erfordert  
Sonderbehandlung



Alternative: ‚Pseudo-Knoten‘ am  
Listenanfang



- Komplexitätsanalyse der einfach verketteten Liste:

|                         |        |
|-------------------------|--------|
| Auslesen Anfang         | $O(1)$ |
| Auslesen Mitte          | $O(n)$ |
| Auslesen Ende           | $O(n)$ |
| Einfügen/Löschen Anfang | $O(1)$ |
| Einfügen/Löschen Mitte  | $O(n)$ |
| Einfügen/Löschen Ende   | $O(n)$ |

- Verweis auf Nachfolger und Vorgänger



```
class DNode {
 Object obj;
 DNode pred; //Vorgänger
 DNode succ; //Nachfolger
}
```

```
class DList {
 DNode first; //Anfang
 DNode last; //Ende
}
```

- Listenklasse merkt sich head und tail

- Komplexitätsanalyse der einfach und doppelt verketteten Liste:

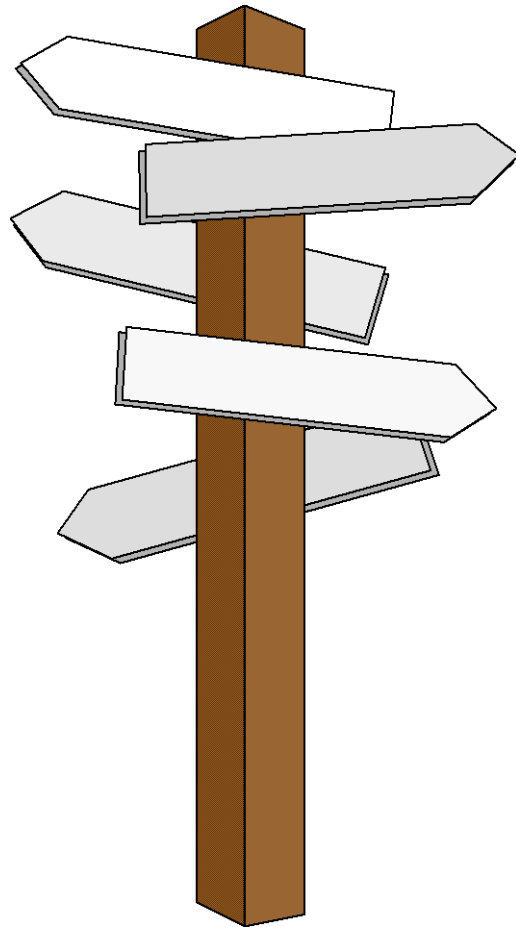
|                         | 1fach  | 2fach       |
|-------------------------|--------|-------------|
| Java-Klasse             | -      | Linked List |
| Auslesen Anfang         | $O(1)$ | $O(1)$      |
| Auslesen Mitte          | $O(n)$ | $O(n)$      |
| Auslesen Ende           | $O(n)$ | $O(1)$      |
| Einfügen/Löschen Anfang | $O(1)$ | $O(1)$      |
| Einfügen/Löschen Mitte  | $O(n)$ | $O(n)$      |
| Einfügen/Löschen Ende   | $O(n)$ | $O(1)$      |

## ArrayList - LinkedList

---

- ArrayList ist in den meisten Fällen schneller.
- ArrayList ist insbesondere schneller, wenn oft auf Elemente in der Mitte der Liste zugegriffen werden muss.
- LinkedList ist schneller, wenn oft Elemente am Anfang und am Ende der Liste eingefügt oder gelöscht werden.
  - **Wichtig für Queues oder Deques.**
  - **Hier kann auch ein zirkuläres dynamisches Feld verwendet werden.**
- Das Einfügen eines einzelnen Elements ist bei einer ArrayList im schlechtesten Fall  $O(n)$ .





## Vorlesung 5

### Binäre Suchbäume

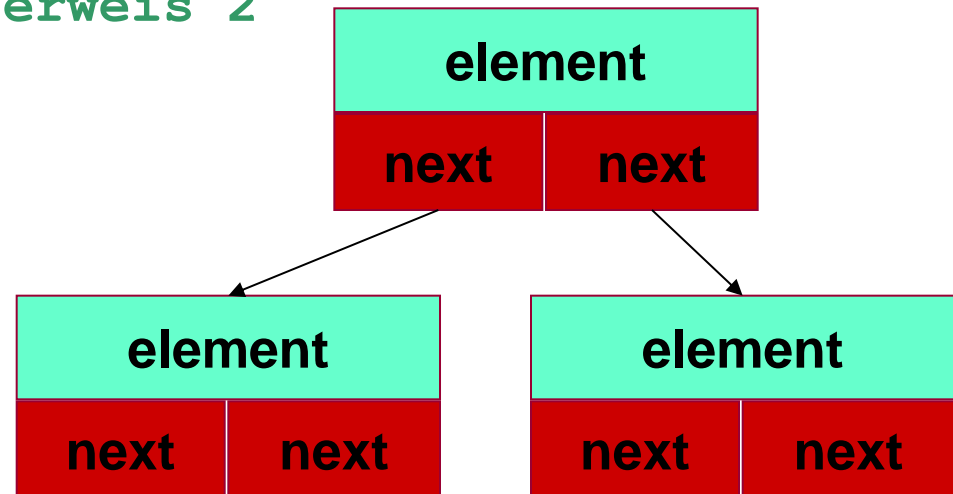
Zunächst betrachten wir die Definition von Bäumen

Anschließend untersuchen wir Bäume als Implementation von Mengen bzw. assoziativen Feldern.

- Alternative zu Hashtabellen (Vorlesungen 3/4)
- Vorteil: Daten sind in Bäumen sortiert.

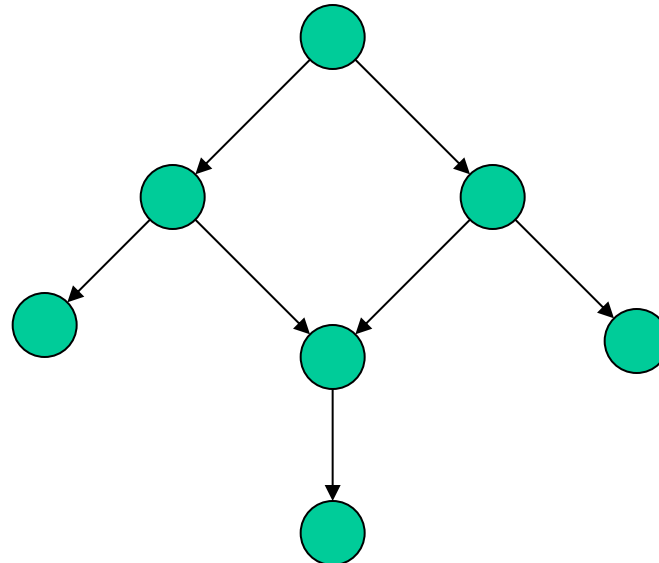
Gibt es in einem Knoten nicht einen, sondern mehrere Verweise, entstehen statt verketteten Listen **Bäume**.

```
class Node {
 Object element; // Datenkomponente
 Node nextLeft; // Verweis 1
 Node nextRight; // Verweis 2
}
```

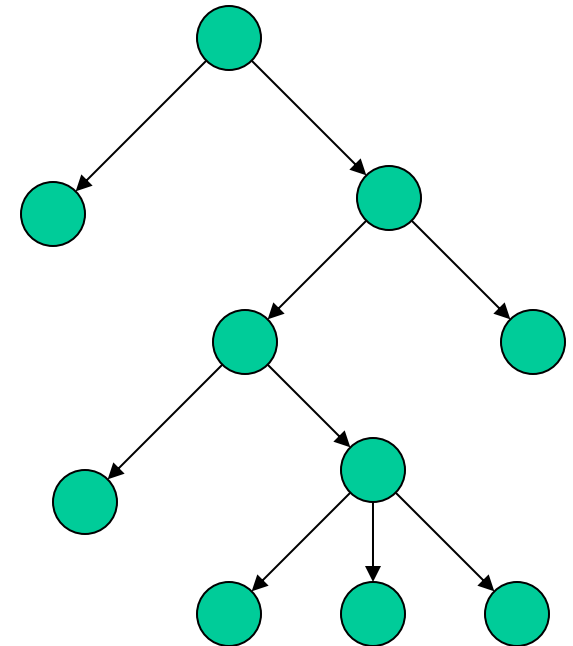


- **Bedingung für Bäume:**
  - **Zwei beliebige Knoten sind durch genau einen (einfachen) Pfad verbunden.**

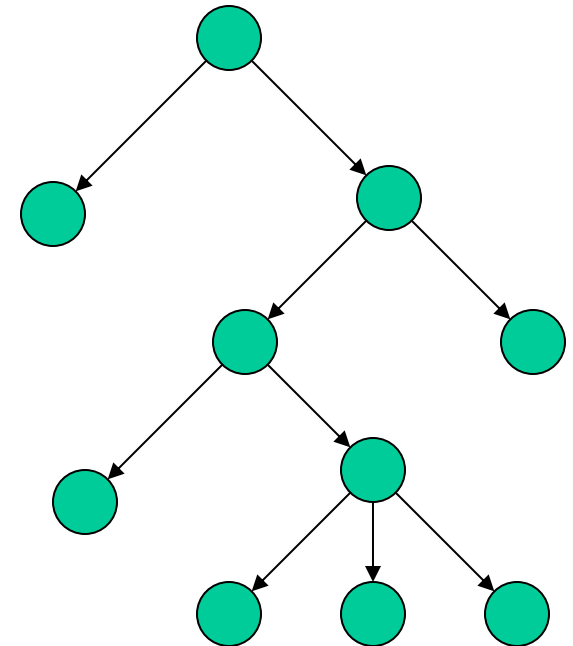
Graph, aber kein Baum:

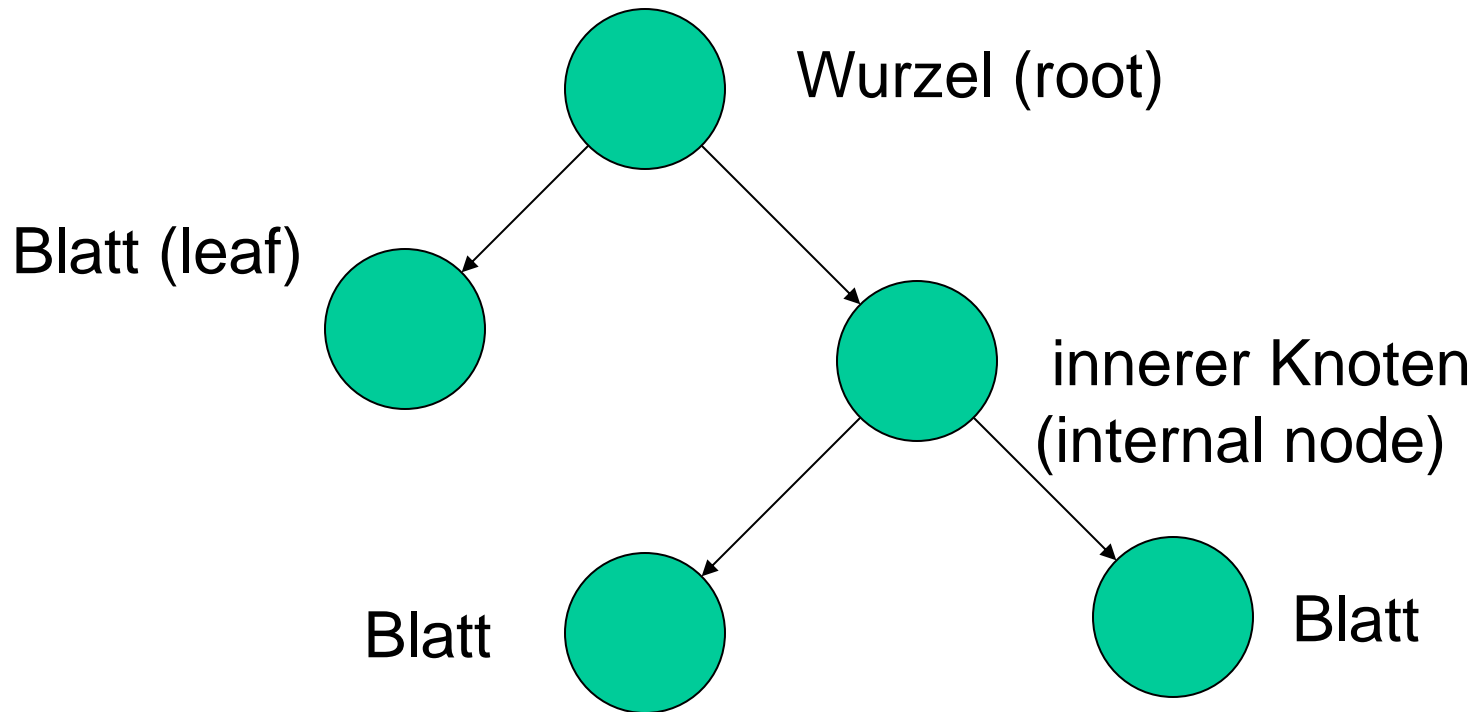


- Baum = Hierarchische (rekursive) Datenstruktur
  - alle Wege gehen von einer **Wurzel** aus
  - A heißt **Vorgänger** von B, wenn A auf einem Weg von der Wurzel zu B liegt. B heißt dann **Nachfolger** von A.
  - A heißt **Vater** von B, wenn  $(A, B) \in E$ , B heißt **Sohn** („Kind“) von A.
  - Haben B und C denselben Vater, so heißen sie **Brüder**.
  - Knoten ohne Söhne heißen **Blätter** („Terminalknoten“).
  - Knoten mit Söhnen heißen **innere Knoten**.

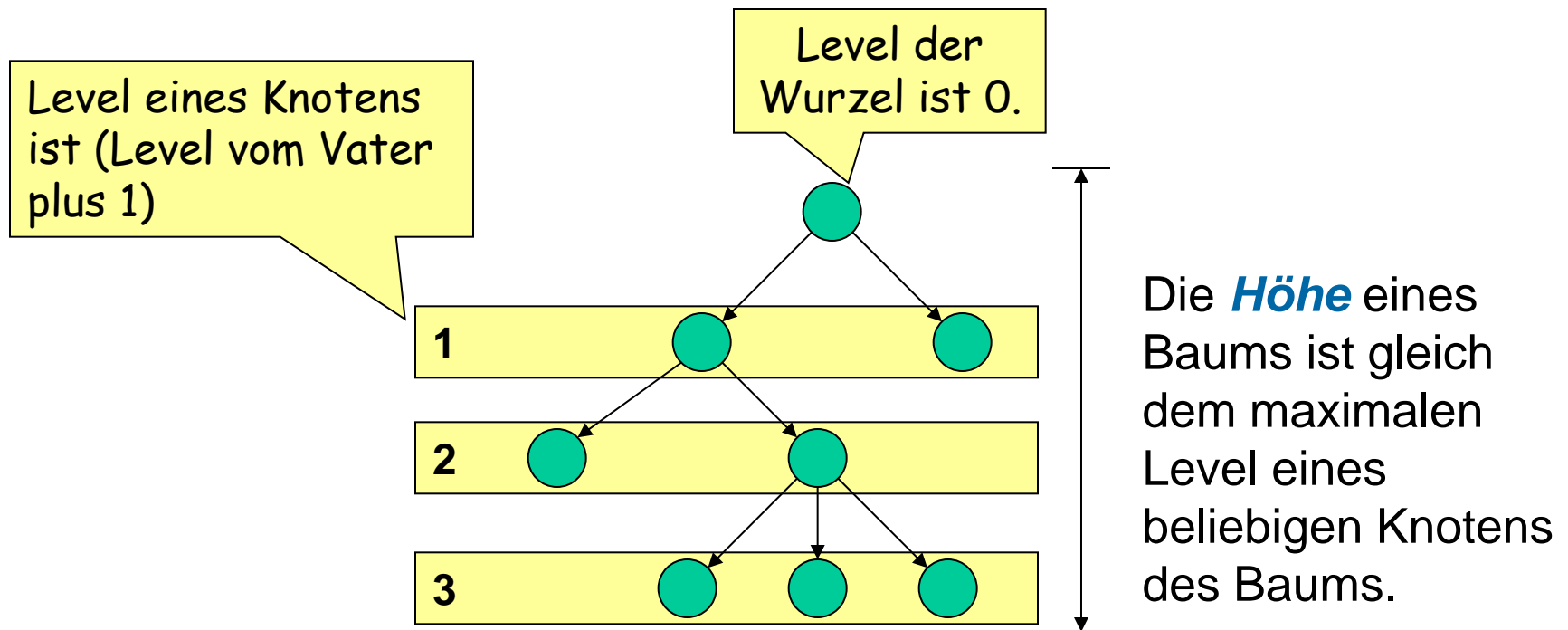


- Ein Knoten  $S$  mit allen Nachfolgern wird **Teilbaum** eines Baumes  $T$  genannt, falls  $S$  nicht die Wurzel von  $T$  ist.
- Der **Verzweigungsgrad** eines Knotens ist die Anzahl seiner Kinder





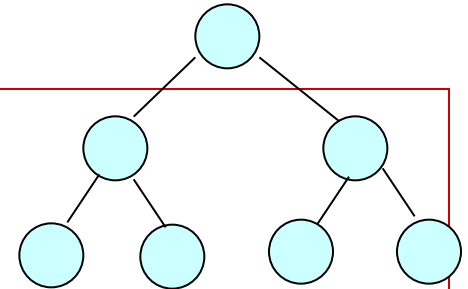
Jeder Knoten in einem Baum liegt auf einem bestimmten **Level** (Länge des Pfades von der Wurzel zu diesem Knoten).





## Definition:

Die Knoten eines **Binärbaums** („binary tree“) haben höchstens den Verzweigungsgrad 2 (haben höchstens 2 Söhne).



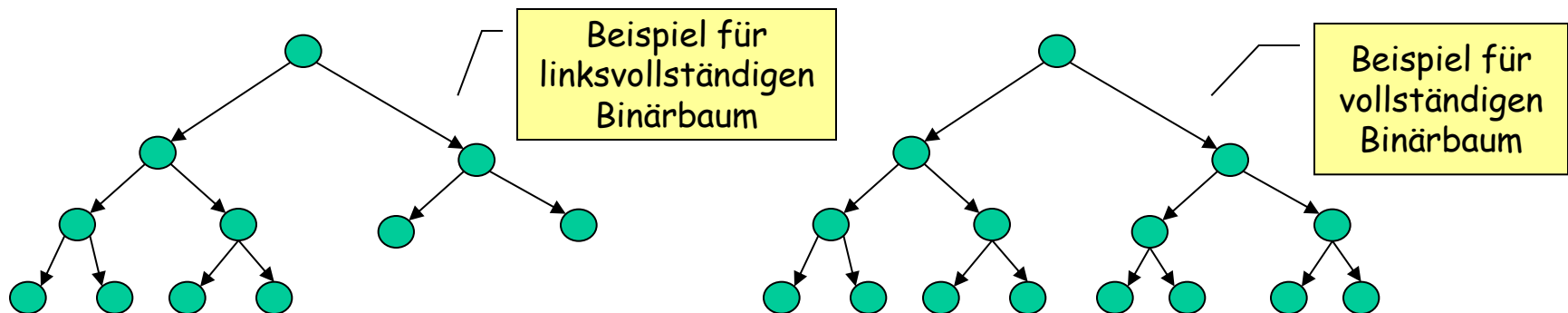
Bei einem **geordneten Binärbaum** ist die Reihenfolge der Söhne durch die Indizes eindeutig festgelegt ( $T_1$  = linker Sohn, linker Teilbaum;  $T_2$  = rechter Sohn, rechter Teilbaum).

### Definition:

Ein Binärbaum heißt *minimal* (bezogen auf die Höhe), wenn kein Binärbaum mit gleicher Knotenzahl aber kleinerer Höhe existiert.

Ein *links-vollständiger Binärbaum* ist ein minimaler Binärbaum, in dem die Knoten auf dem untersten Level so weit wie möglich links stehen.

Alle Blätter eines *vollständigen* Binärbaums haben den gleichen Level.



### Einführung:

<http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>

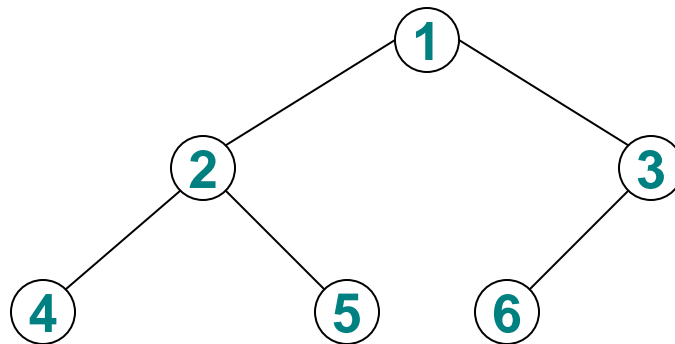
Kapitel Suchbaum (1)

### **binärer Suchbaum:**

Ein binärer Suchbaum ist ein Binärbaum, bei dem für jeden Knoten des Baumes gilt:

- Alle Schlüssel im linken Teilbaum sind kleiner, alle im rechten Teilbaum sind größer oder gleich dem Schlüssel in diesem Knoten.

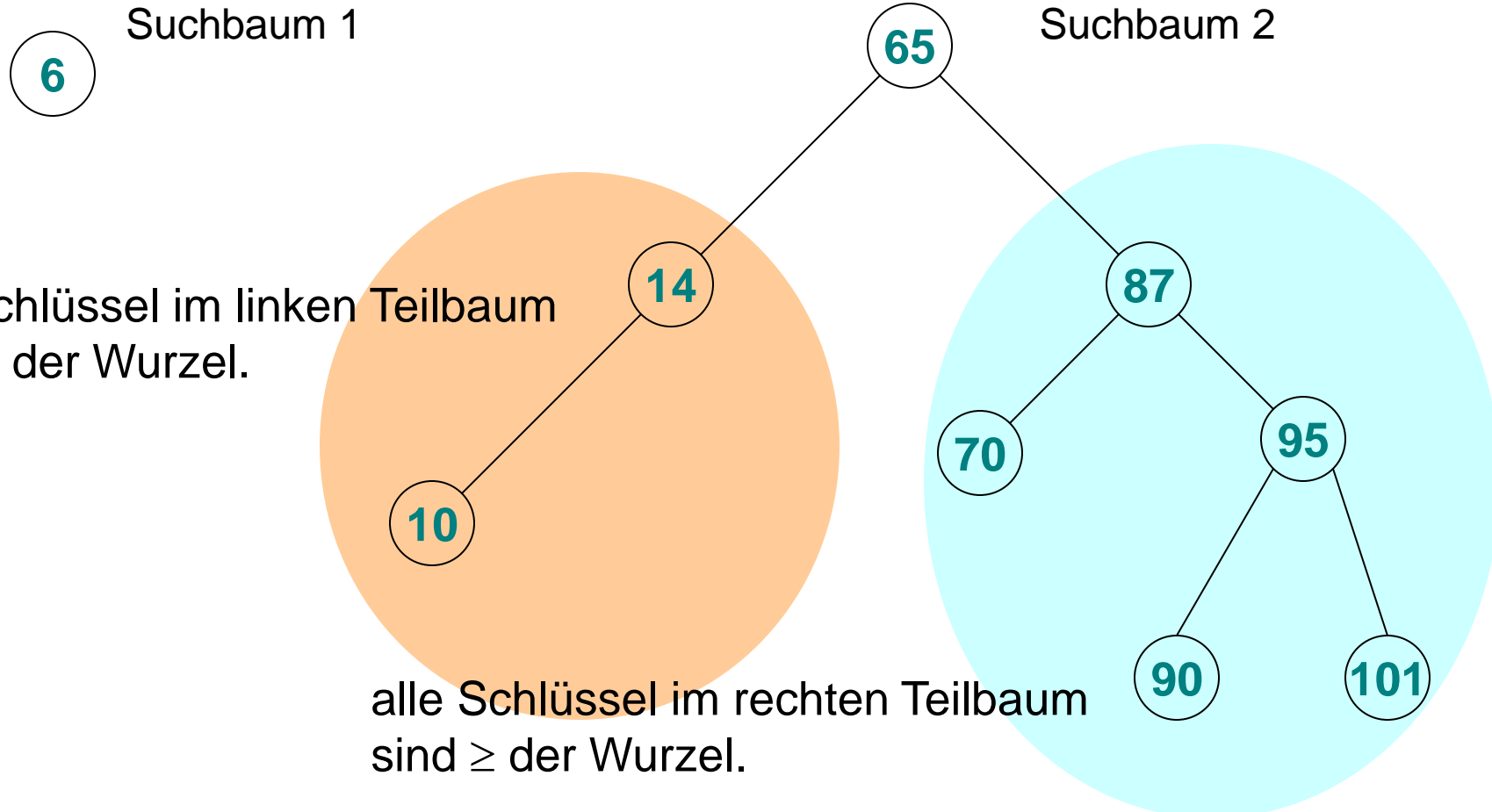
Eignet sich am besten zur Darstellung links-vollständiger Binärbäume, z.B.:



Man nummeriert die Knoten ebenenweise durch. Das Feldelement 0 bleibt frei. Dann berechnen sich die Positionen von Vorgängern und Nachfolgern so:

$$\text{Vorgänger : Pred}(n) = \lfloor n/2 \rfloor$$

$$\text{Nachfolger : Succ}(n) = \begin{cases} 2n & \text{linker Sohn} \\ 2n+1 & \text{rechter Sohn} \end{cases} \quad (\text{falls diese existieren})$$



### Einführung:

<http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>

Kapitel Suchbaum (2)

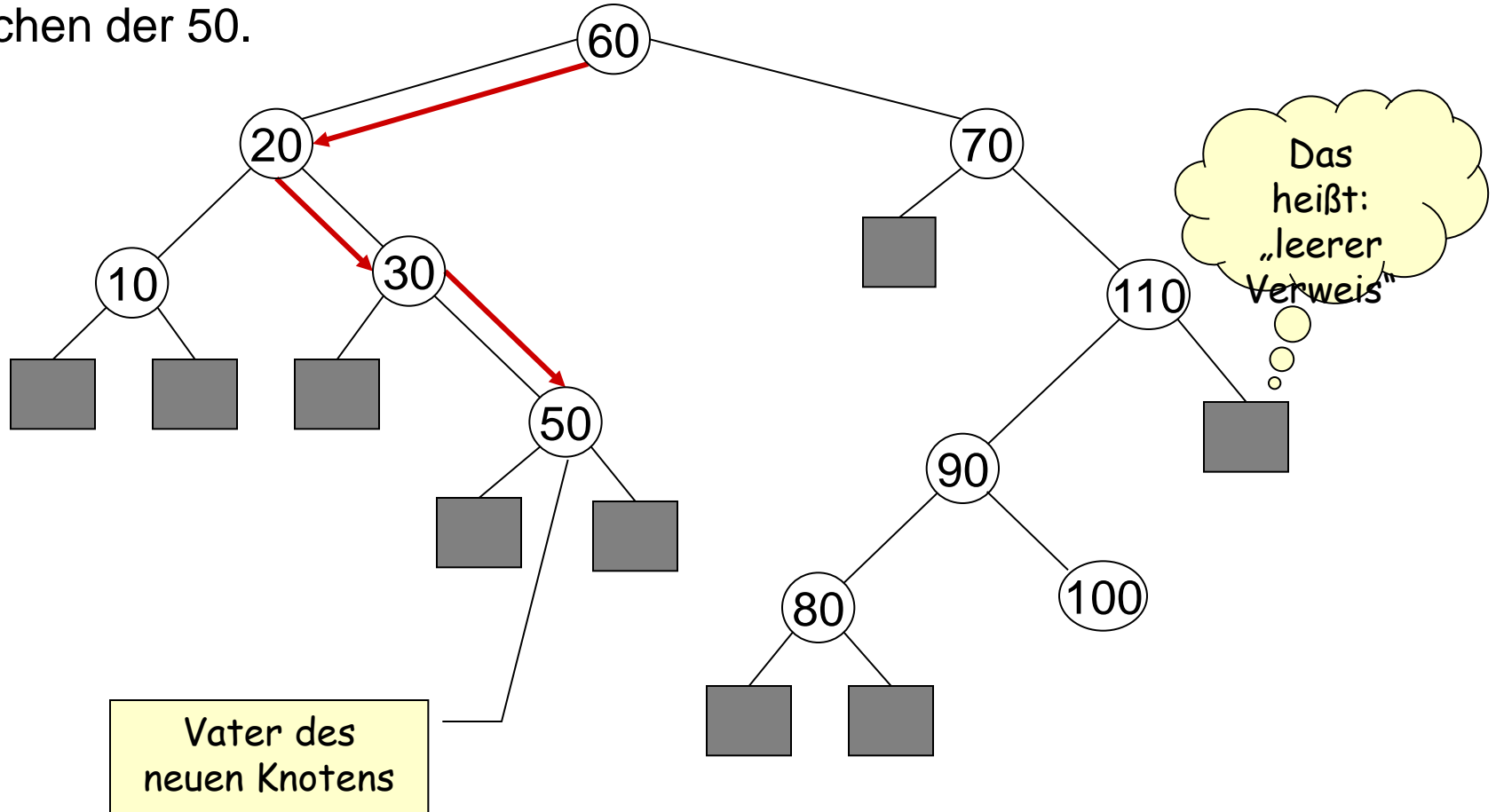
### binärer Suchbaum:

Ein binärer Suchbaum ist ein Binärbaum, bei dem für jeden Knoten des Baumes gilt:

- Alle Schlüssel im linken Teilbaum sind kleiner, alle im rechten Teilbaum sind größer oder gleich dem Schlüssel in diesem Knoten.

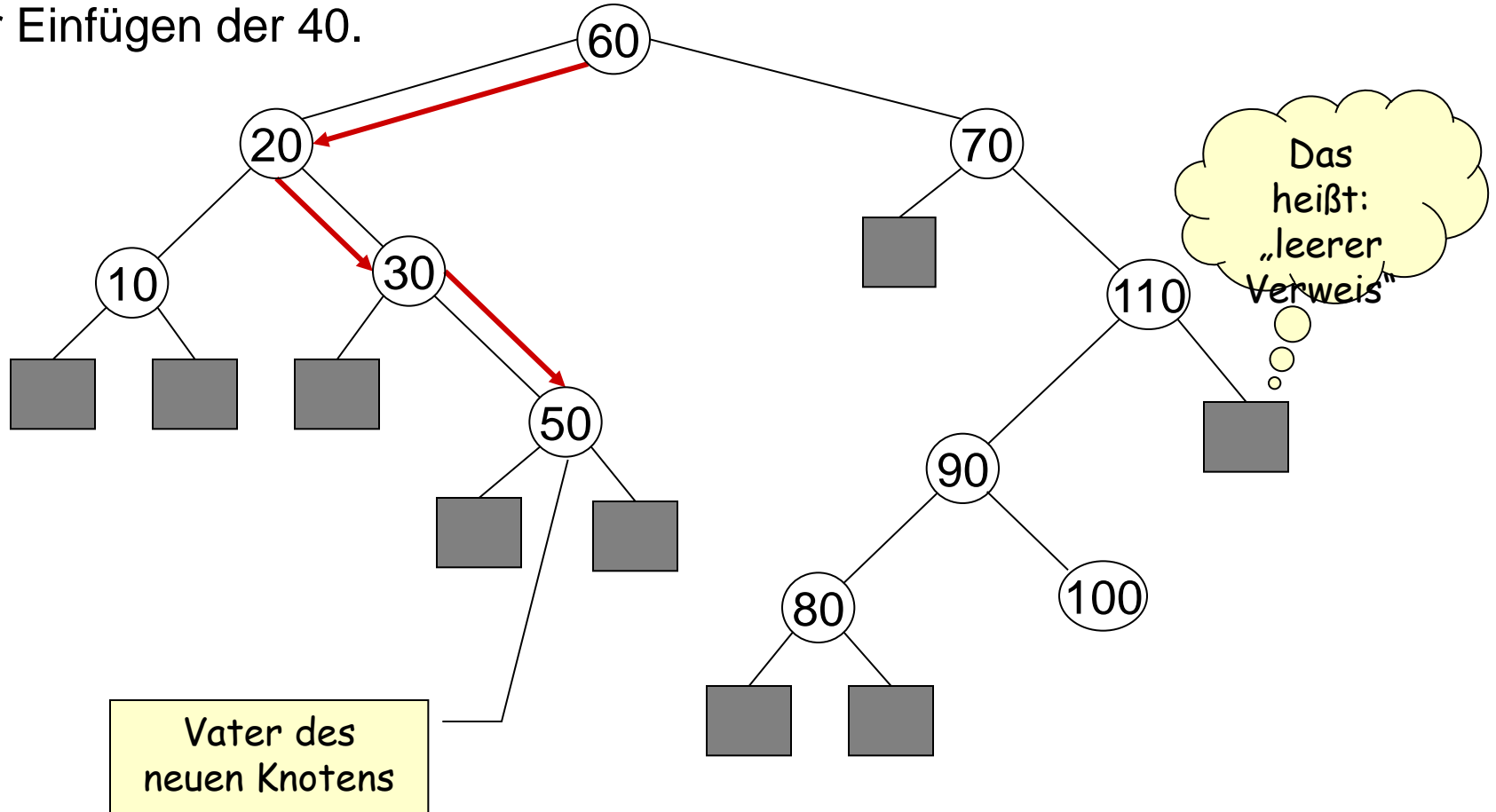
# Suchen im Suchbaum

Suchen der 50.



# Einfügen im Suchbaum (1)

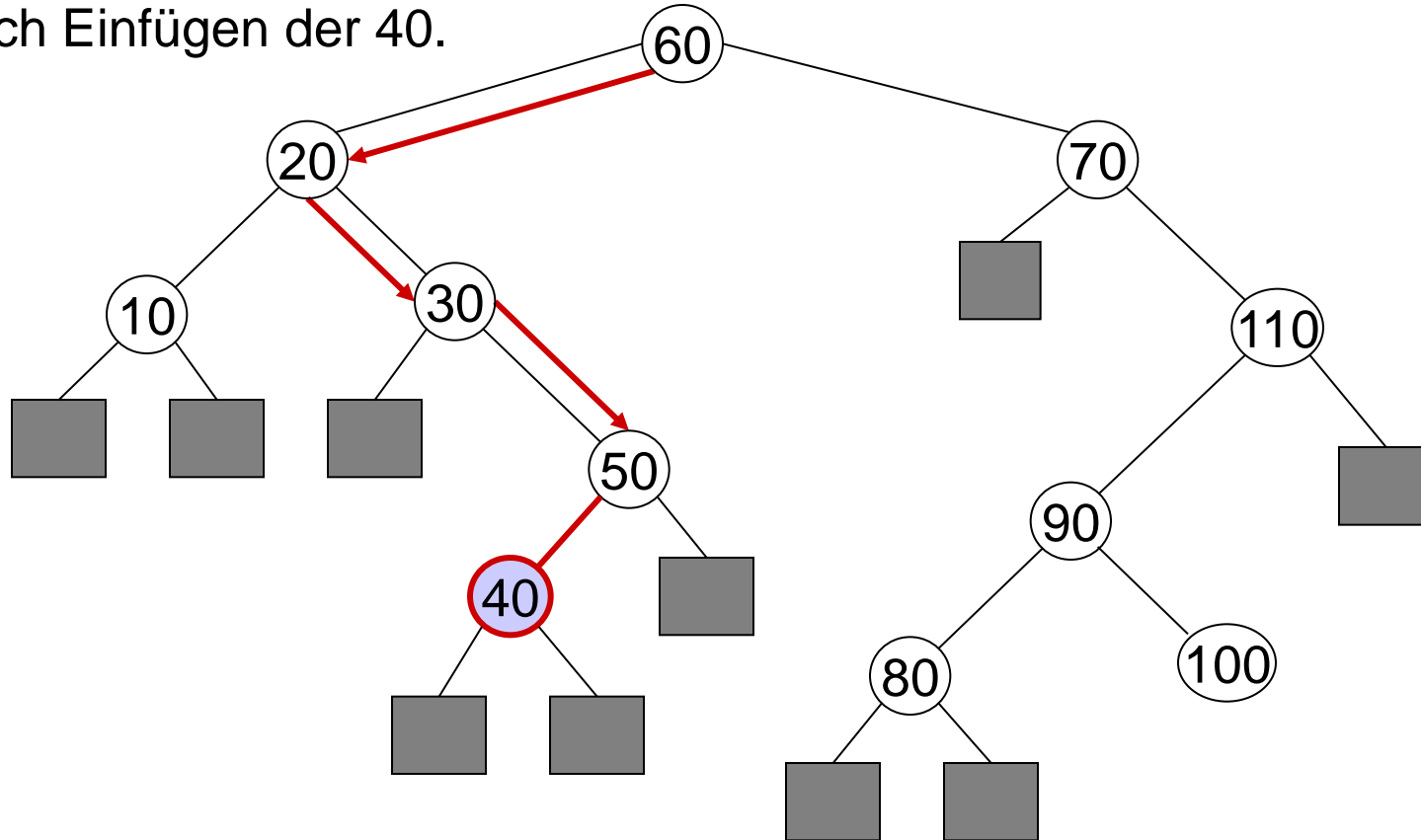
Vor Einfügen der 40.



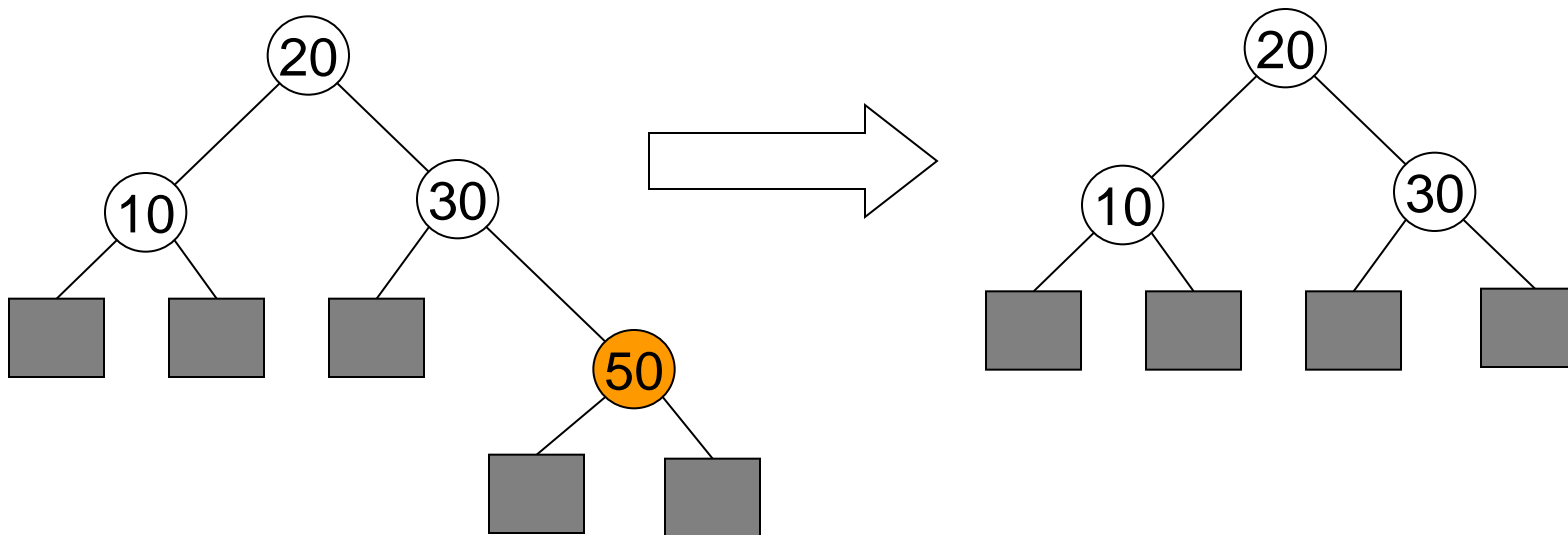


## Einfügen im Suchbaum (2)

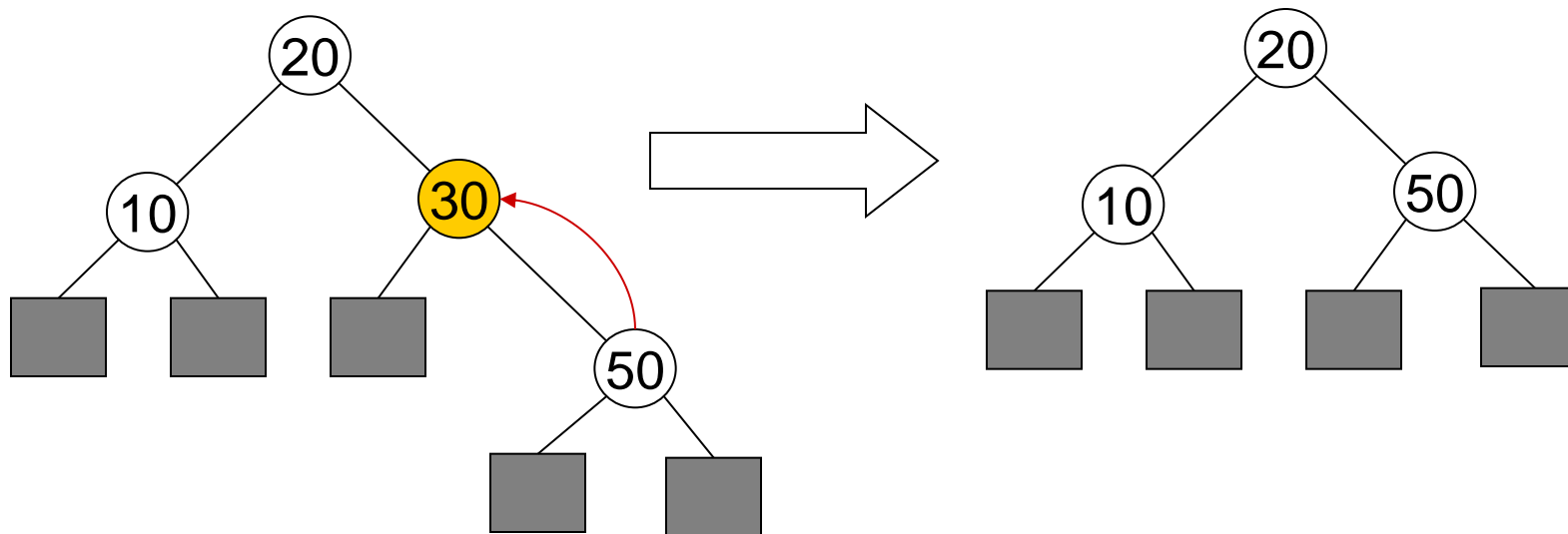
Nach Einfügen der 40.



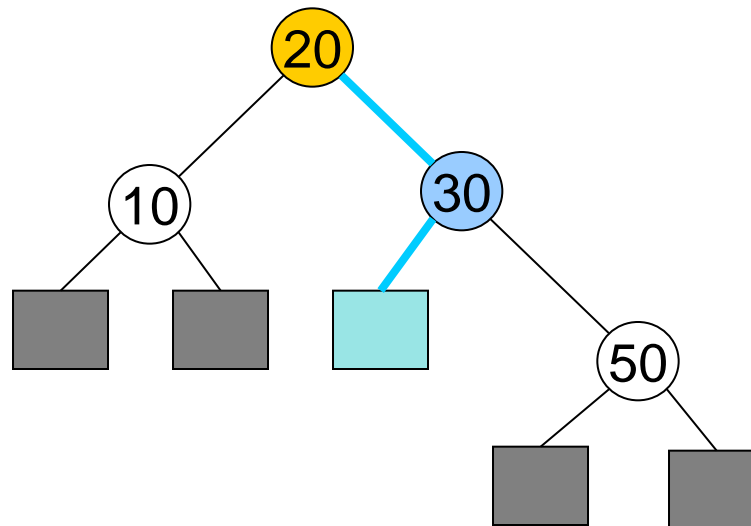
- **Fall 1: zu löschender Knoten ist Blatt (hat keine Kinder)**
- Knoten kann problemlos gelöscht werden
- Beispiel: Löschen der 50



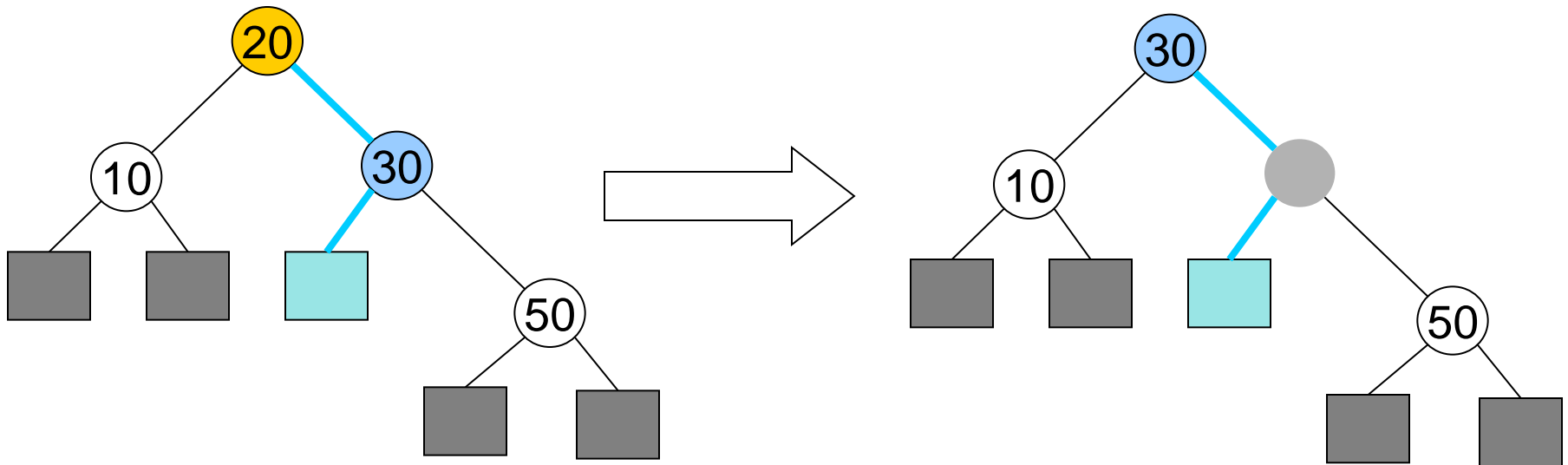
- **Fall 2: zu löschender Knoten hat ein Kind**
- Kind-Knoten rückt an die Stelle des Knotens vor
- Beispiel: Löschen der 30



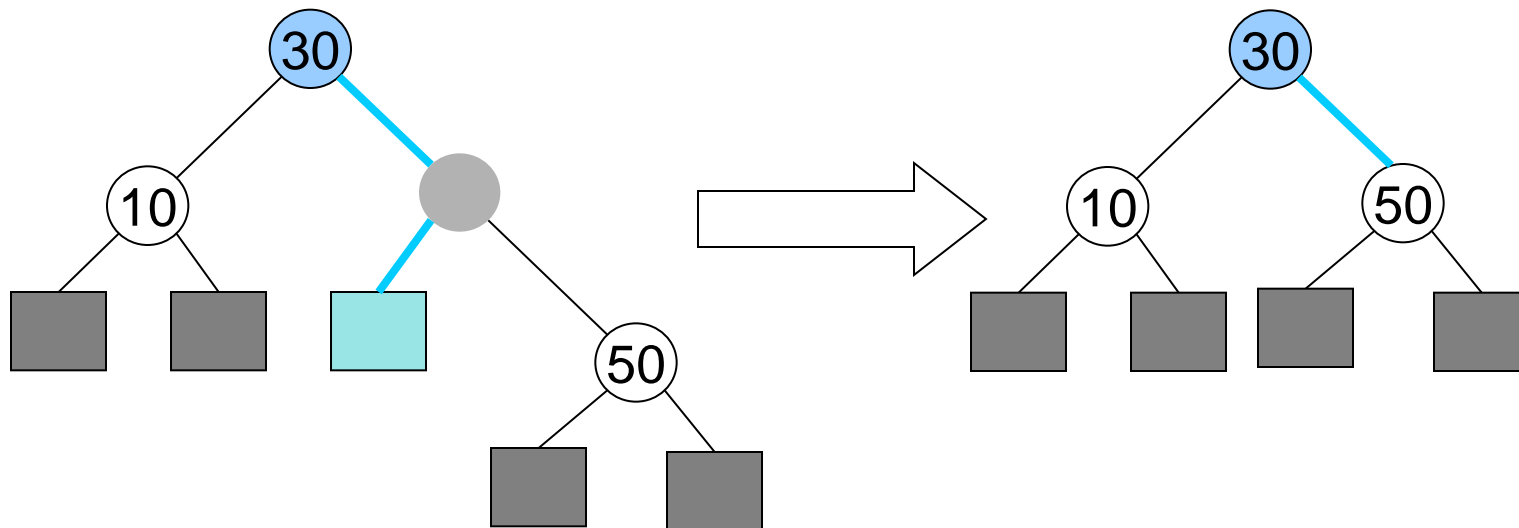
- **Fall 3: zu löschender Knoten hat zwei Kinder**
- Suchen des nächstgrößeren Knoten im rechten Baum
- Erst einen Schritt nach rechts
- Dann solange nach links, bis es links nicht weitergeht.
- Beispiel: Löschen der 20 → nächstgrößere Knoten: 30



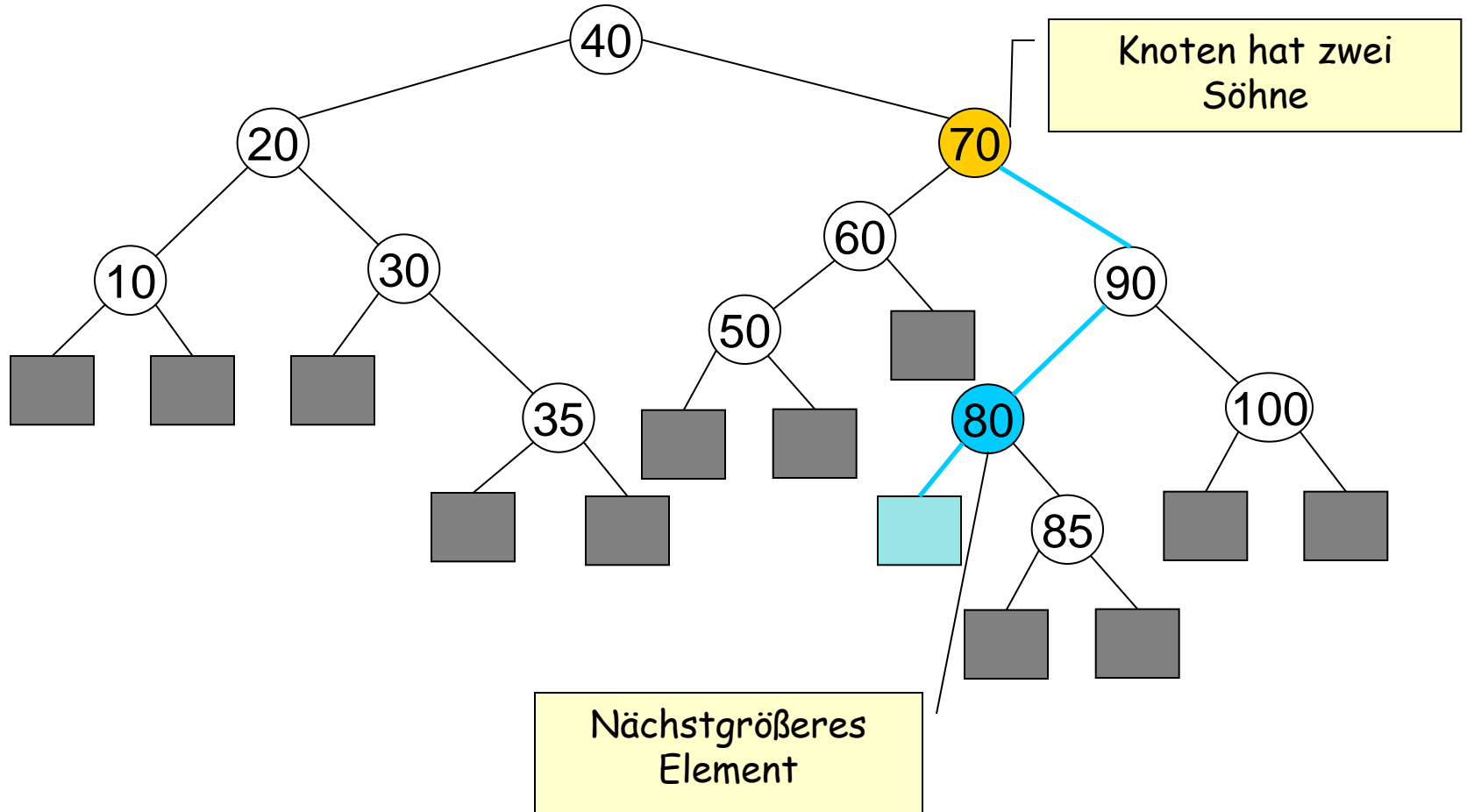
- **Fortsetzung Fall 3**
- Ersetze den zu löschenden Knoten durch den nächstgrößeren.



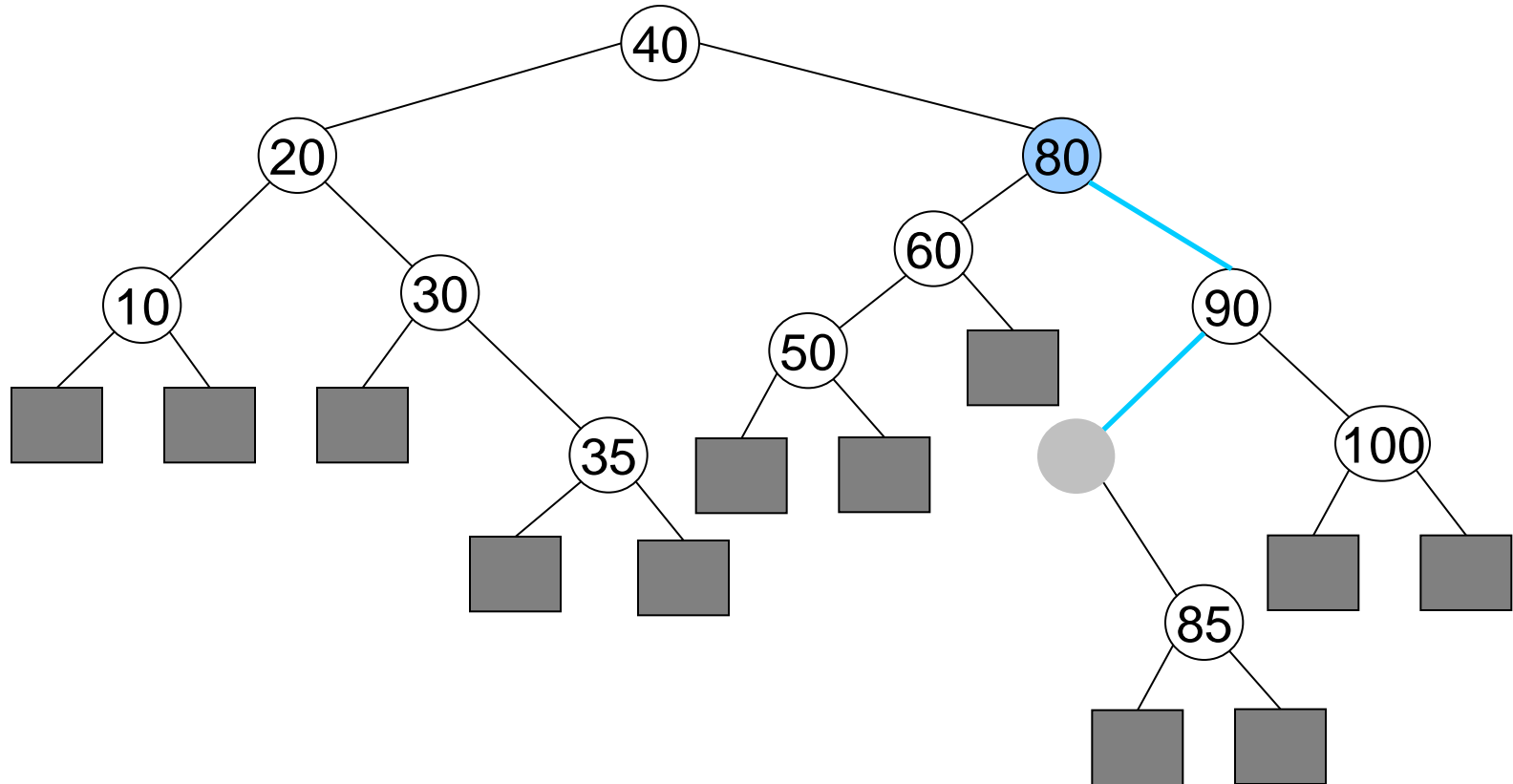
- **Fortsetzung Fall 3**
- Lasse rechte Seite des freien Platzes vorrücken.



# Löschen der „70“ (1)

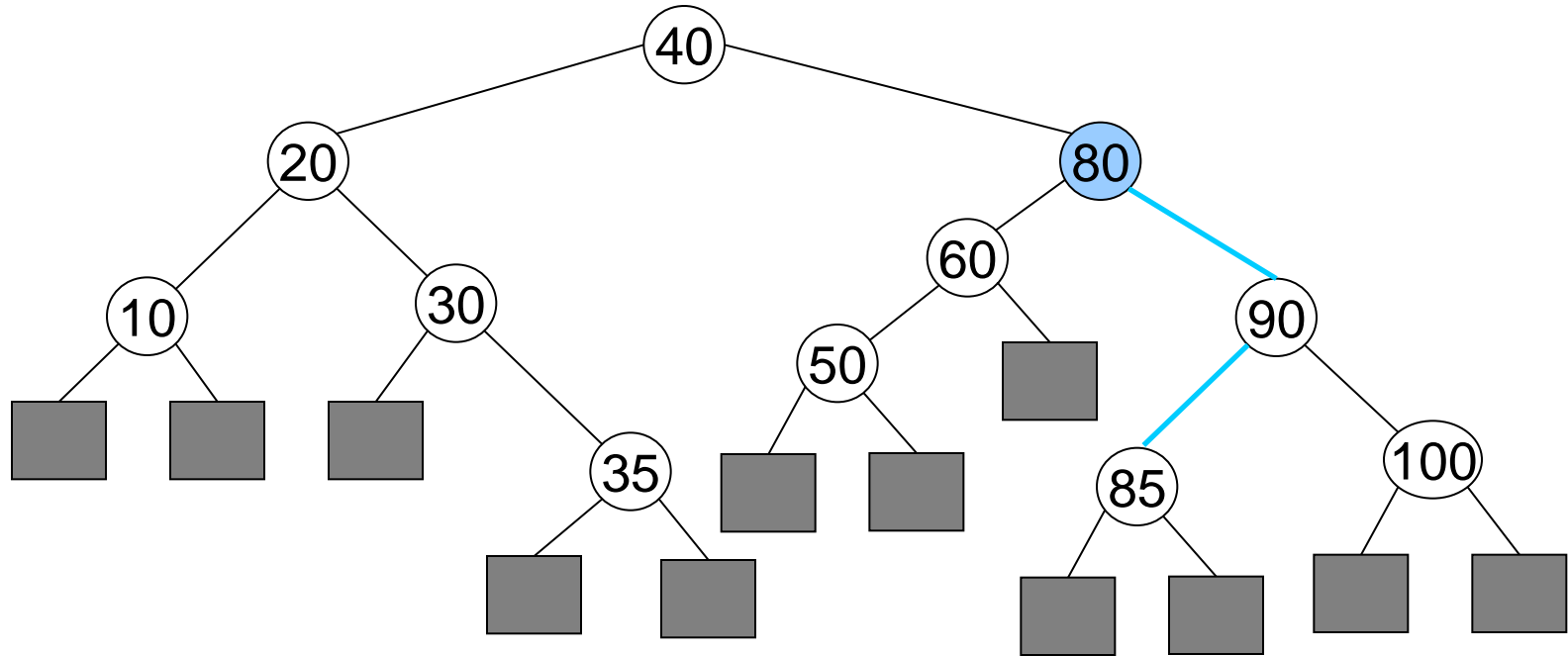


# Löschen der „70“ (2)





# Löschen der „70“ (3)

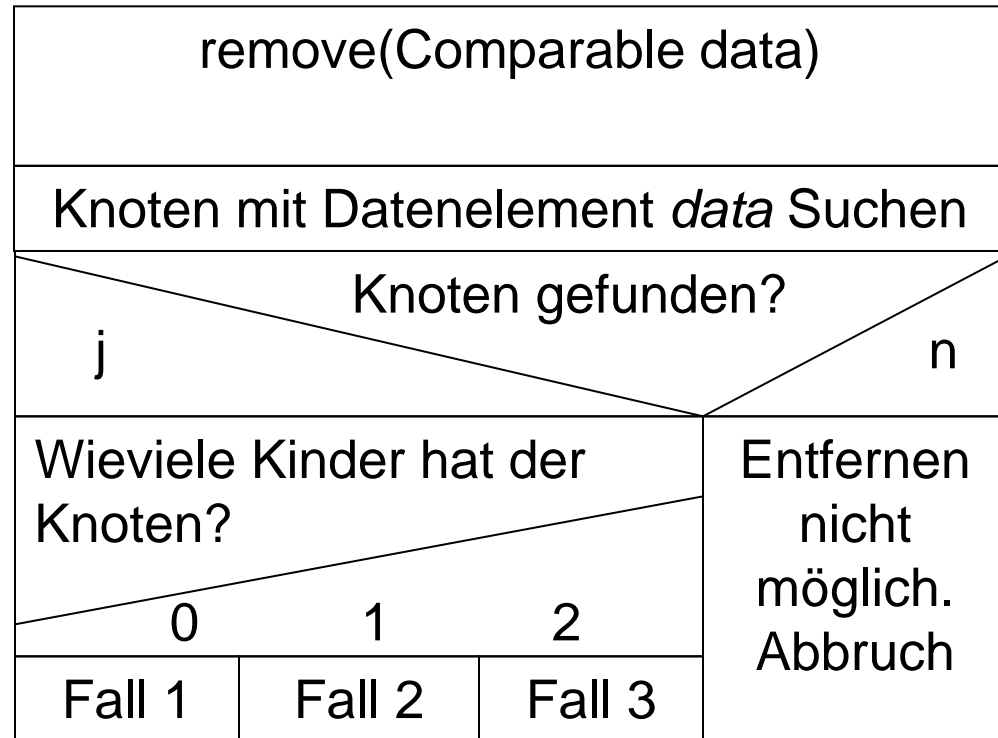


<http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>

Kapitel Suchbaum (3)

Datenstruktur:

```
class Node {
 Node parent;
 Node right;
 Node left;
 Comparable data;
}
```



## Unterrouninen zum Löschen (1)

Datenstruktur:

```
class Node {
 Node parent;
 Node right;
 Node left;
 Comparable data;
}
```

CopyNode: Kopiert Daten und Referenzen zu den Nachfolgern von Node n1 in Node n2

```
public void copyNode
 (Node n1, Node n2) {
 n2.data = n1.data;
 n2.left = n1.left;
 n2.right = n1.right;
}
```

copyData:

Kopiert *data* von Node n1 in Node n2

```
public void copyData(Node n1, Node n2) {
 n2.data = n1.data;
}
```

clearNode(Node n):

Löscht n aus dem Baum. Darf nur aufgerufen werden, wenn n ein Blatt ist.

```
public void clearNode(Node n) {
 if (n==root) {
 root = null;
 return;
 }
 if (n.parent.left == n) {
 n.parent.left = null;
 } else {
 n.parent.right = null;
 }
}
```

## Löschen: Fall 1 und 2

- Zu löschender Node: Node n
- **Fall 1: keine Kinder**

```
clearNode (n)
```

- **Fall 2: ein Kind**

```
if (n.left==null) {
 copyNode (n.right, n);
} else {
 copyNode (n.left, n);
}
```

- Fall 3: zwei Kinder

```
//minimales Element auf der rechten Seite suchen
```

```
Node minR = n.right;
```

```
while (minR.left != null) {
```

```
 minR = minR.left;
```

```
}
```

```
//n durch minR ersetzen
```

```
copyData(minR, n);
```

```
//rechte Seite von minR nach minR schieben
```

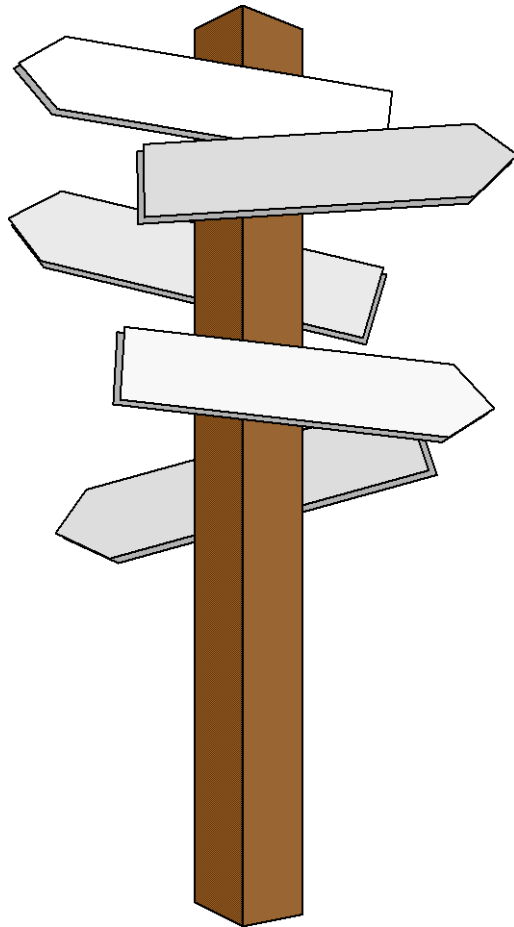
```
if (minR.right!=null) {
```

```
 copyNode(minR.right, minR);
```

```
} else {
```

```
 clearNode(minR);
```

```
}
```

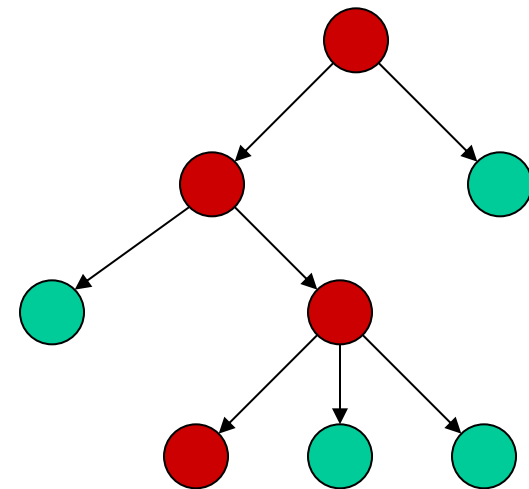


## Vorlesung 6

## AVL-Bäume



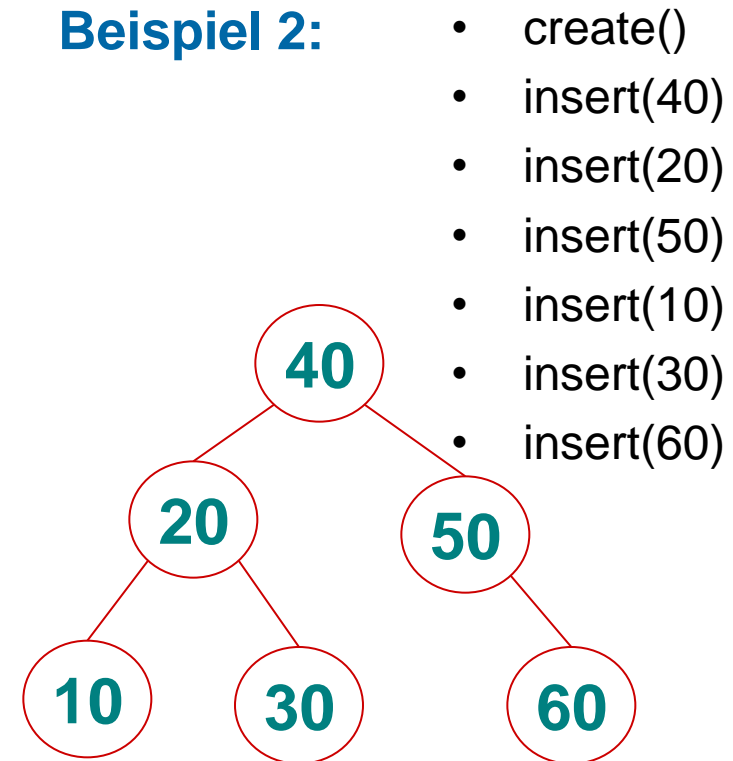
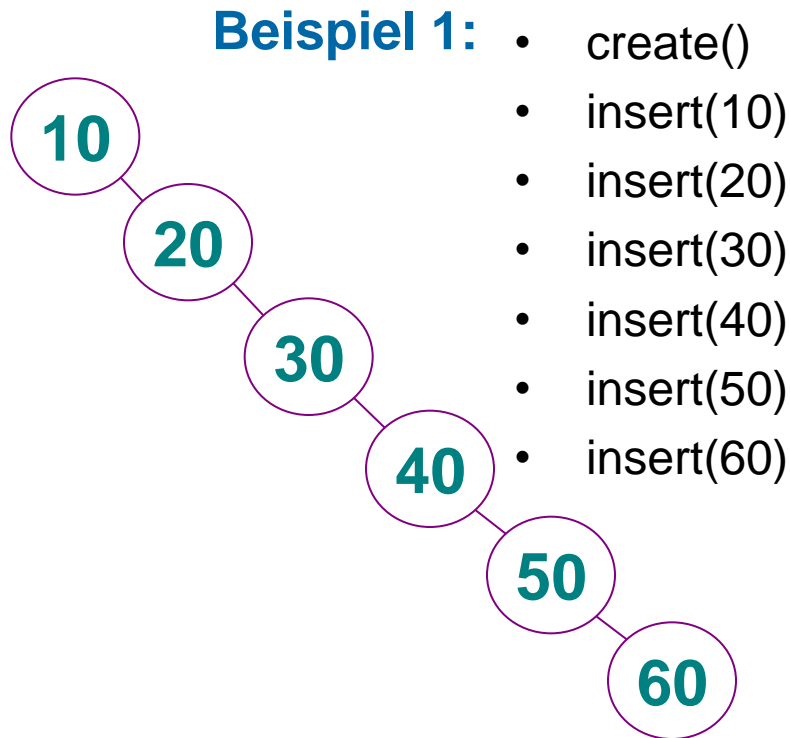
- Die Komplexität der Funktionen **Suchen**, **Löschen** und **Einfügen** werden durch die Komplexität des Suchens eines Elements bestimmt
- Im schlechtesten Fall ist die Anzahl der zu durchsuchenden Elemente gleich der Höhe des Baums+1.
- Die Höhe hängt stark von der Reihenfolge der Einfüge-Operationen ab.



- Einführendes Applet:

<http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>

Kapitel AVL (1)



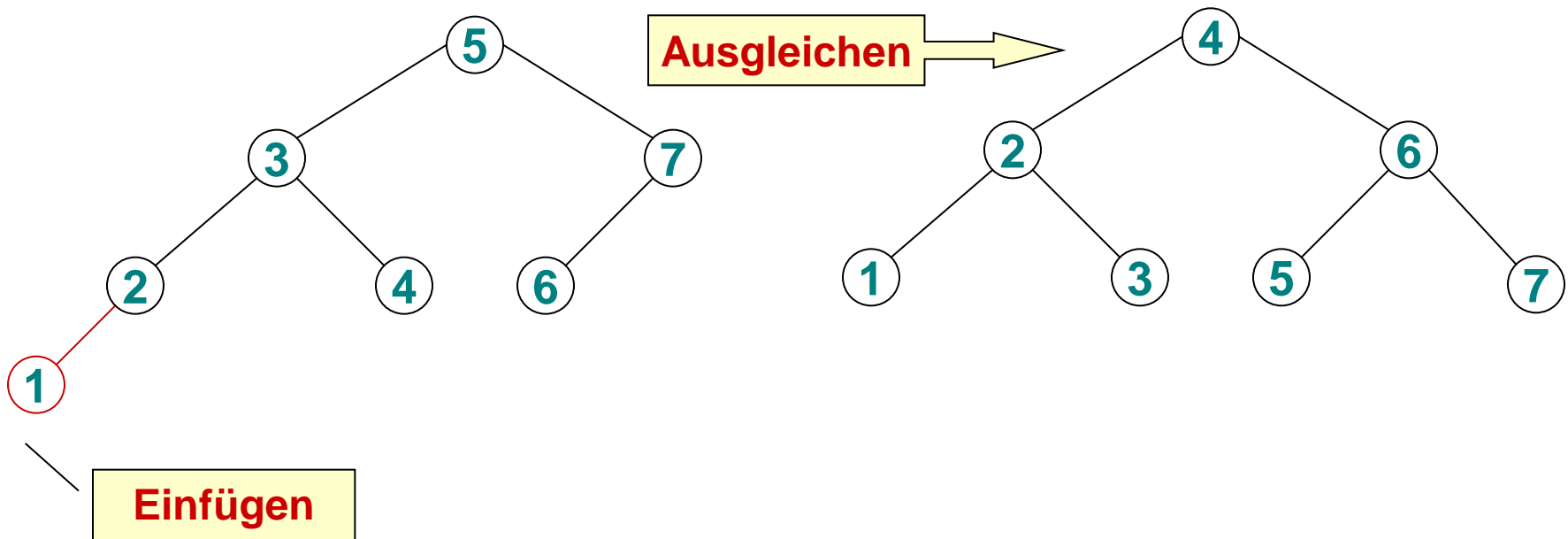
- **Balanciertheit** (Ausgeglichenheit) der Knoten-Verteilung nicht garantiert
- bei ungünstiger Reihenfolge der Einfügeoperationen können zu linearen Listen **entartete Bäume** entstehen

- Ein vollständig gefüllter Binärbaum (voller Baum, in dem die letzte Ebene voll besetzt ist) der **Höhe H** hat
  - $n = 1+2+4+\dots+2^H = (2^{H+1}-1)/(2-1)$  **Knoten**, (geometrische Reihe)
- Ein Binärbaum mit n Knoten hat im **besten Fall** (optimal balanciert) die Höhe  $\lceil \log_2(n+1) \rceil - 1 = \lfloor \log_2(n) \rfloor$ 
  - ⇒ **Suchen ist  $O(\log n)$**
- Ein Binärbaum mit n Knoten hat im **schlechtesten Fall** („entarteter“/„degenerierter“ Baum) die Höhe n-1.
  - ⇒ **Suchen ist  $O(n)$**

## Wie kann Entartung verhindert werden?

- Eine Möglichkeit: Baum nach jedem insert/remove durch Umstrukturierung ausgleichen.

⇒ kann aufwändig sein, z.B.:



(mindestens) 3 Lösungsideen (Balance-Kriterien):

1. abgeschwächtes Kriterium für ausgeglichene Höhe  
⇒ **lokale Umordnungsoperationen reichen aus.**
- **Verschiedene Varianten, die jeweils unterschiedliche Kriterien haben:**
  - **AVL-Bäume:** Werden ausführlich behandelt.
  - **Rot-Schwarz-Bäume:** Werden in der Java-Klassenbibliothek benutzt. Ähnlichkeiten mit B-Bäumen (siehe 3. Lösungsidee). Kurze Vorstellung im Anschluss an B-Bäume.

2. Jeder neue Knoten wandert an die Wurzel des Baums.
  - **Vorteil: Zuletzt eingefügte Elemente lassen sich schneller finden.**
  - **Durch ein spezielles Einfügeverfahren wird der Baum zusätzlich (teilweise) ausgeglichen.**
- **Splay-Bäume:** Werden in der Vorlesung nicht weiter behandelt.

3. Unausgeglichener Verzweigungsgrad ermöglicht ausgeglichene Höhe.
  - **B-Bäume:**
    - B-Bäume besitzen **ausgeglichene Höhe**, lassen aber **unausgeglichene Verzweigungsgrad** zu.
    - Varianten von B-Bäumen werden speziell in Datenbanksystemen als Indexstrukturen eingesetzt.
    - Werden ausführlich behandelt.



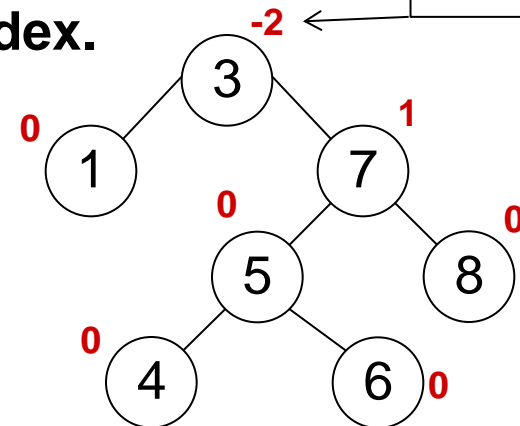
- Bei einem AVL-Baum unterscheiden sich die Höhen zweier Teilbäume des gleichen Knotens maximal um 1.

- Der sogenannte **Balance-Index** ist die Differenz

Höhe(rechter Teilbaum) – Höhe(linker Teilbaum)

- Jeder Knoten hat einen Balance-Index.

- Er darf nur die Werte **-1, 0 oder 1** annehmen.



Dieser Baum ist kein AVL-Baum

- <http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>

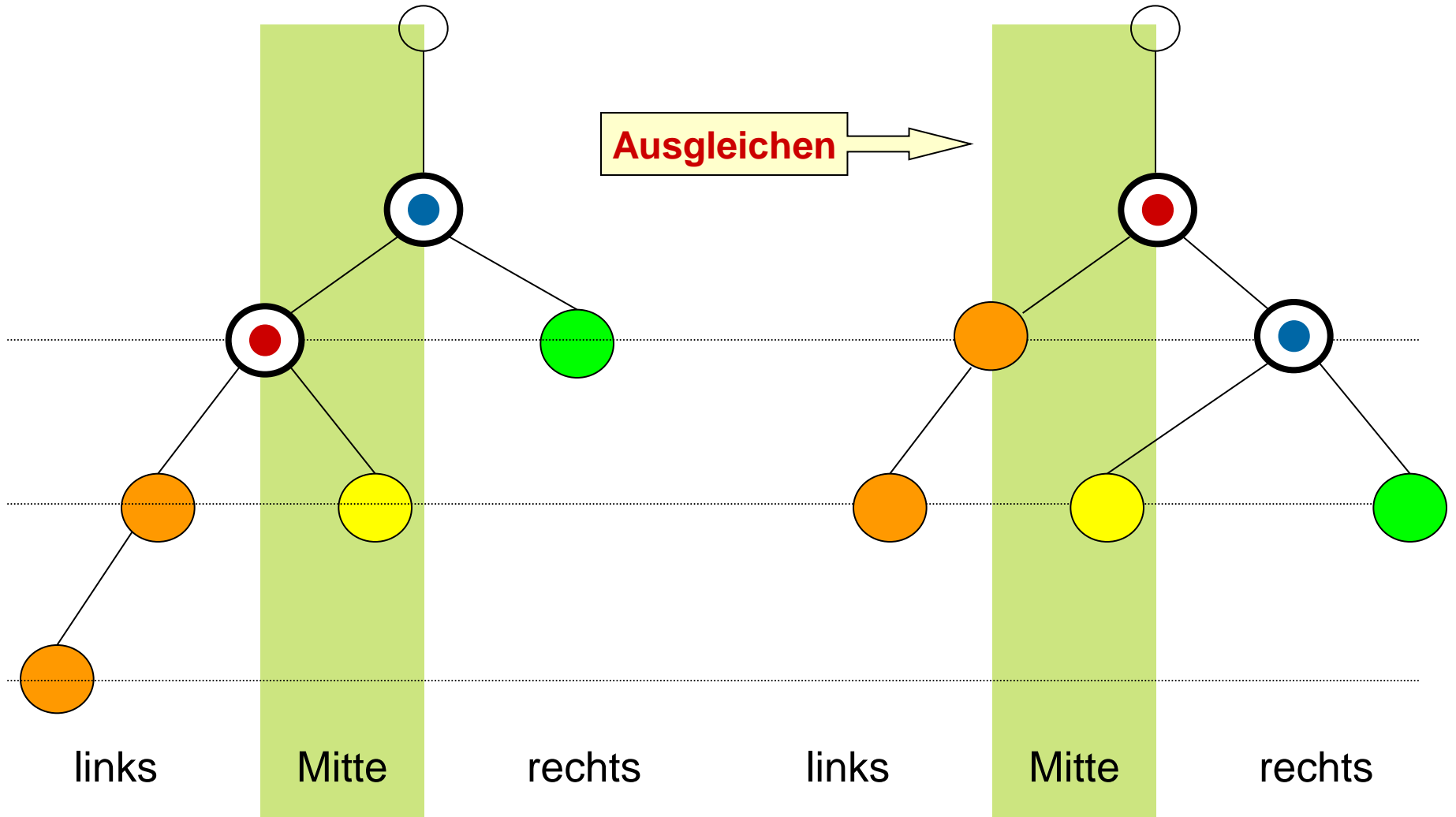
Kapitel AVL (2)

Wenn durch eine Einfüge- oder Lösche-Operation die AVL-Bedingung verletzt wird, muss (mit lokalen Operationen) **rebalanciert** (ausgeglichen) werden.

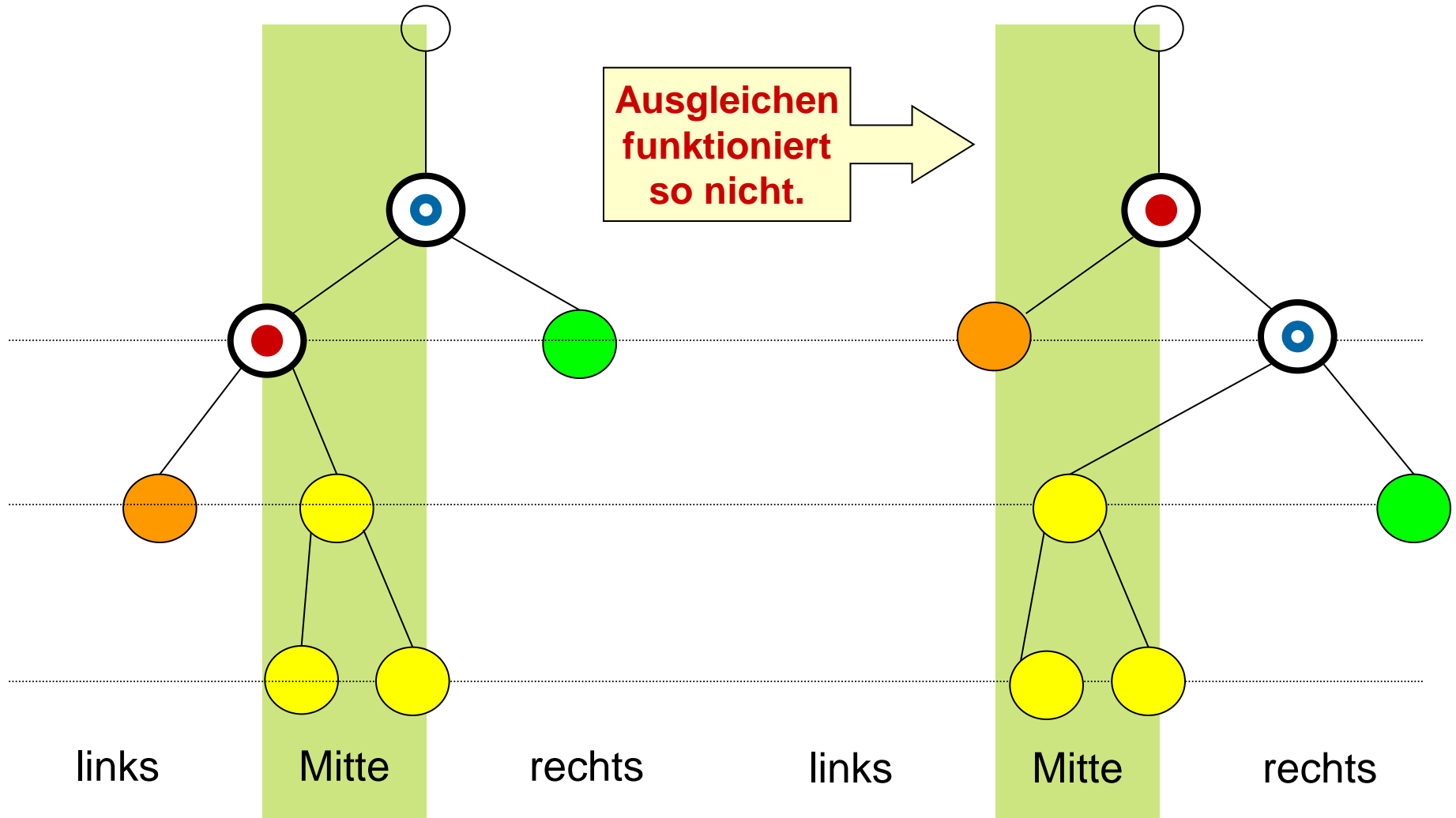
Je nach Situation wendet man dazu entweder eine **Rotation** oder eine **Doppelrotation** an.

- **Wie die Rotationen aussehen, wird auf den folgenden Folien erklärt, ist aber nicht klausurrelevant.**

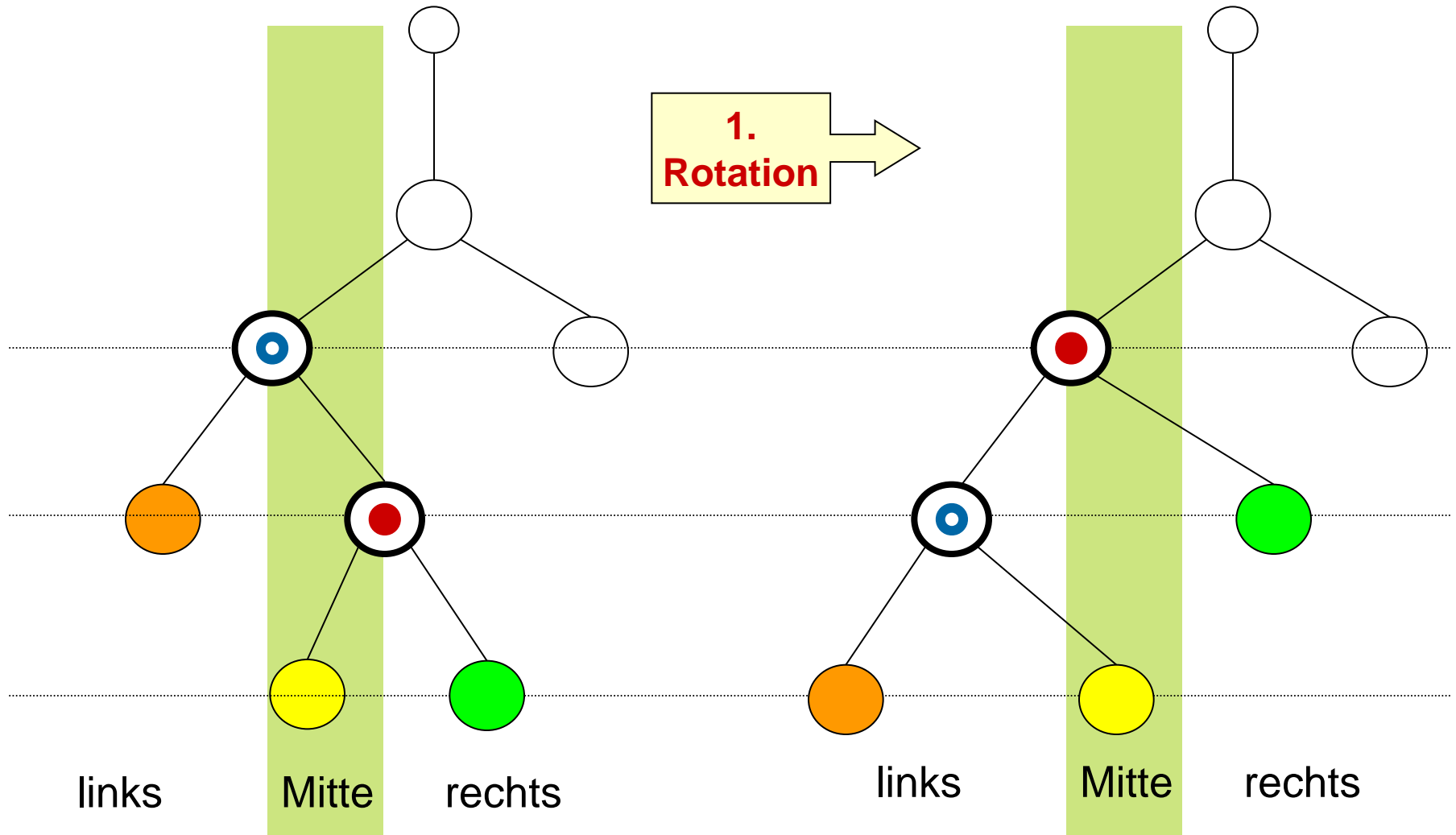
# Ausgleichen im AVL-Baum (eine Rotation)



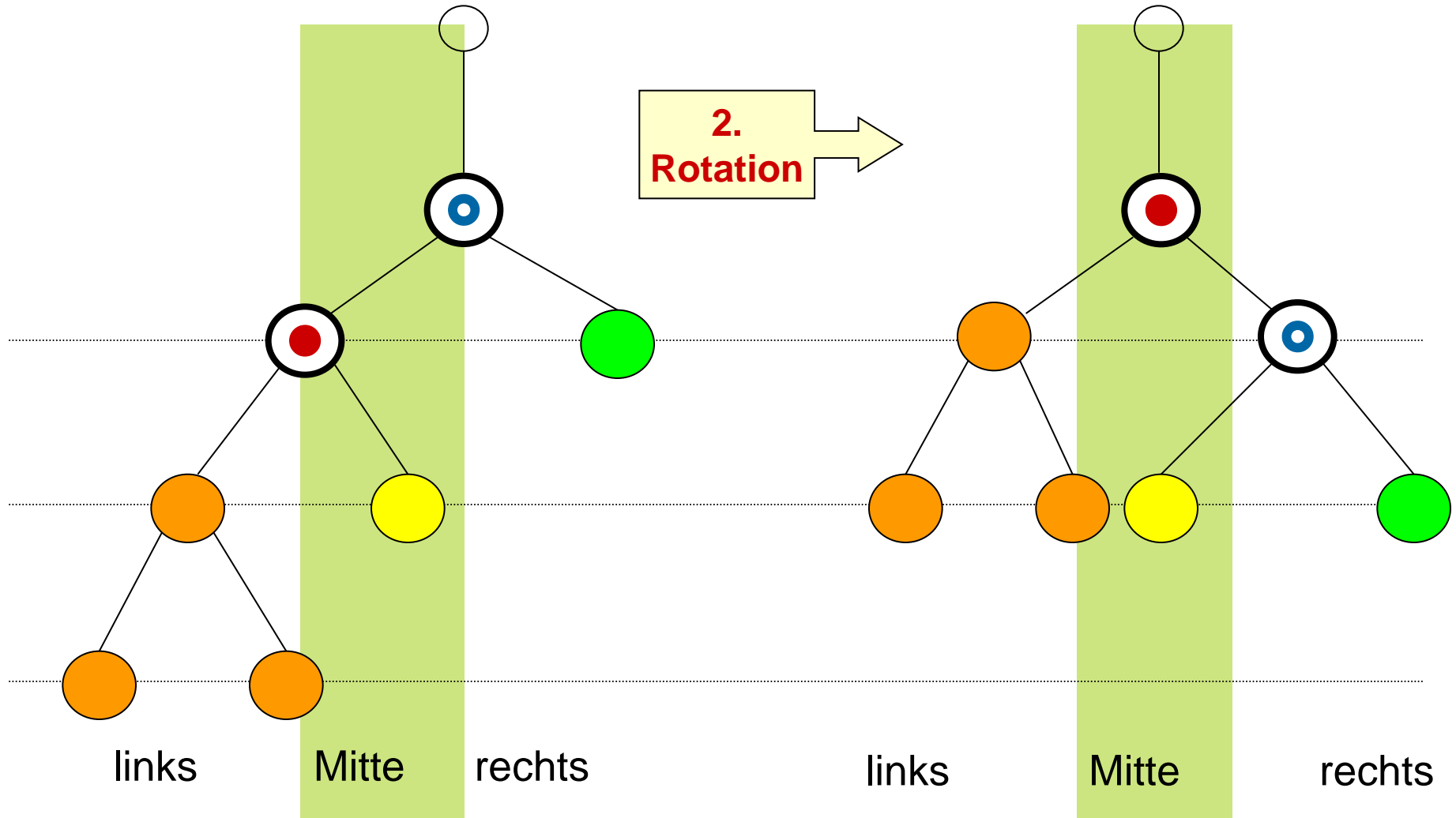
# Ausgleichen im AVL-Baum (eine Rotation)



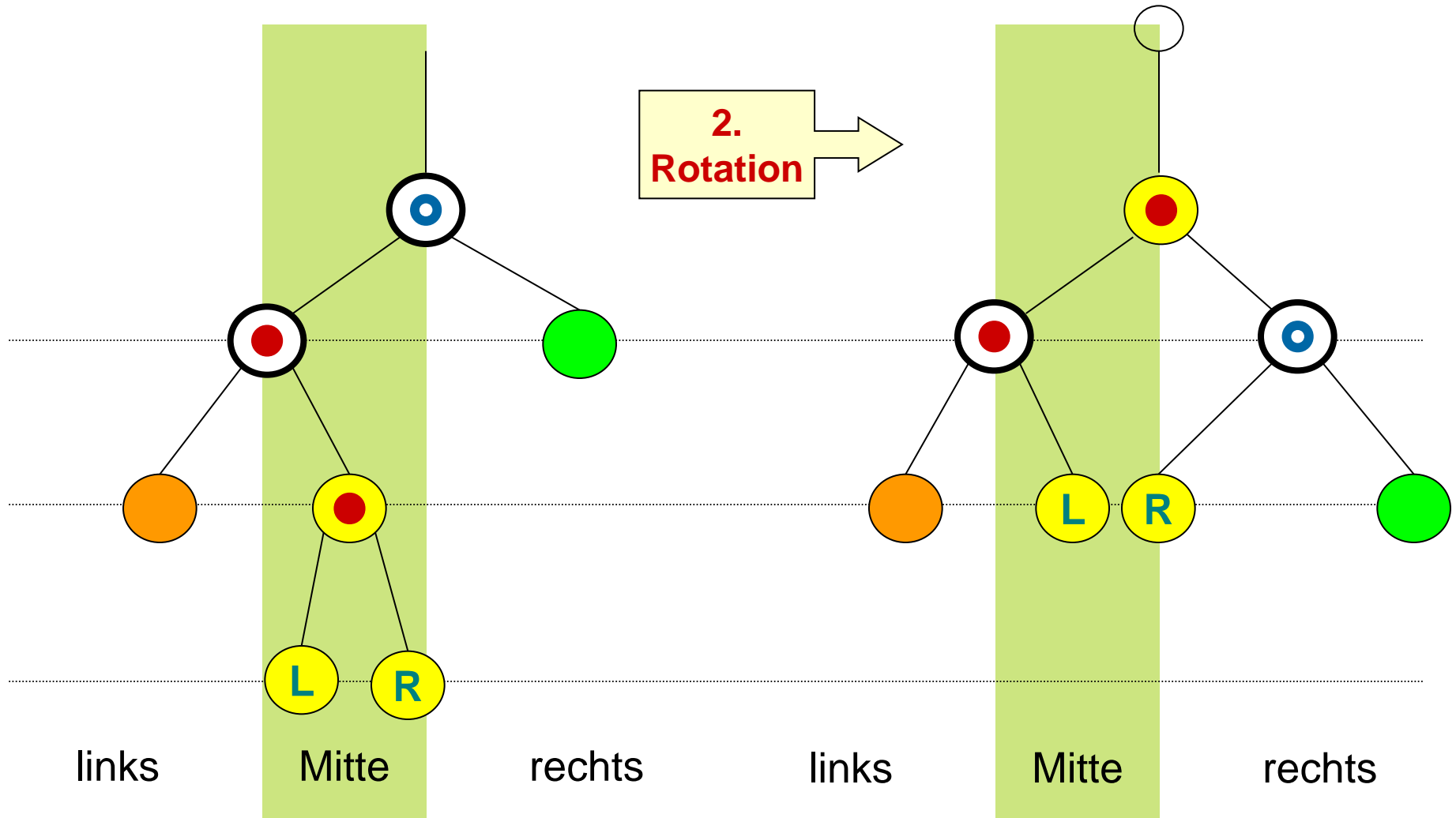
# Ausgleichen im AVL-Baum (Doppelrotation)



# Ausgleichen im AVL-Baum (Doppelrotation)



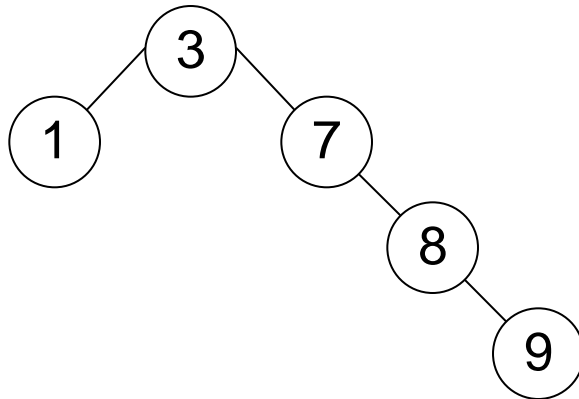
# Ausgleichen im AVL-Baum (Doppelrotation)



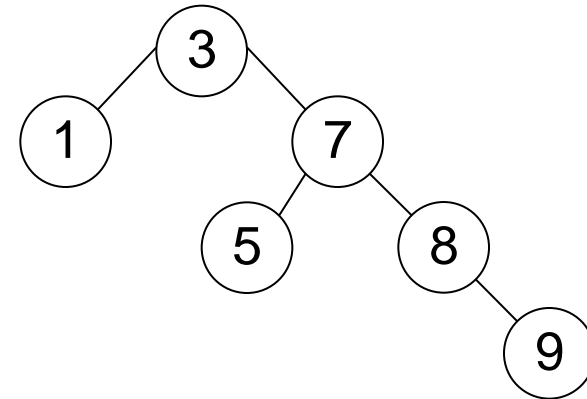


Gleiche die folgenden Bäume durch Rotationen aus

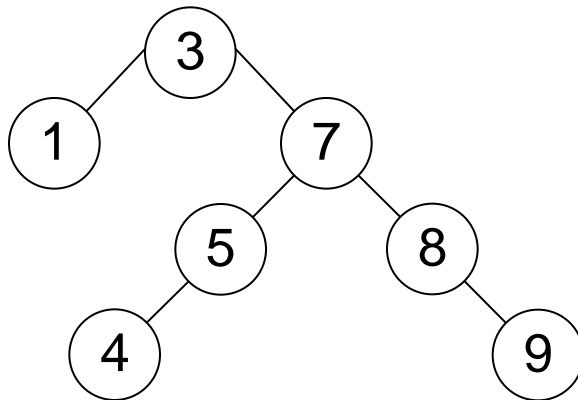
a)



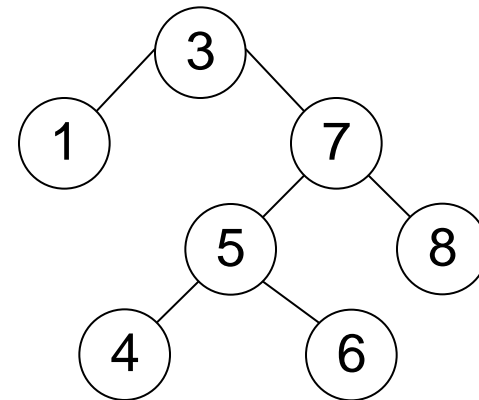
b)



c)

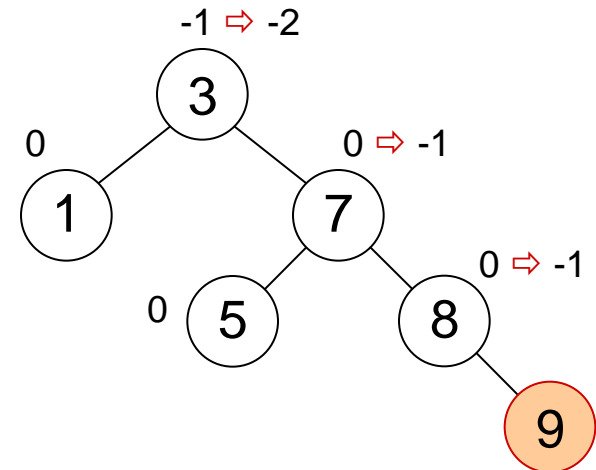


d)



- <http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>  
Kapitel AVL (3) und AVL (4)

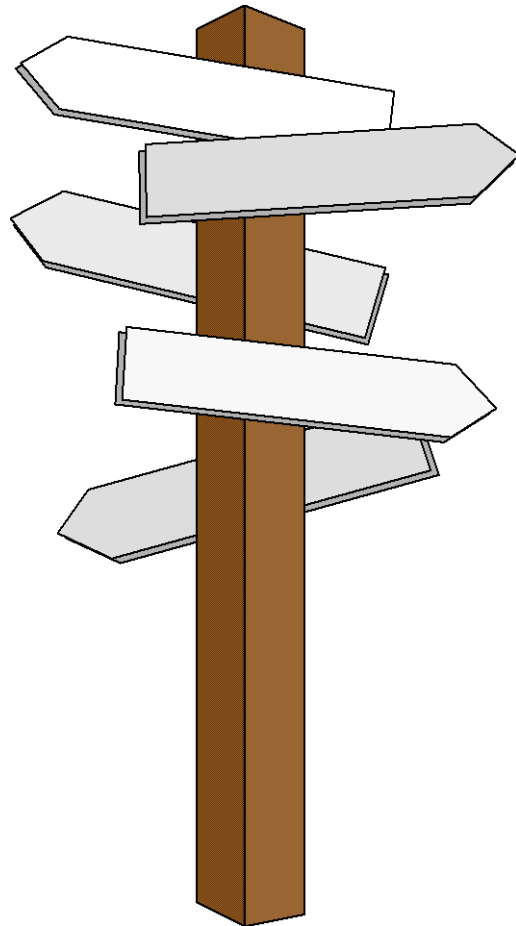
- Die Höhe eines AVL-Baums ist um maximal 45% höher als die eines optimal balancierten Baums.
  - Suchoperation hat  **$O(\log n)$** .
- Nach dem Einfügen wird maximal um alle Knoten zwischen der Wurzel und dem eingefügten Knoten rotiert.
  - Ähnlich beim Löschen.
  - Ausgleichen nach dem Einfügen oder Löschen hat  **$O(\log n)$** .



- **Einfügen:**
  - Element muss gesucht werden ( $O(\log n)$ ).
  - Element muss angehängt werden ( $O(1)$ ).
  - Baum muss ausgeglichen werden ( $O(\log n)$ ).  $O(\log n)$
- **Löschen:**
  - Element muss gesucht werden ( $O(\log n)$ ).
  - Das nächstgrößere Element muss gesucht werden (im worst case  $O(\log n)$ ).
  - Elemente müssen umkopiert werden ( $O(1)$ ).
  - Baum muss ausgeglichen werden ( $O(\log n)$ ).  $O(\log n)$
- **Prüfen/Auslesen:**
  - Element muss im Baum gesucht werden.  $O(\log n)$

## Komplexität von Mengen

|          | Unsortierte Liste | Sortierte Liste | Baum                          | (Dyn.) Hashtabelle    |
|----------|-------------------|-----------------|-------------------------------|-----------------------|
| Suchen   | $O(n)$            | $O(\log n)$     | <b><math>O(\log n)</math></b> | $O(1)$                |
| Einfügen | $O(n)$            | $O(n)$          | <b><math>O(\log n)</math></b> | $O(1)$<br>wst: $O(n)$ |
| Löschen  | $O(n)$            | $O(n)$          | <b><math>O(\log n)</math></b> | $O(1)$                |

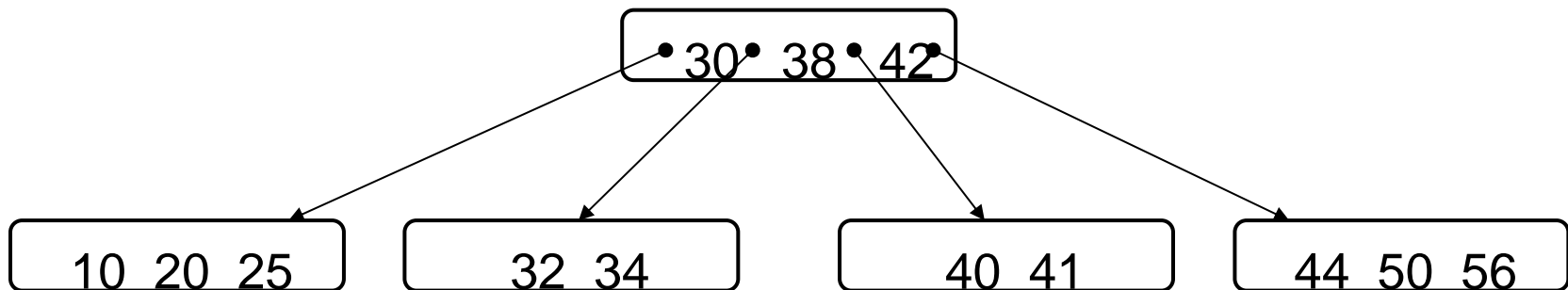


## Vorlesung 7

### B-Bäume

- **B-Bäume** wurden 1972 (1969?) von Rudolf **Bayer** und Edward M. **McCreight** eingeführt
- Ziel: effiziente Indexstrukturen für **riesige Datenbestände** (z.B. bei Datenbanken), die überwiegend auf externen Datenträgern (z.B. auf **Festplatten**) gespeichert sind.
- Viele Datenbanken (z.B. MySQL, Oracle, Access) benutzen (als Default) B-Bäume (bzw. eine Abart davon).

Jeder Knoten in einem **B-Baum der Ordnung  $d$**  enthält  $d$  bis  $2d$  Elemente.  
Die Wurzel bildet die einzige Ausnahme, sie kann 1 bis  $2d$  Elemente enthalten.  
Die Elemente in einem Knoten sind aufsteigend sortiert.  
Die Anzahl der Söhne in einem B-Baum ist entweder 0 (Blatt) oder um eins größer als die Anzahl der Elemente, die der Knoten enthält.



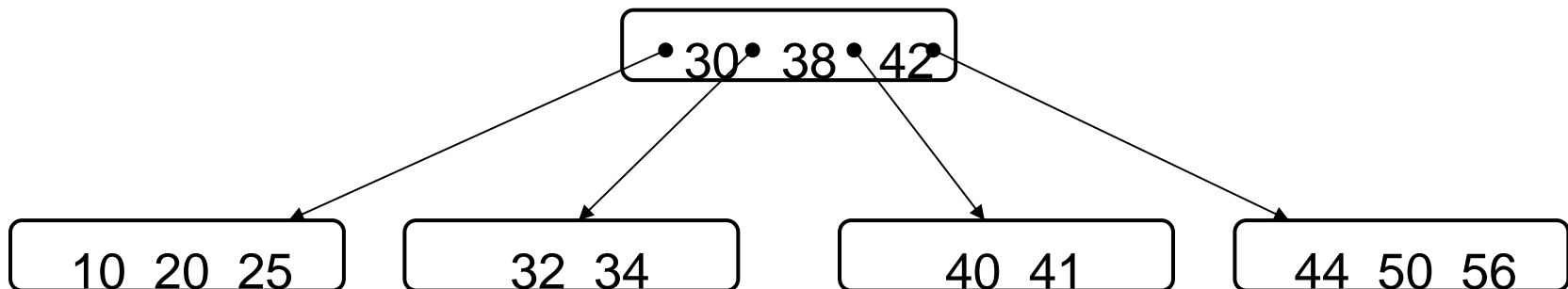


Alle Blätter liegen auf demselben Level.

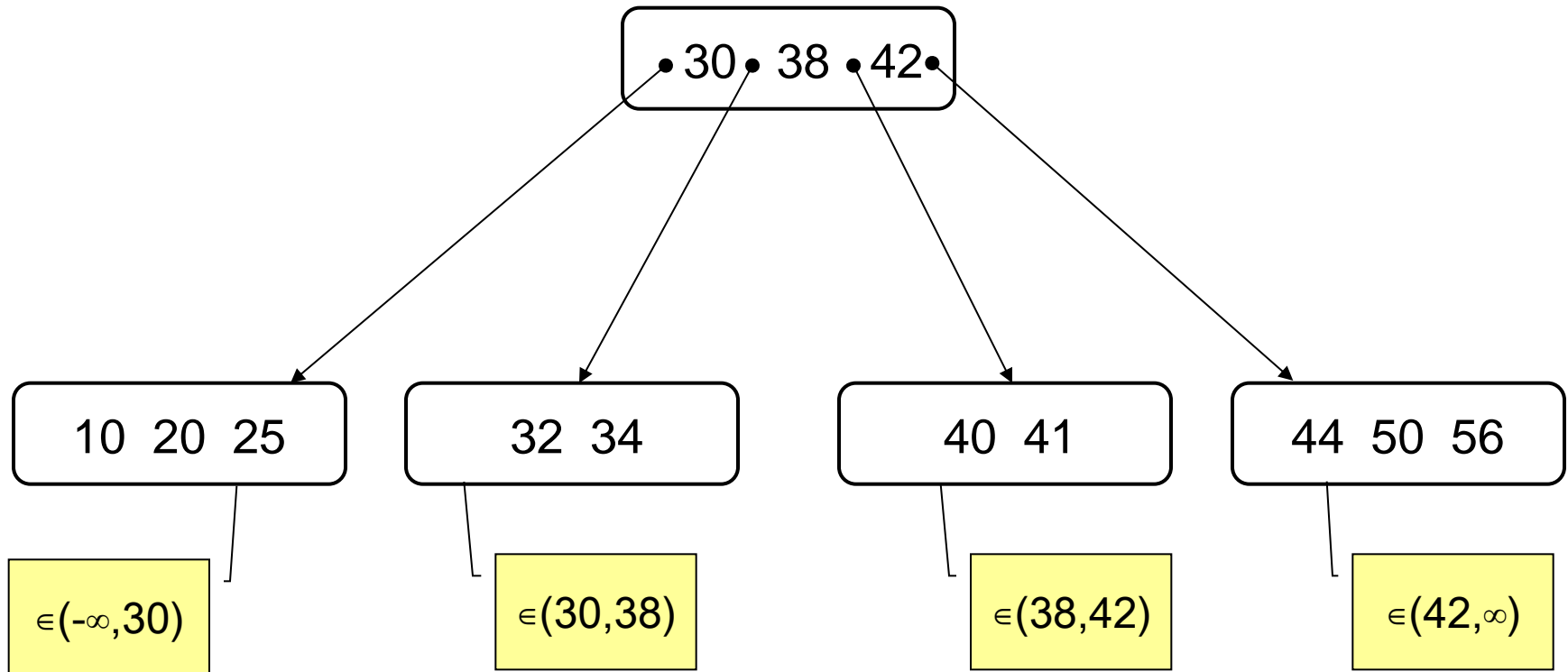
- Garantierte Zugriffszeiten.
- Bei realistischen Parametern (z.B. Ordnung 1000) sind nur sehr wenige ( $<5$ ) Zugriffe auf das externe Medium nötig.

B-Bäume besitzen ausgeglichene Höhe, lassen aber unausgeglichene Verzweigungsgrad und Knotenfüllgrad zu.

Der **längste Weg** in einem B-Baum der Ordnung  $d$  ist in  $O(\log_{d+1} n)$  (ohne Rechnung).



- Bisherige Definition: Jeder Knoten in einem B-Baum der Ordnung  $d$  enthält  **$d$  bis  $2d$**  Elemente.
- Es gibt eine weitere Variante des B-Baums. Jeder Knoten in einem B-Baum der Ordnung  $d$  enthält hier  **$d-1$  bis  $2d-1$**  Elemente.
- Wir behandeln nur die erste Variante. Bei der zweiten Variante ist das Einfügen geringfügig anders. Das bekannteste Beispiel der 2. Variante sind die  $(2,3,4)$ -Bäume (mit 1,2 oder 3 Element-Knoten).



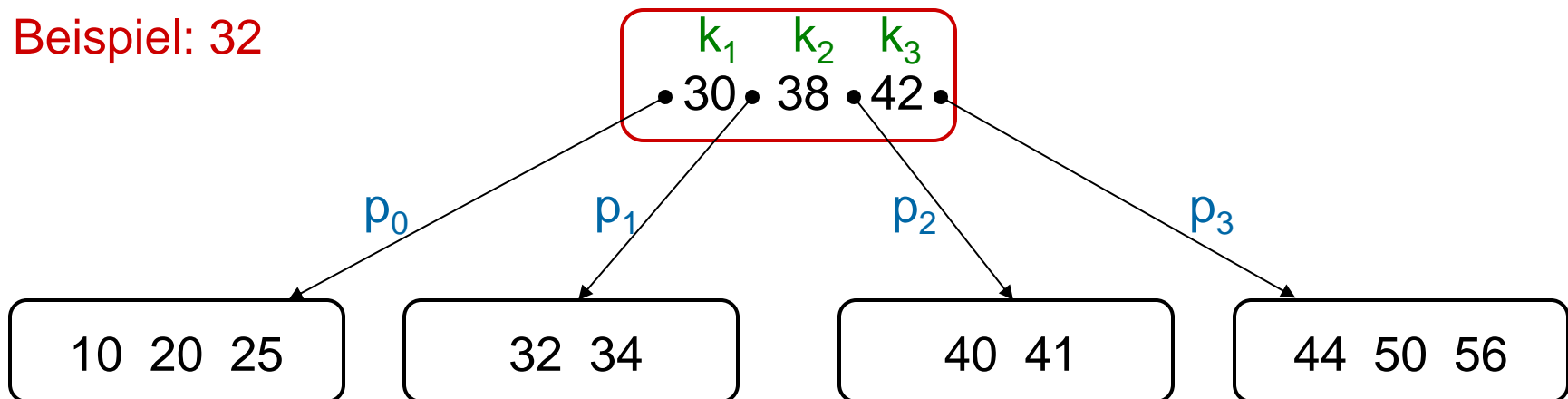
Ausgehend von der Wurzel:

1. Prüfe ob der gerade betrachtete Knoten den gesuchten Schlüssel  $m$  enthält

Suche innerhalb eines Knotens: linear oder binär.

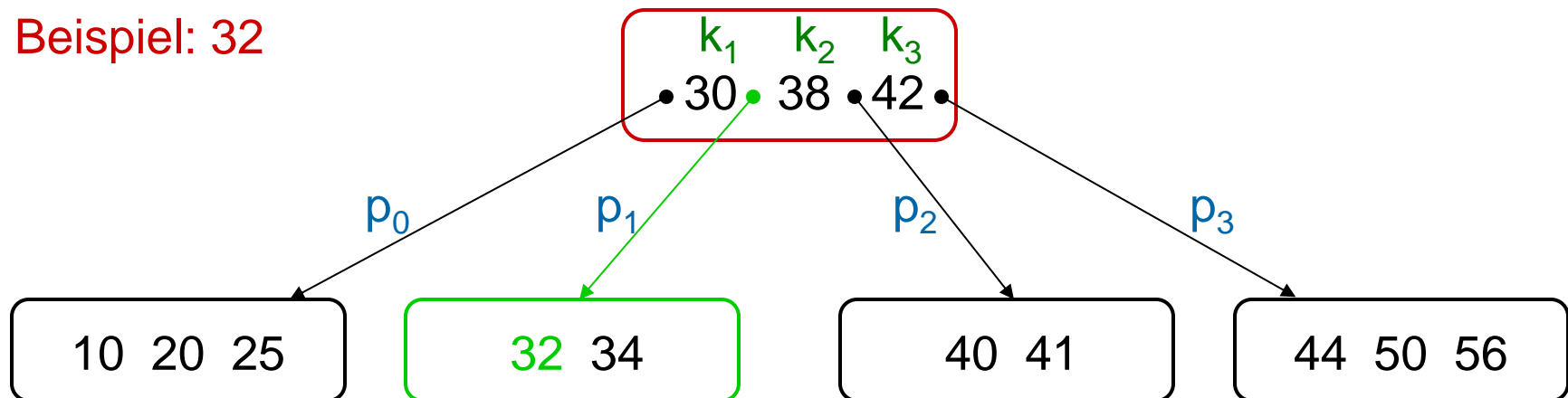
(Nicht wichtig, weil Laufzeit hauptsächlich von Anzahl der Zugriffe auf Hintergrundspeicher (Festplatte) abhängt.)

Beispiel: 32



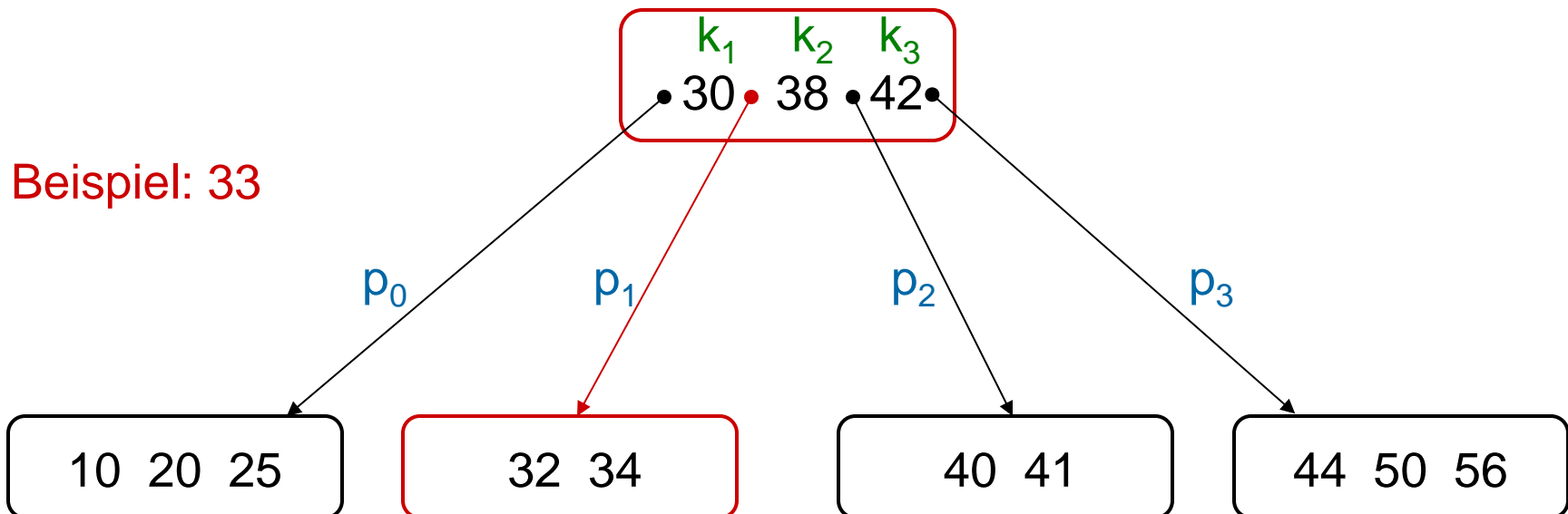
- Falls nicht, bestimme den kleinsten Schlüssel  $k_i$ , der größer als  $m$  ist.
  - $k_i$  gefunden: Weiter bei Schritt 1. mit Sohn links von  $k_i$ :  $p_{i-1}$ .
  - $k_i$  nicht gefunden: Weiter bei Schritt 1. mit letztem Sohn:  $p_n$ .

Beispiel: 32



## Suchen in einem B-Baum (3)

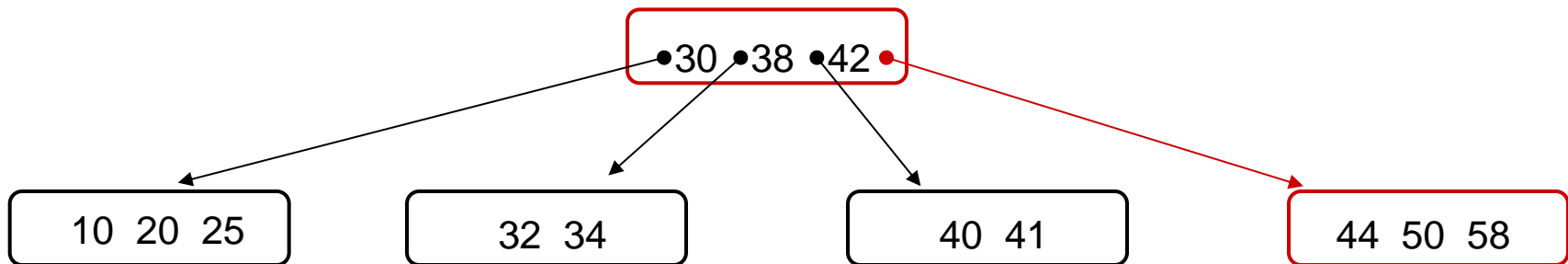
Erfolgreiche Suche endet in einem Blatt.



1. Suche nach Schlüssel endet (scheitert) in einem Blattknoten „*node*“.
2. Schlüssel in *node* in Sortierreihenfolge **einfügen** (und neuen leeren Verweis einfügen).
3. Falls *node* nun **überfüllt** ist ( $2d+1$  Elemente): *node* **aufteilen**  
*k* sei mittlerer Eintrag von *node*.
  1. **Neuen Knoten** „*current*“ anlegen und mit den  $d$  größeren Schlüsseln (rechts von *k*) belegen. Die  $d$  kleineren Schlüssel (links von *k*) bleiben in *node*.
  2. Falls *node* **Wurzelknoten** war:
    - neue Wurzel „*parent*“ anlegen.
    - Verweis ganz links in *parent* mit *node* verbinden.
  3. *k* in Vaterknoten „*parent*“ von *node* **verschieben**.
  4. Verweis rechts von *k* in *parent* mit *current* **verbinden**.
4. Falls *parent* nun überfüllt ist: Schritt 3. mit *parent* **wiederholen**.

## Beispiel: Einfügen von 56.

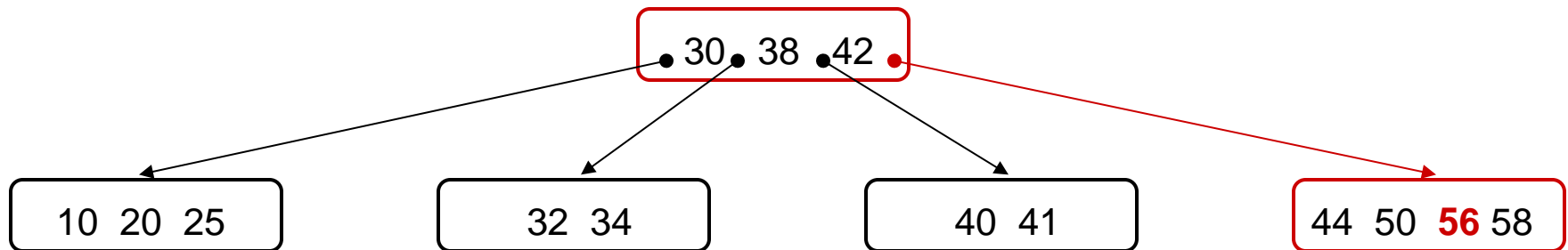
1. **Suche** nach Schlüssel endet (scheitert) in einem Blattknoten „*node*“.





### Beispiel: Einfügen von 56.

2. Schlüssel in *node* in Sortierreihenfolge **einfügen** (und neuen leeren Verweis einfügen).

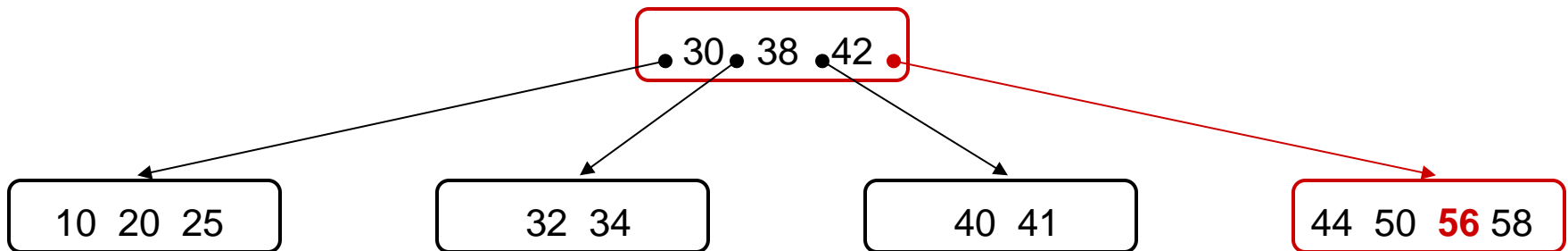


### Beispiel: Einfügen von 56.

3. Falls *node* nun **überfüllt** ist ( $2d+1$  Elemente):

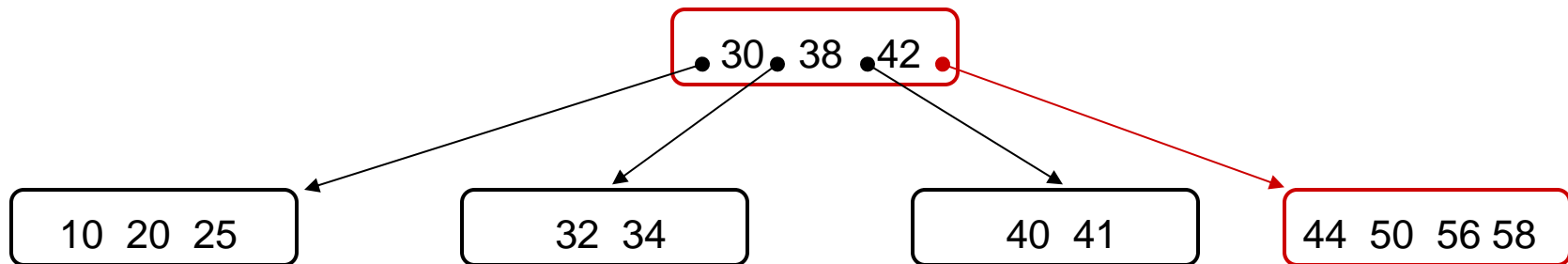
→ ist nicht der Fall

→ Ende



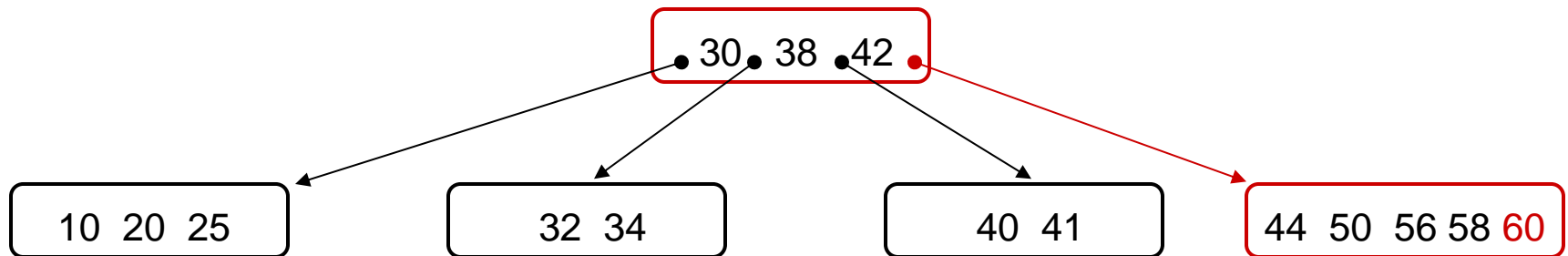
### Beispiel: Einfügen von 60.

1. **Suche** nach Schlüssel endet (scheitert) in einem Blattknoten „*node*“.



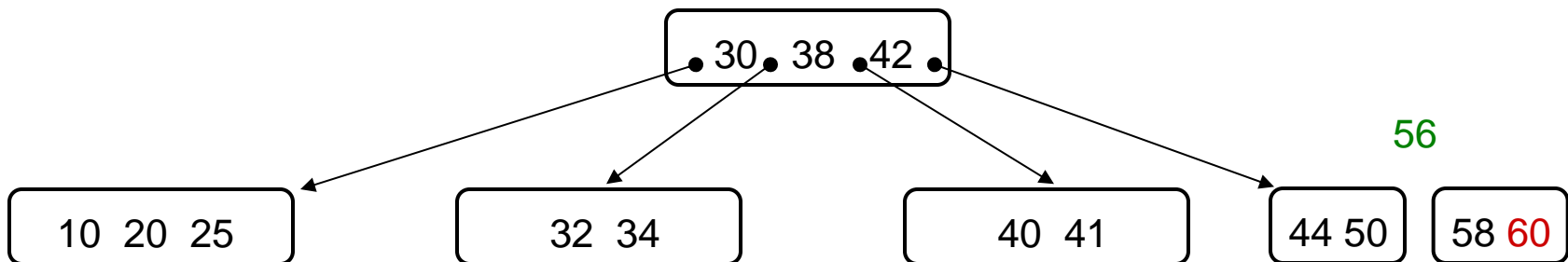
### Beispiel: Einfügen von 60.

2. Schlüssel in *node* in Sortierreihenfolge **einfügen** (und neuen leeren Verweis einfügen).



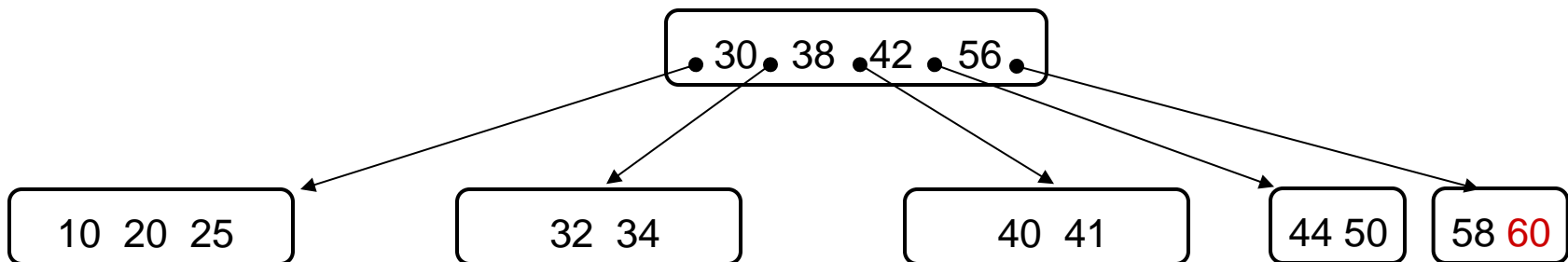
## Beispiel: Einfügen von 60.

3. Falls *node* nun **überfüllt** ist ( $2d+1$  Elemente): *node* **aufteilen**  
*k* sei mittlerer Eintrag von *node*.
  1. **Neuen Knoten** „*current*“ anlegen und mit den *d* größeren Schlüsseln (rechts von *k*) belegen. Die *d* kleineren Schlüssel (links von *k*) bleiben in *node*.
  2. Falls *node* **Wurzelknoten** war:
    - neue Wurzel „*parent*“ anlegen.
    - Verweis ganz links in *parent* mit *node* verbinden.
  3. *k* in Vaterknoten „*parent*“ von *node* **verschieben**.
  4. Verweis rechts von *k* in *parent* mit *current* **verbinden**.



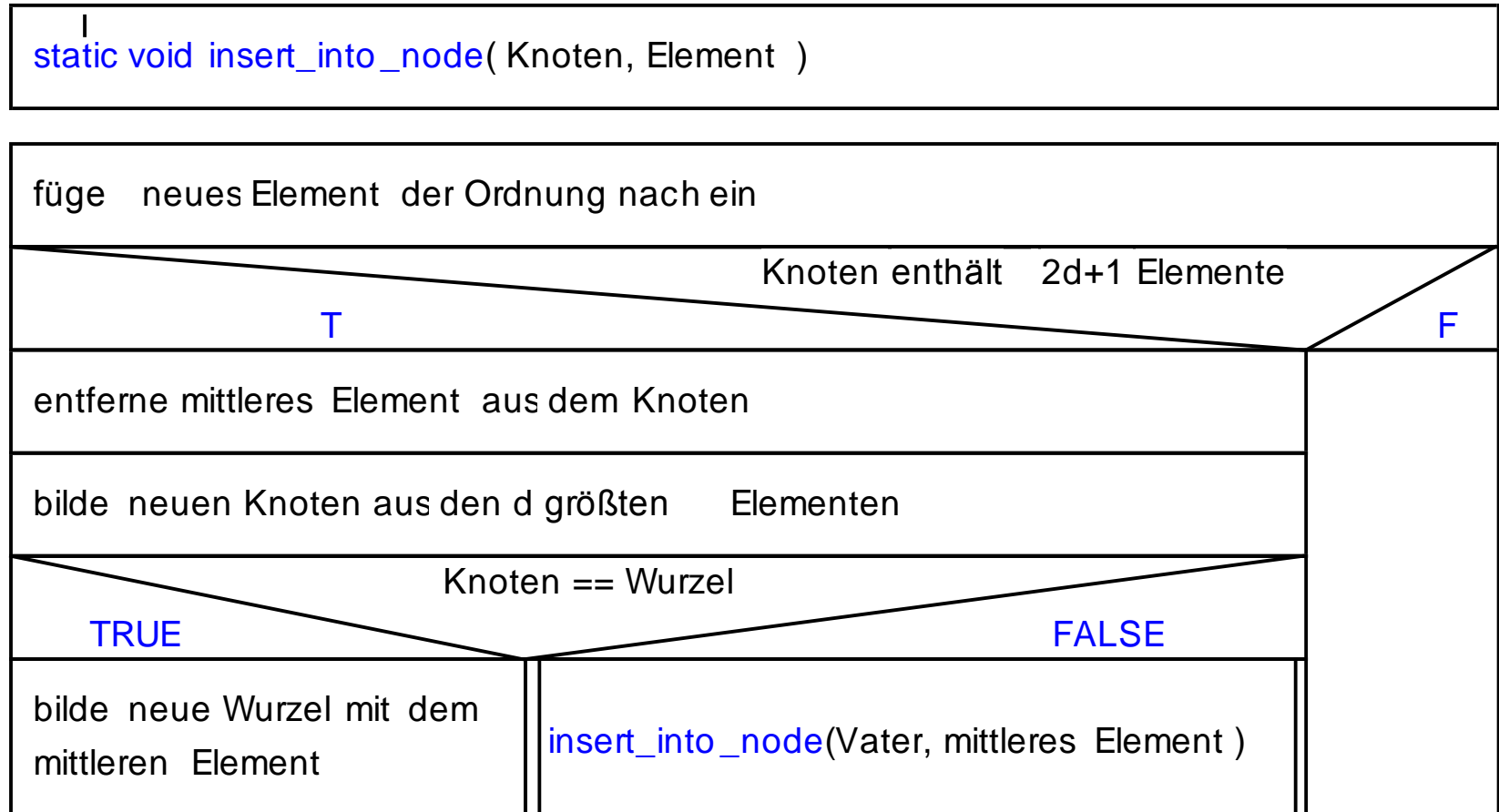
## Beispiel: Einfügen von 60.

3. Falls *node* nun **überfüllt** ist ( $2d+1$  Elemente): *node* aufteilen  
*k* sei mittlerer Eintrag von *node*.
  1. **Neuen Knoten** „*current*“ anlegen und mit den *d* größeren Schlüsseln (rechts von *k*) belegen. Die *d* kleineren Schlüssel (links von *k*) bleiben in *node*.
  2. Falls *node* **Wurzelknoten** war:
    - neue Wurzel „*parent*“ anlegen.
    - Verweis ganz links in *parent* mit *node* verbinden.
  3. *k* in Vaterknoten „*parent*“ von *node* **verschieben**.
  4. Verweis rechts von *k* in *parent* mit *current* **verbinden**.



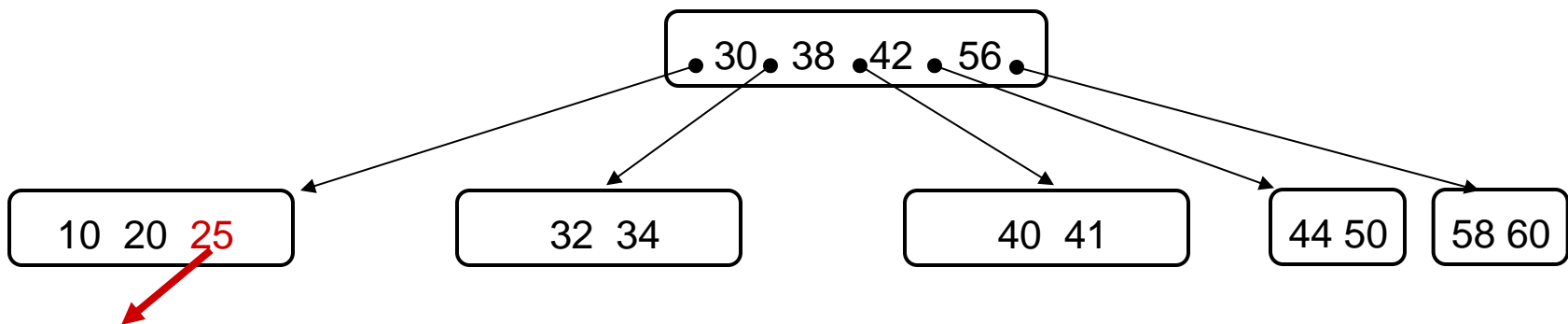
3. Falls *parent* nun überfüllt ist: Schritt 3. mit *parent* **wiederholen**.  
→ in diesem Beispiel nicht nötig.

# Element in Knoten einfügen



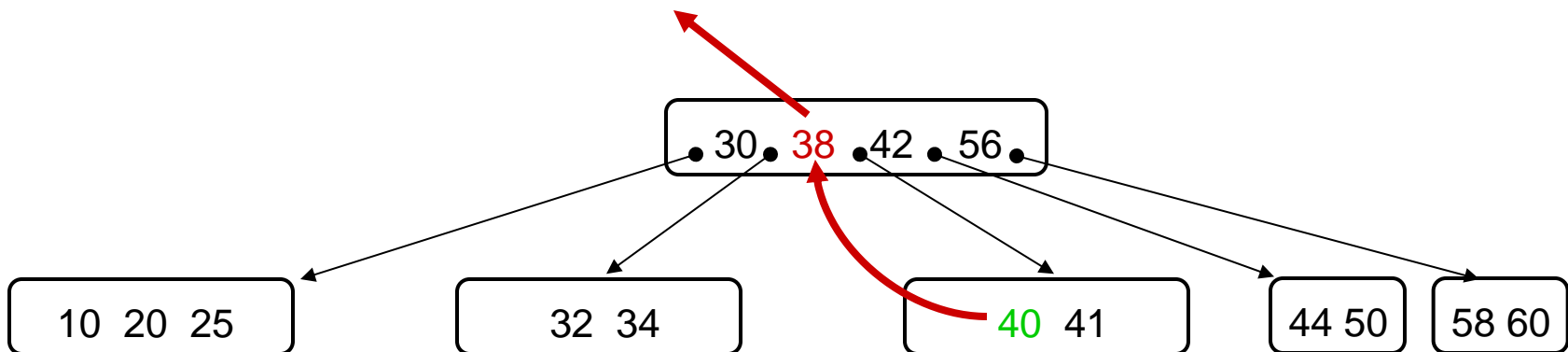


- Unterscheidung: der zu löschende Schlüssel  $x=k_i$  liegt
  - a) in einem Blatt mit Struktur  $(\text{null}, k_1, \dots, \text{null}, k_i, \text{null}, \dots, k_n, \text{null})$ : Entfernen von  $x=k_i$  und der darauf folgenden null-Referenz. Ein  $\rightarrow$  Underflow tritt auf, falls  $n=d$  war.

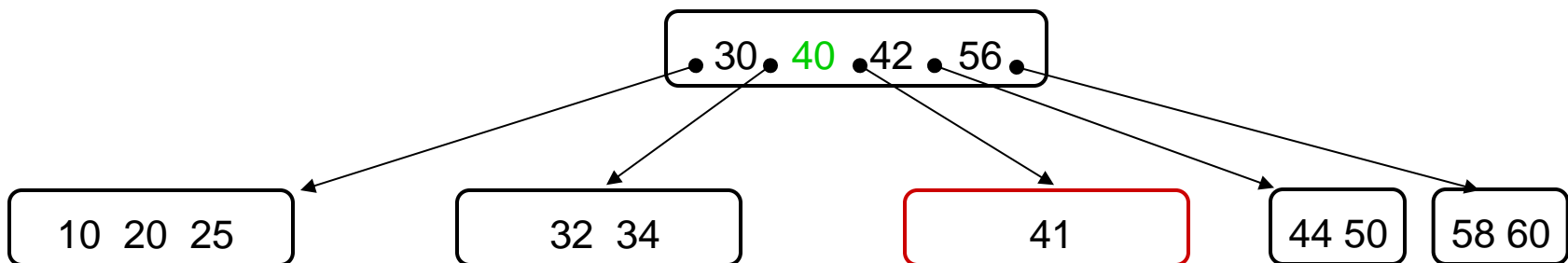


b) in einem inneren Knoten  $(p_0, k_1, \dots, k_i, p_i, \dots, k_n, p_n)$ :

- alle Referenzen haben einen Wert ungleich *null*.
- analog zu Löschen aus binären Suchbäumen: Finde kleinsten Schlüssel *s* im durch  $p_i$  referenzierten Teilbaum (in einem Blatt).
- Ersetze  $k_i$  durch *s* und lösche *s* aus dem Blatt (Fall a).



- **Underflow-Problem:** zu wenig  $(d-1)$  Schlüssel im Knoten



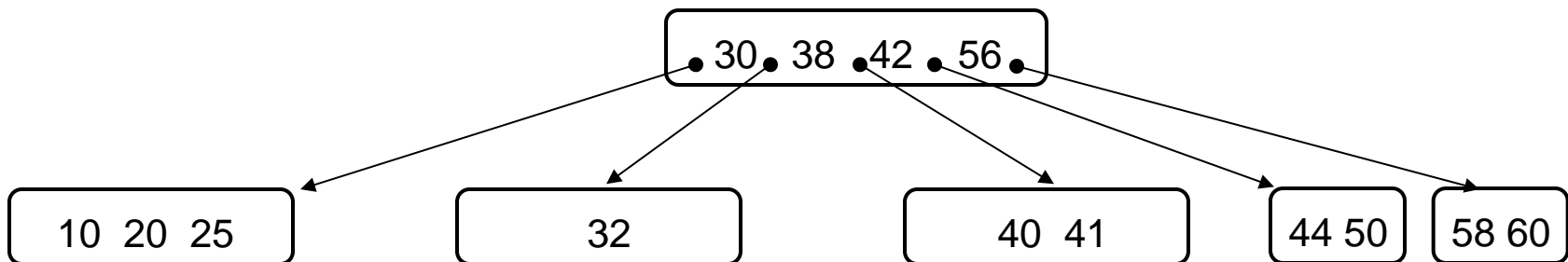
- Bei der Behandlung des Underflows sind wieder verschiedene Fälle möglich.
- Wir benötigen zwei Operationen:
  1. **Ausgleichen** (zwischen zwei Bruder-Knoten)
  2. **Verschmelzen** (von zwei Bruder-Knoten)

- Voraussetzung für Ausgleichsoperation bei Underflow in Knoten  $q$ :  $q$  hat **benachbarten** Bruder-Knoten  $p$  mit mehr als  $d$  Schlüsseln

- Annahme:

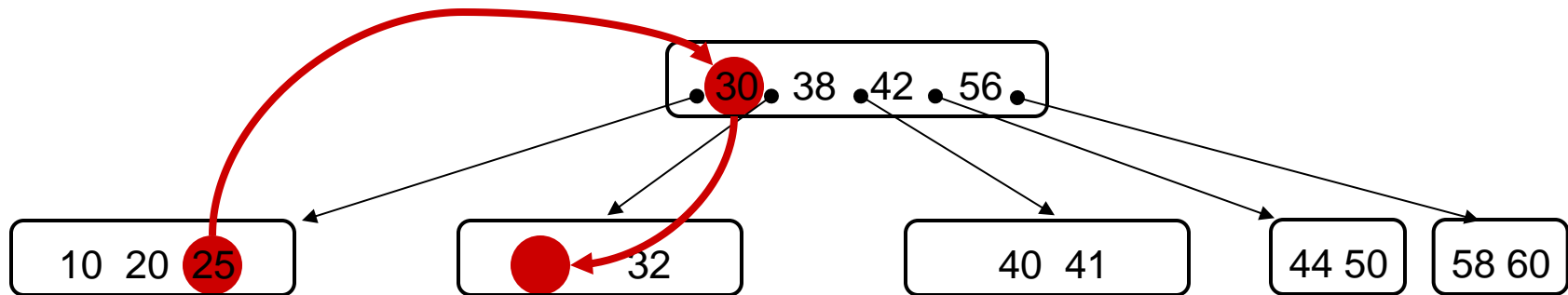
- $p$  ist linker Bruder von  $q$ .
- im Vater  $parent$  (von  $p$  und  $q$ ) trennt der Schlüssel  $t$  die Verweise auf  $p$  und  $q$ .

Falls vorhanden und ausreichend besetzt. Sonst analog mit rechtem Bruder.



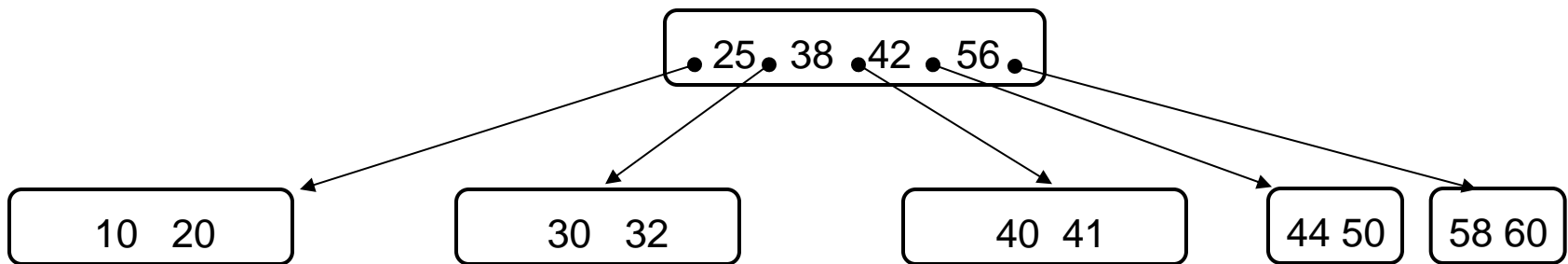
## Ausgleich zwischen Bruder-Knoten (2)

- Idee:  $p$  schenkt  $q$  ein Element („Umweg“ über Vater)



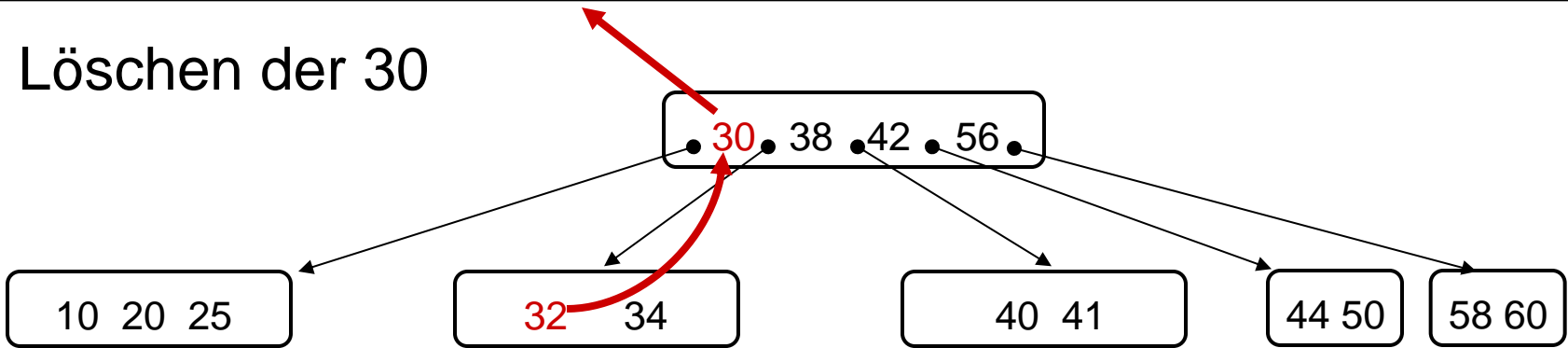
## Ausgleich zwischen Bruder-Knoten (3)

- Ergebnis:

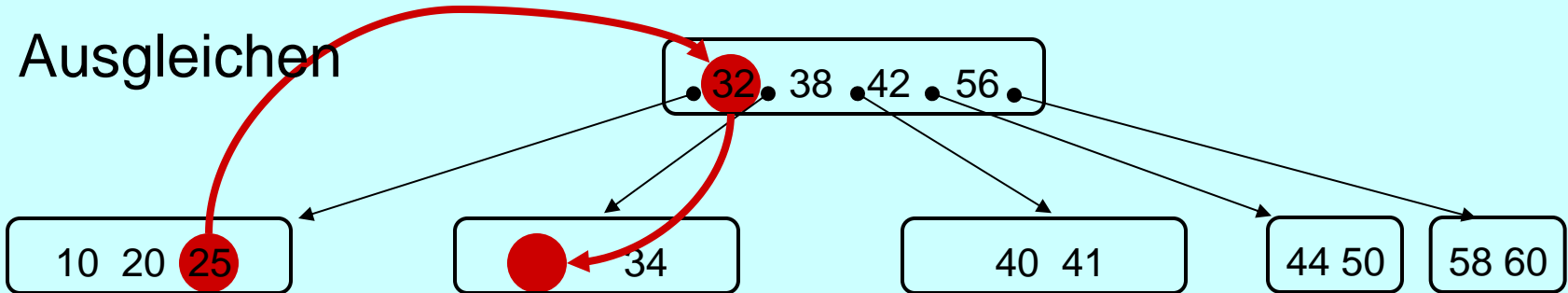


# Beispiel: Löschen aus innerem Knoten

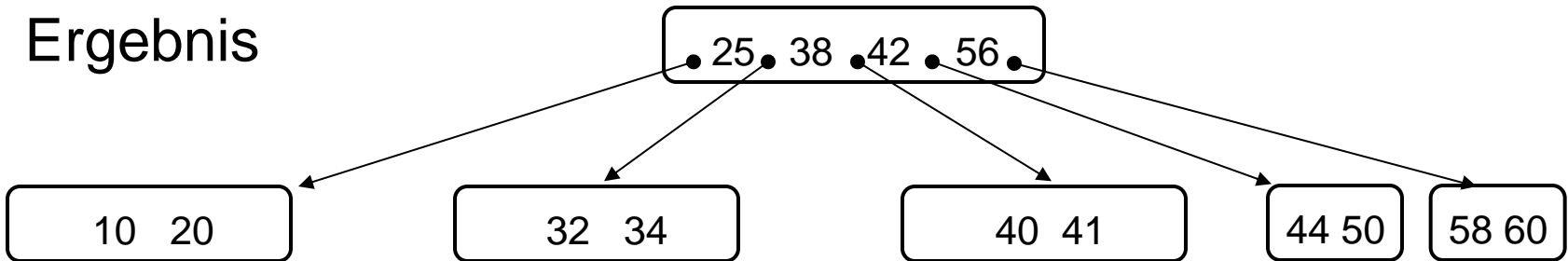
Löschen der 30



Ausgleichen



Ergebnis

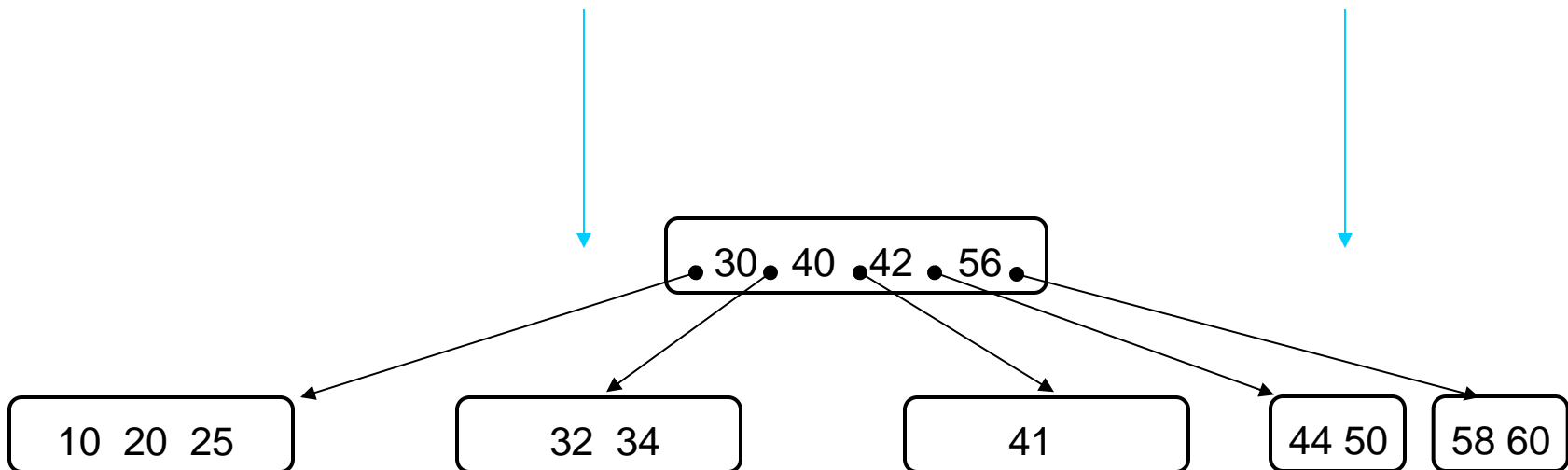




## Verschmelzen von Bruder-Knoten (1)

- Verschmelzung mit Bruder-Knoten  $p$  bei Underflow in Knoten  $q$ : Alle benachbarten Brüder von  $q$  haben nur  $d$  Schlüssel.

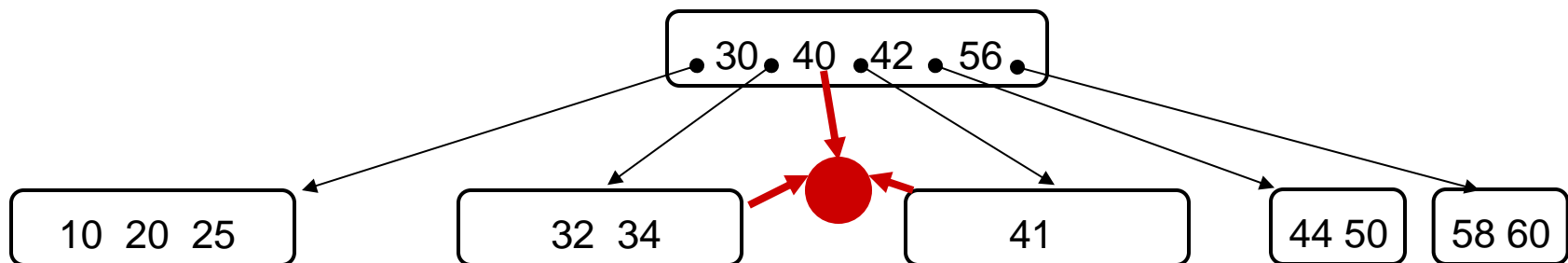
Nachbarknoten haben nur 2 Schlüssel



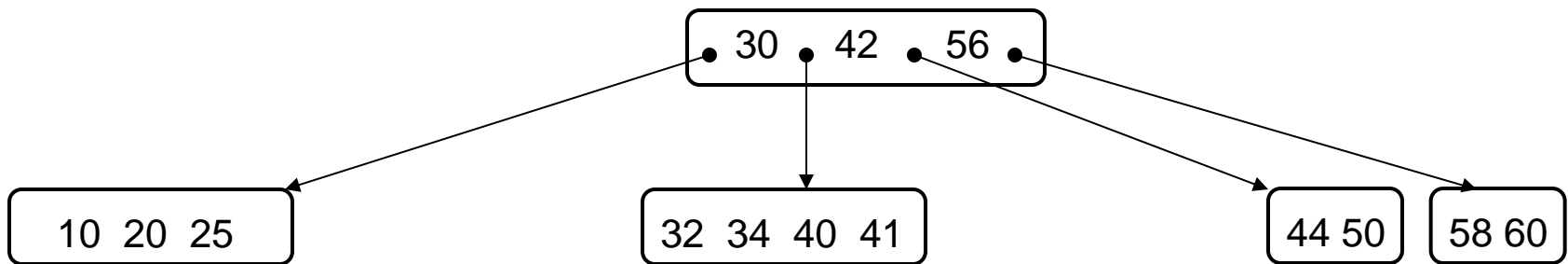
## Verschmelzen von Bruder-Knoten (2)

- Annahme:
  - $p$  ist linker Bruder von  $q$ .
  - im Vater  $parent$  (von  $p$  und  $q$ ) trennt der Schlüssel  $t$  die Verweise auf  $p$  und  $q$ .
- Idee:  $p$  und  $q$  mit dem trennenden Element aus  $parent$  verschmelzen.

Falls vorhanden.  
Sonst analog mit  
rechtem Bruder.



- **Ergebnis:**



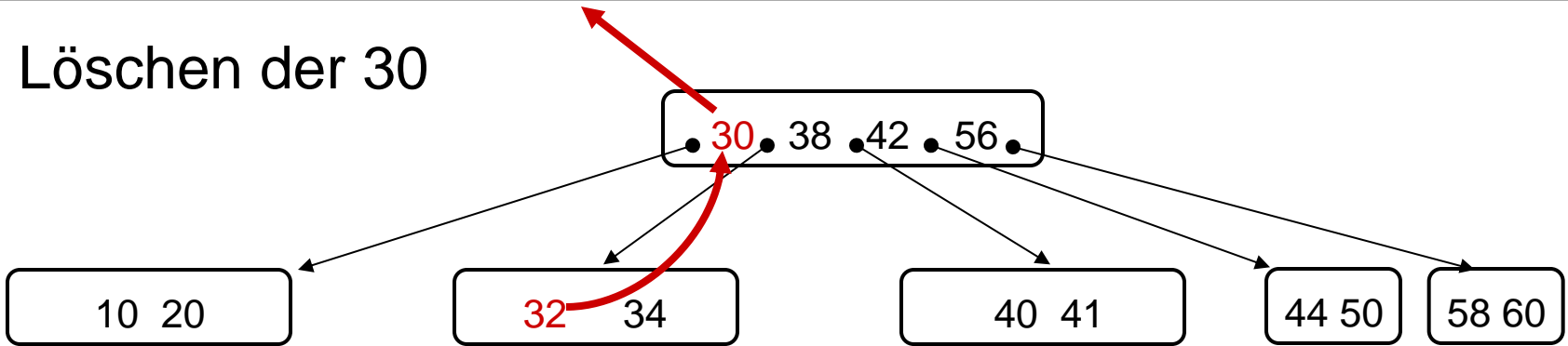
- Jetzt eventuellen Underflow in `parent` behandeln...
- Underflow-Behandlung rekursiv nach oben fortsetzen.
- Die Wurzel darf auch einen einzigen Schlüssel enthalten.
- Falls zuletzt der letzte Schlüssel aus der Wurzel gelöscht wird (bei Verschmelzung der beiden letzten Söhne der Wurzel zu einem einzigen Knoten):

**Einzigster Nachfolger der Wurzel wird neue Wurzel.**

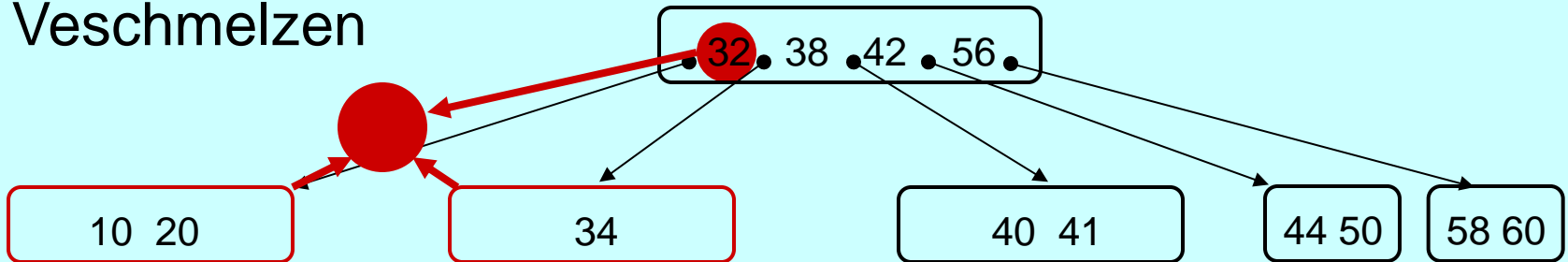
**⇒ Höhe des B-Baums um 1 verringert.**

# Beispiel: Löschen aus innerem Knoten

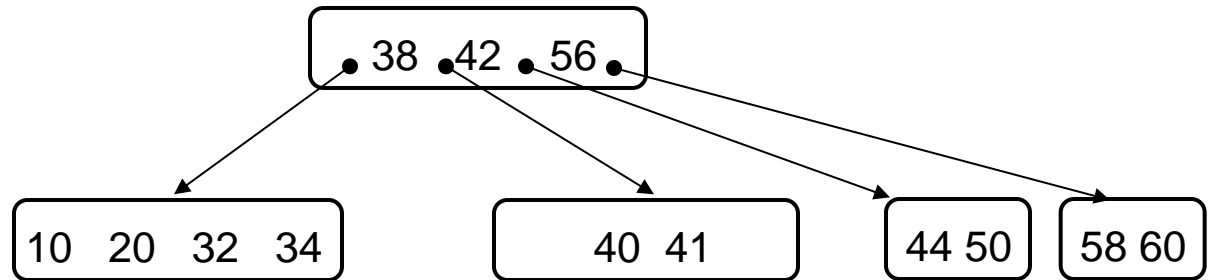
Löschen der 30



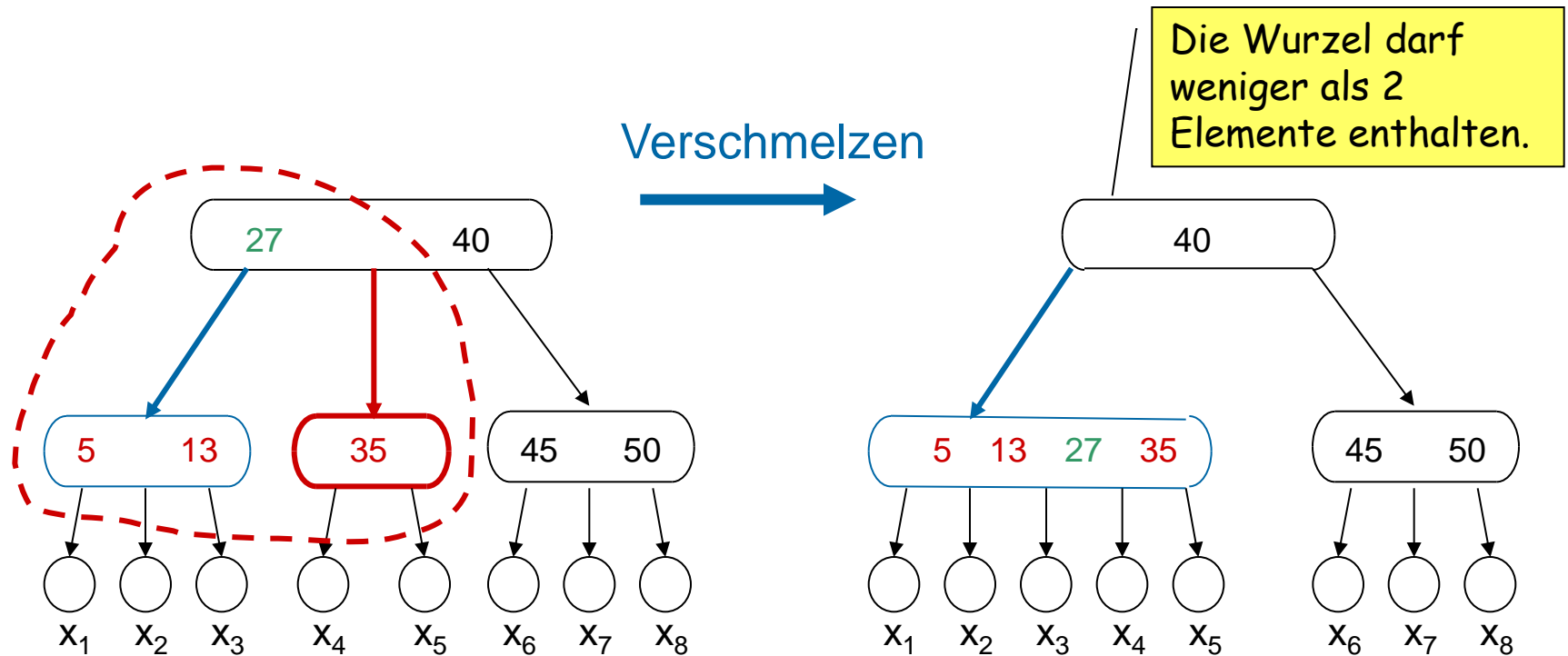
Veschmelzen



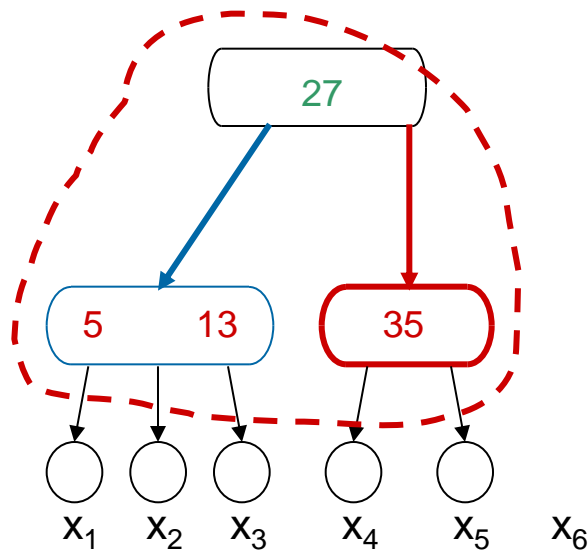
Ergebnis



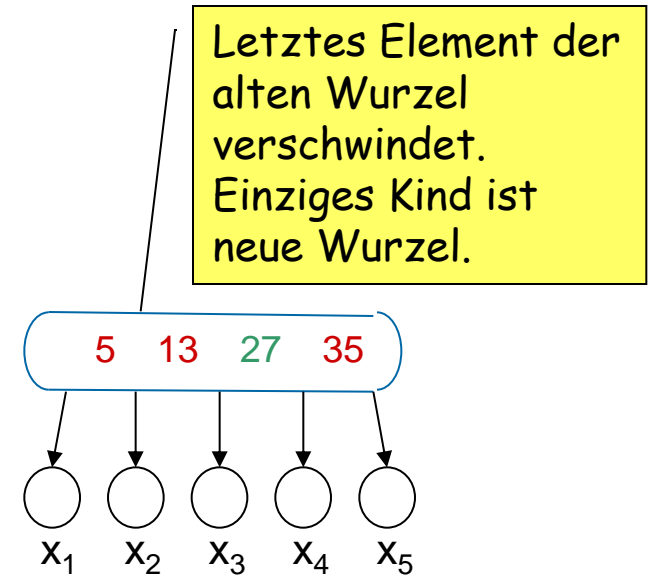
# Spezialfall Wurzel (1)

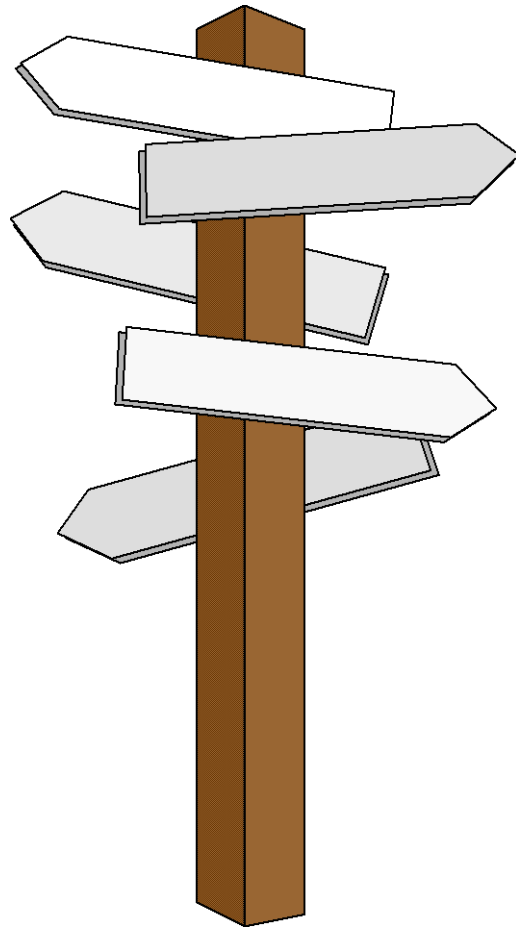


# Spezialfall Wurzel (2)



Verschmelzen





## Vorlesung 8

### Varianten von B-Bäumen Baumdurchlauf (1)

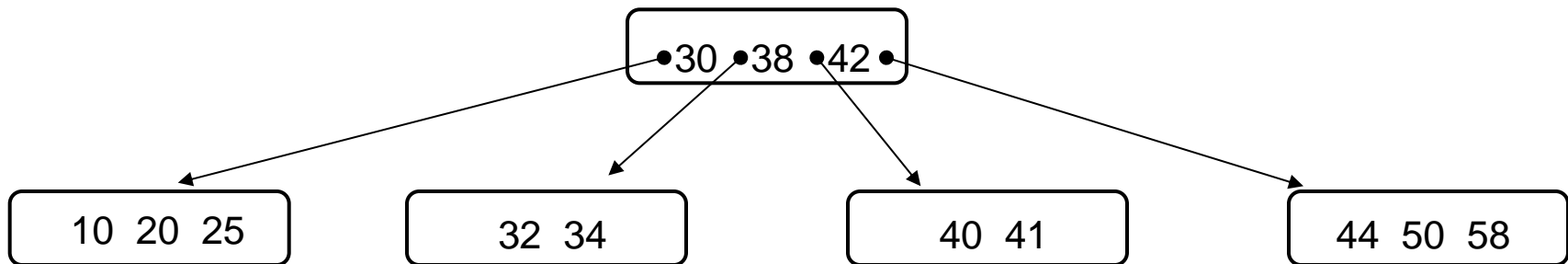


- Einfügen, Löschen, Prüfen und Auslesen sind auch hier  $O(\log n)$ .
- Vorteile haben B-Bäume bei sehr großen Datenmengen, die auf der Festplatte liegen.
  - **Es sei die Ordnung  $d=1024$**
  - **Bei einer Höhe von  $h=4$  können bereits  $1024^4-1 \approx 10^{12}$  Schlüssel gespeichert werden.**
  - **Dabei müssen für jede Suchanfrage maximal 5 Baumknoten inspiziert werden.**
- Bei kleinen Datenmengen werden eher Hashtabellen verwendet.

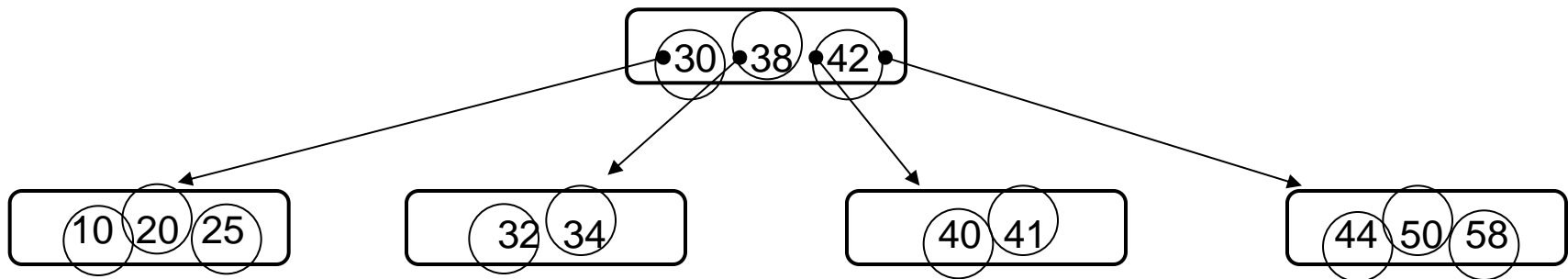
- Die wichtigsten Unterschiede zwischen B-Bäumen und B+-Bäumen sind:
- B-Bäume
  - trennen Datensätze nicht von Schlüsseln.
- B+ -Bäume
  - speichern in den inneren Knoten nur Schlüssel.
  - Die eigentlichen Datensätze befinden sich ausschließlich in den Blättern.
  - Dies ist bei der Anwendung für Datenbanken naheliegend und sinnvoll.

- <http://www.seanster.com/BplusTree/BplusTree.html>

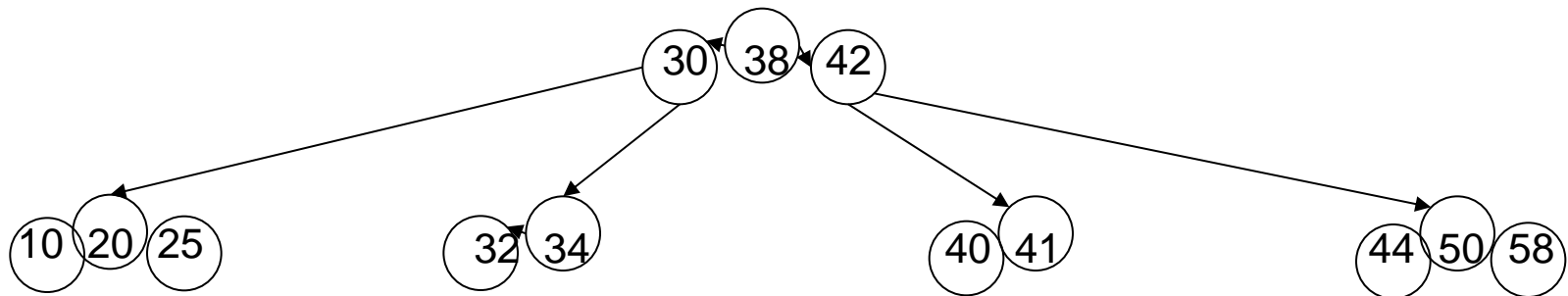
- Rot-Schwarz-Bäume kombinieren (2,3,4)-Bäume (eine Variante der B-Bäume) mit Binärbäumen.
- Gehen wir zunächst vom untenstehenden (2,3,4)-Baum aus.



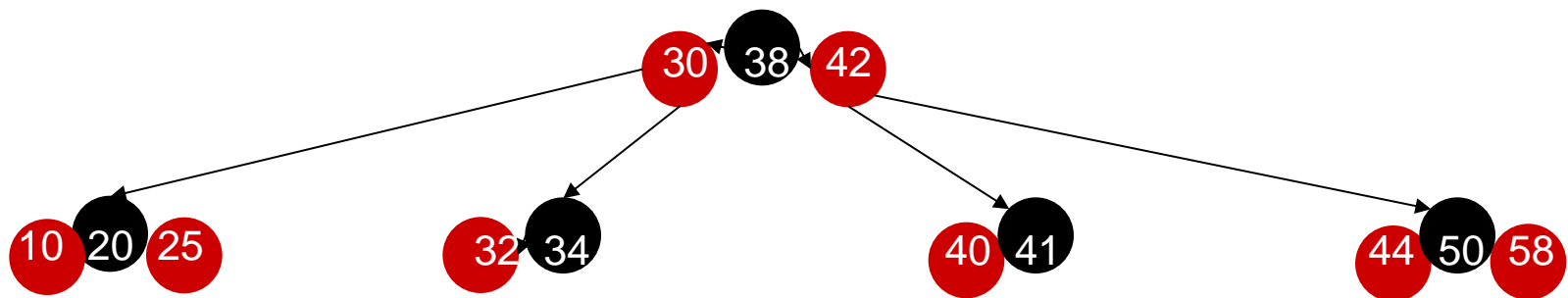
- Die einzelnen B-Baum-Knoten werden als kleine Binärbäume betrachtet, so dass jeder Knoten wieder nur 1 Element besitzt.



- Die einzelnen B-Baum-Knoten werden als kleine Binärbäume betrachtet, so dass jeder Knoten wieder nur 1 Element besitzt.



- Jeder Knoten erhält ein zusätzliches boolean-Attribut, das sagt, ob der Knoten „rot“ oder „schwarz“ ist.
- Die Wurzeln der kleinen Binärbäume sind schwarz; die anderen Knoten rot.

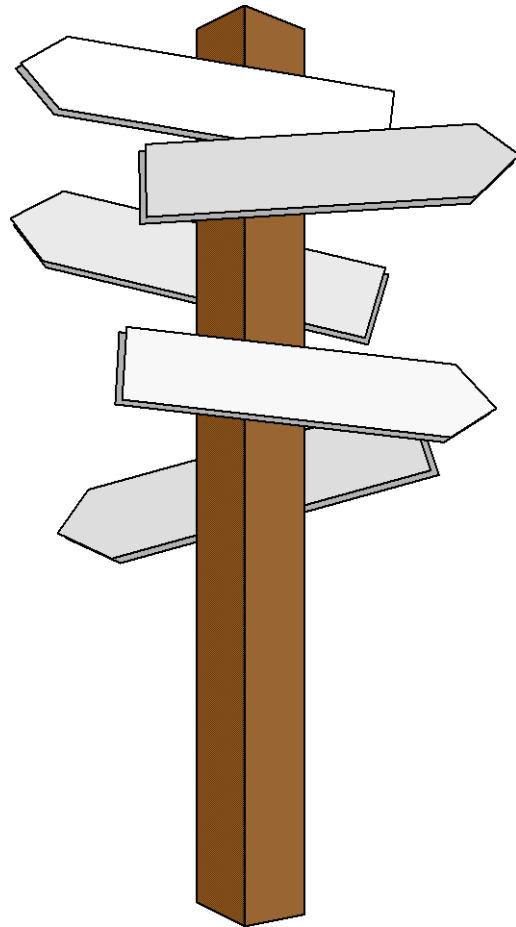


- Einfügen und Löschen erfolgt wie bei B-Bäumen.
- Zusätzliche Regel: Enthält eine „kleiner“ Binärbaum 3 Elemente, dann wird das mittlere Element automatisch zur Wurzel.
  
- Die binären Suchbäume in der Java-Bibliothek sind Rot-Schwarz-Bäume
  - **Z.B. `java.util.TreeMap`, `java.util.TreeSet`.**



- Das Laufzeitverhalten ist ähnlich wie bei AVL-Bäumen.
- Die Höhe von Rot-Schwarz-Bäumen ist etwa 15-30% größer als bei AVL-Bäumen.
  - **AVL-Bäume sind also etwas kompakter.**
  - **Dadurch verringert sich bei Rot-Schwarz-Bäumen die Zeit zum Einfügen/Löschen leicht.**
  - **Die Zeit zum Suchen erhöht sich leicht.**
- Trotz des sehr ähnlichen Verhaltens findet man Rot-Schwarz-Bäume weit häufiger als AVL-Bäume.
  - **AVL-Bäume sind meist Lehrbeispiel.**

- Ein Vorteil von Suchbäumen ist, dass durch einen In-Order-Durchlauf (siehe späteres Kapitel) die Daten leicht in einer **sortierten Reihenfolge** durchlaufen werden können.
  - **Mit einigen Modifikationen könnte man einen binären Suchbaum auch als **Liste** verwenden:**
  - **Lesen, Schreiben, Löschen und Einfügen mit  **$O(\log n)$** .**
  - **Dies ist aber unüblich.**
- Dies ist mit Hash-Tabellen nicht möglich.



## Vorlesung 9

### Heaps Heap-Sort

- Auch **Vorrangwarteschlange** oder **Priority Queue** genannt.
- Eine Warteschlange, deren Elemente einen Schlüssel (Priorität) besitzen.
- Wichtige Operationen bei Prioritätswarteschlangen:
  - Element in Schlange einfügen
  - Element mit der höchsten Priorität entnehmen.
    - Dies ist gewöhnlich das Element mit dem kleinsten Schlüssel,
    - Manchmal ist es auch das Element mit dem größten Schlüssel.

- Ereignissimulation
  - **Die Schlüssel sind die Zeitpunkte von Ereignissen, die in chronologischer Reihenfolge zu verarbeiten sind.**
- Verteilung der Rechenzeit auf mehrere Prozesse
- Graphalgorithmen
  - **Dijkstra, Vorlesung 15**
- Sortierverfahren
  - **Alle Elemente in Prioritätswarteschlange einfügen**
  - **Nach der Reihe die größten Elemente entnehmen**
  - **Heap-Sort**

- Wartezimmer eines Arztes
- Reihenfolge des Aufrufs wird durch Prioritätswarteschlange bestimmt.
- Die Priorität wird ermittelt aufgrund:
  - **Ankunftszeit**
  - **Termin / kein Termin**
  - **Privatpatient / Kassenpatient**
  - **Notfall / kein Notfall**

- Einfache Implementierung mit Array (Feld).
- Effiziente Implementierung mit AVL-Baum.
- Andere effiziente Implementierung: Partiell geordneter Baum (**Heap**).

- Kein Interface
- Nur eine einzige Implementation einer Prioritätswarteschlange: `java.util.PriorityQueue`
  - (wird auch in `java.util.concurrent.PriorityBlockingQueue` benutzt).
- Die zugrundeliegende Datenstruktur ist ein Heap.



- Das Wort Heap (Halde) hat zwei Bedeutungen:
  1. **Heap**: Besonderer Speicherbereich, in dem Objekte und Klassen gespeichert werden.
  2. **Heap**: Datenstruktur zur effizienten Implementierung einer Prioritätswarteschlange.
- Beide Bedeutungen haben nichts miteinander zu tun. In folgenden Kapitel widmen wir uns ausschließlich der zweiten Bedeutung.
- Dabei betrachten wir ausschließlich den **binären Heap**. Es gibt z.B. noch den Binominal-Heap und den Fibonacci-Heap.

## Einführung:

<http://www.matheprisma.uni-wuppertal.de/Module/BinSuch/index.html>

Kapitel Heap (1+2)

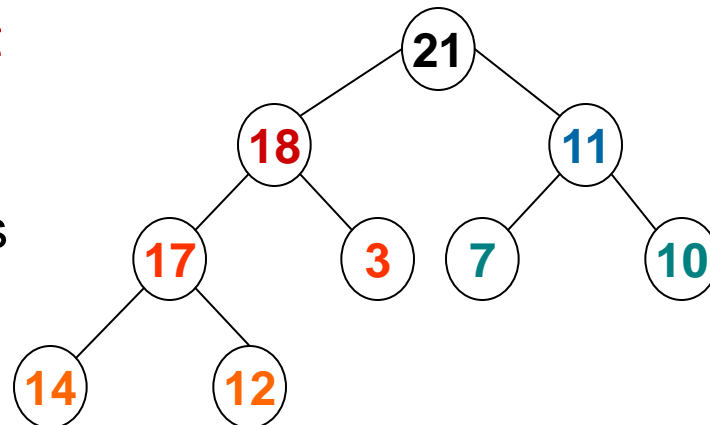
## Definition:

Ein *Heap* ist ein Binärbaum mit folgenden Eigenschaften:

- Er ist links-vollständig
- Die Kinder eines Knotens sind höchstens so groß wie der Knoten selbst.

⇒ das größte Element befindet sich an der Wurzel des Heaps

Achtung: In der Literatur gibt es auch die umgekehrte Definition



Ein Heap ist also ein Binärbaum mit den beiden Eigenschaften:

**Form :**



(linksvollständig)

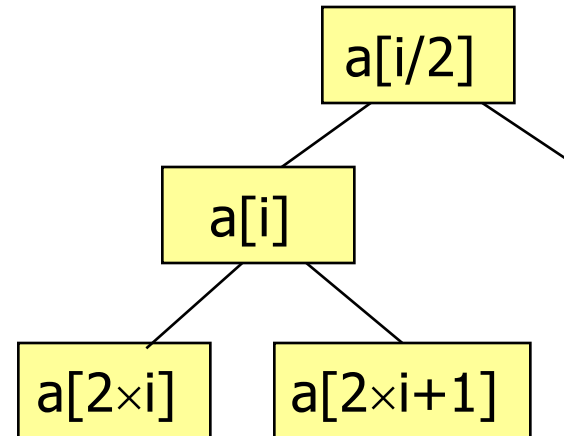
und

**Ordnung:**

Entlang jedes Pfades von einem Knoten zur Wurzel sind die Knoteninhalte aufsteigend sortiert.

Vorteil der Linksvollständigkeit: Feld-Einbettung leicht möglich.

Vater und Söhne zu  $a[i]$ :



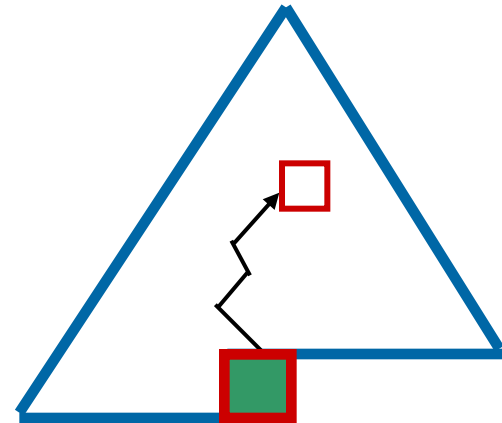
**Anmerkung:** In Java werden Felder beginnend mit 0 indiziert.

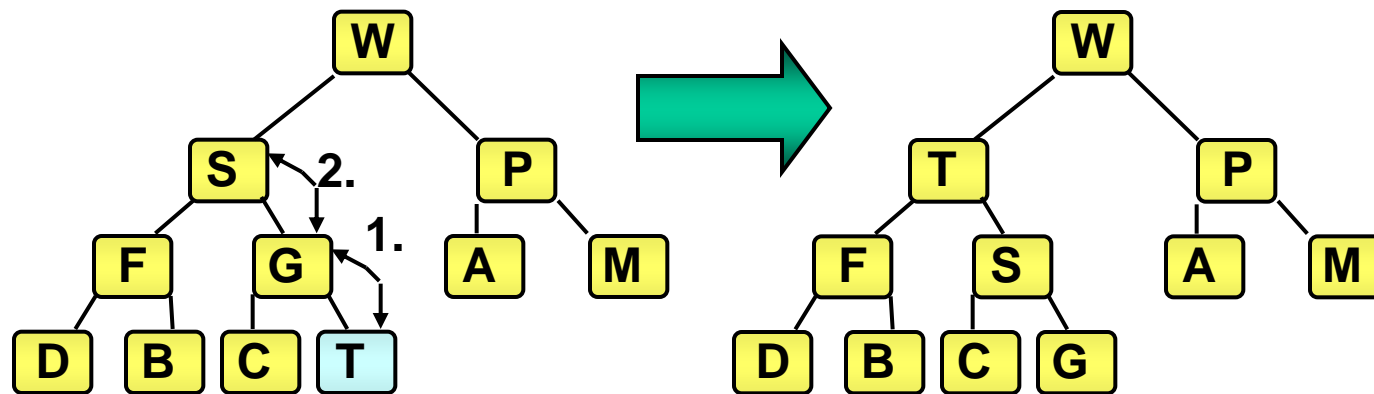
⇒ entweder:

- der  $k$ -te Knoten wird in dem Feld-Element  $a[k-1]$  gespeichert...
- oder Feld-Index 0 bleibt unbenutzt!

## Element einfügen: upHeap(„swim“)

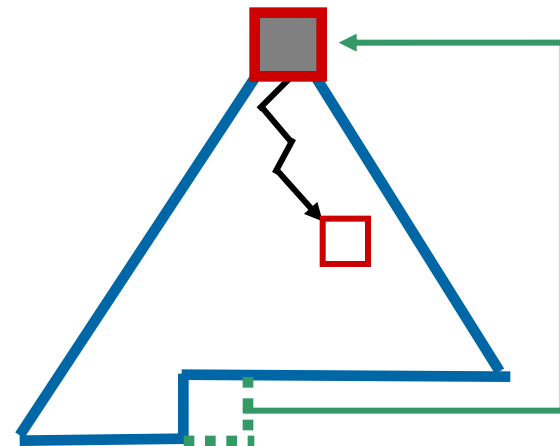
- neues Element in Heap einfügen geht (wegen Linksvollständigkeit) nur an **genau einer Position**.
- Ordnungseigenschaft kann dadurch verletzt werden.
- Algorithmus zum Wiederherstellen der Ordnung: **upHeap**.
- Idee:
  - Knoten mit Vaterknoten vergleichen und ggf. vertauschen.
  - Dies setzt sich nach oben fort (notfalls bis zur Wurzel).





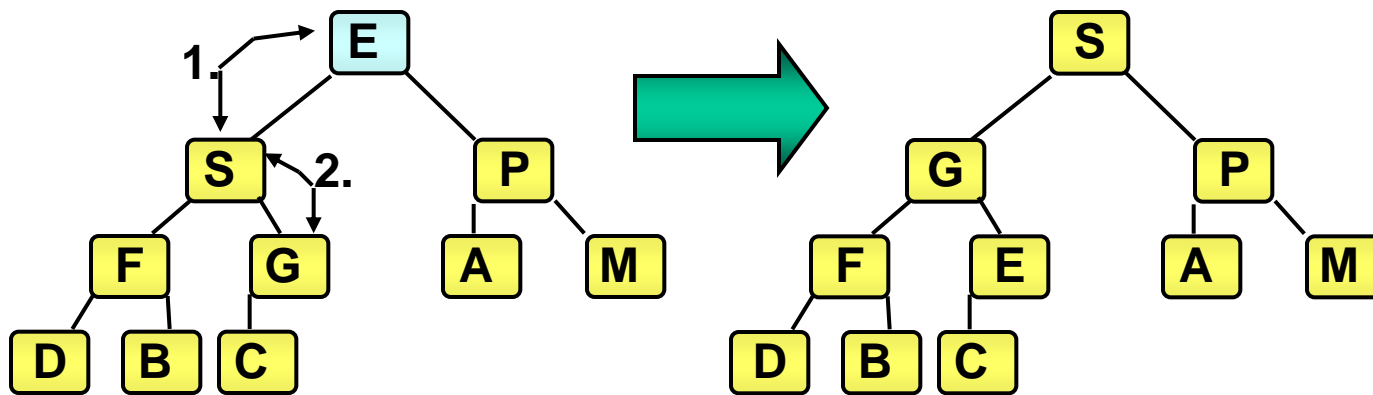
## Wurzel entfernen: downHeap („sink“)

- Nach Entfernen der Wurzel eines Heaps wird am weitesten rechts stehender Blattknoten in der untersten Ebene die neue Wurzel.  
⇒ Form wiederhergestellt.
- Ordnungseigenschaft kann dadurch verletzt werden.
- Algorithmus zum Wiederherstellen der Ordnung: **downHeap**.
- Idee:
  - Neues Wurzel-Element wandert nach unten („versickert“).
  - Dabei wird es jeweils falls nötig mit dem **größeren** Sohn vertauscht.





# downHeap: Beispiel (1)



- Einfache Methode:
  - **Alle Elemente der Reihe nach einfügen und aufsteigen lassen.**
- Falls ein bestehendes (unsortiertes) Array in einen Heap umgewandelt werden soll, ist das **Bottom-up-Verfahren** effizienter:
  - **Die erste Hälfte der Elemente wird nach unten versickert (downHeap).**
  - **Auf dieses Verfahren wird in der Vorlesung nicht weiter eingegangen.**

- sink folgt dem Weg von der Wurzel eines links-vollständigen Binärbaums bis maximal zu seinen Blättern, d.h. es werden höchstens

$$\lceil \log_2(n+1) \rceil$$

Knoten besucht. (minimale Höhe +1)  $\Rightarrow O(\log n)$

- Aufbau eines Heaps (mit Bottom-up) ruft sink für  $n/2$  Elemente auf, also:

$$T^{\text{av}}_{\text{heapcreate}}(n) = T^{\text{w}}_{\text{heapcreate}}(n) =$$

$$n/2 \cdot \lceil \log_2(n+1) \rceil \in O(n \log n)$$

Der Heap ist die Grundlage für ein bekanntes Sortierverfahren

- Zu Beginn: Unsortiertes Feld
- **Phase 1:** Aufbau des Heaps
  - **Einfaches Verfahren: Alle Elemente werden nacheinander eingefügt**
  - **Schneller: Bottom-up-Verfahren (nicht in Vorlesung)**
  - **Resultat ist ein Heap, der in ein Feld eingebettet ist.**
- **Phase 2:** Der Heap wird geleert, indem immer wieder die Wurzel (d.h. das größte Element) entfernt wird.
  - **Die Elemente werden in absteigender Reihenfolge entfernt.**
  - **Der Heap schrumpft immer weiter.**

- Zunächst wird das Feld komplett in einen Heap umgewandelt:



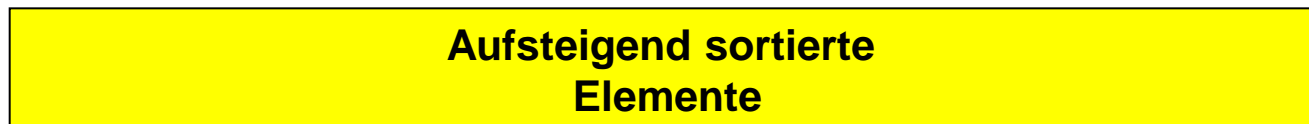
- Wenn die Wurzel entfernt wird, schrumpft der Heap um 1 Element und das letzte Feldelement gehört nicht mehr zum Heap.
- Hier kann das 1. sortierte Element untergebracht werden



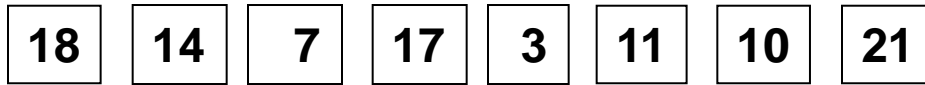
- Wenn weitere Male die Wurzel entfernt wird, schrumpft der Heap immer weiter. Die freiwerdenden Stellen werden mit den entnommenen Werten besetzt, die jetzt aufsteigend sortiert sind.



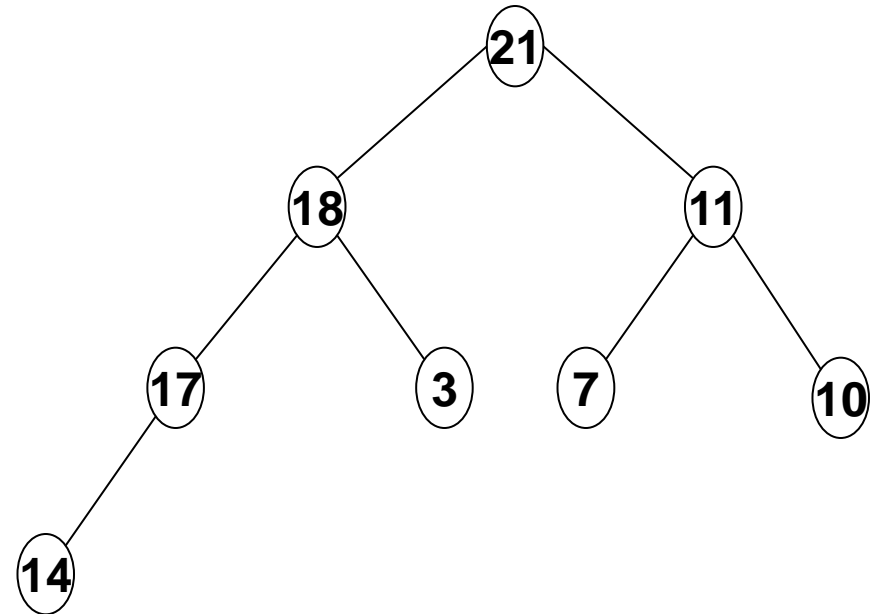
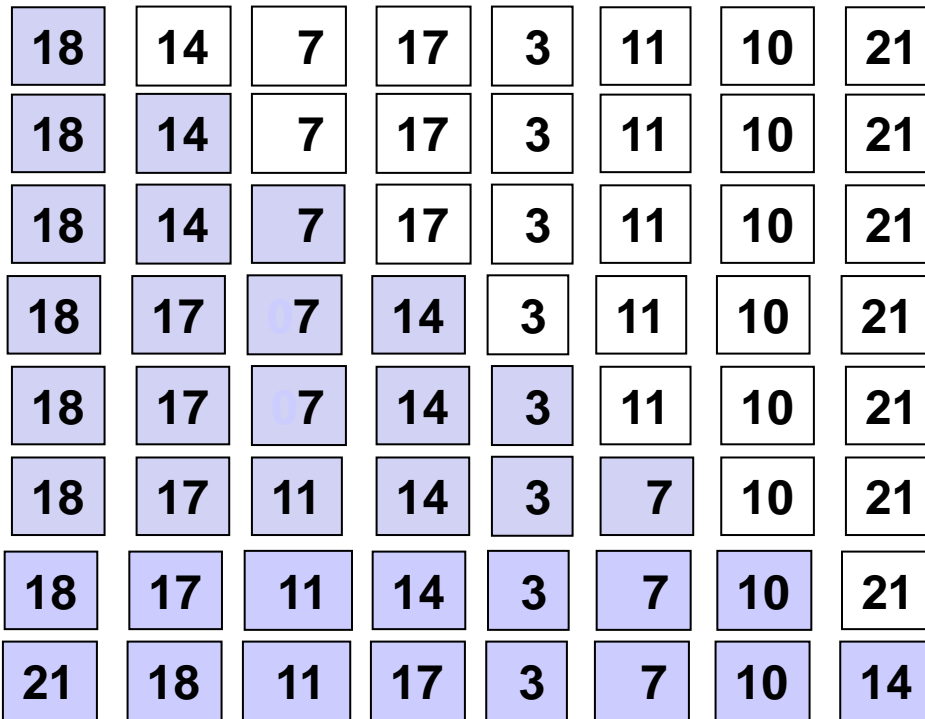
- Zuletzt ist der Heap auf 0 geschrumpft und das Feld ist sortiert.



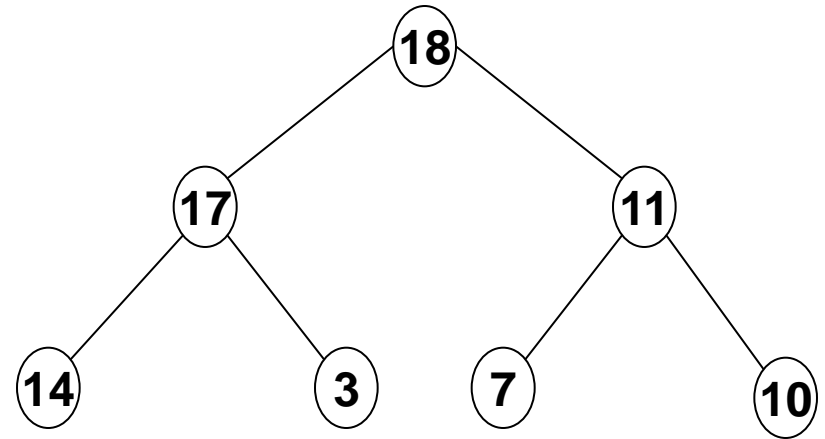
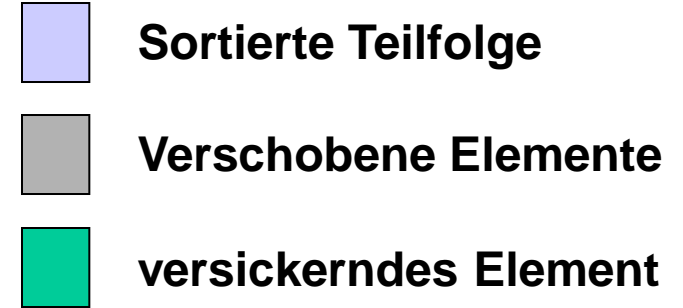
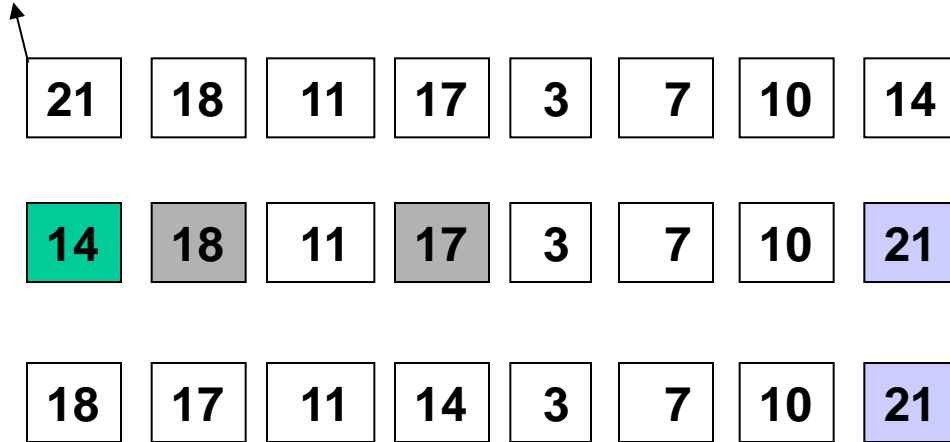
# Beispiel für Phase 1 von HeapSort



↓ Heap aufbauen

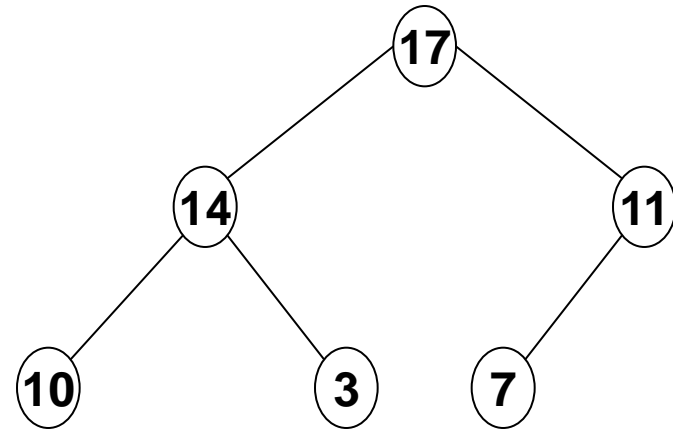
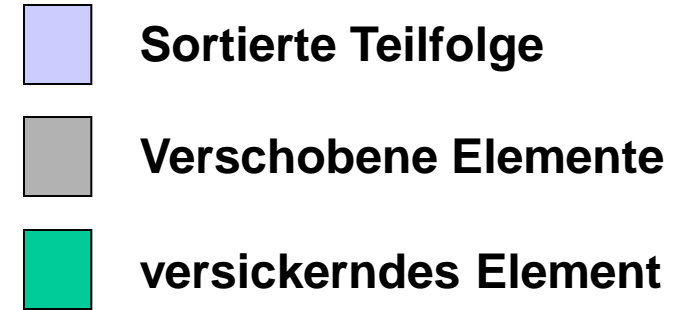
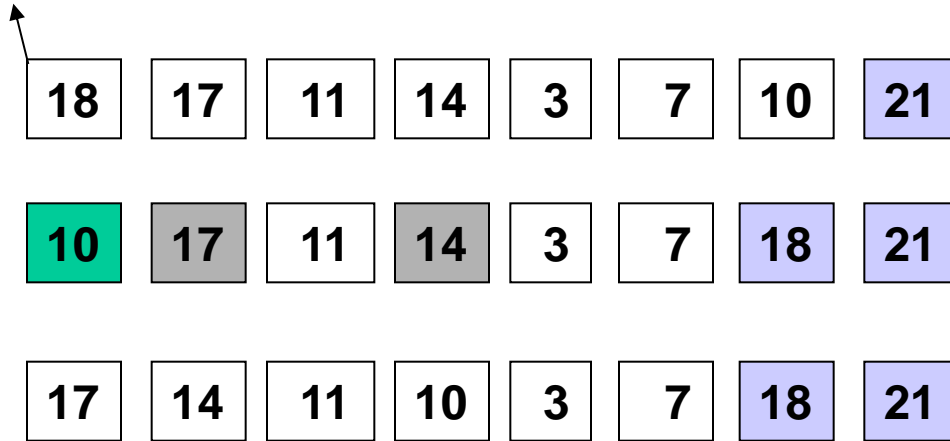


# Beispiel für Phase 2 von HeapSort

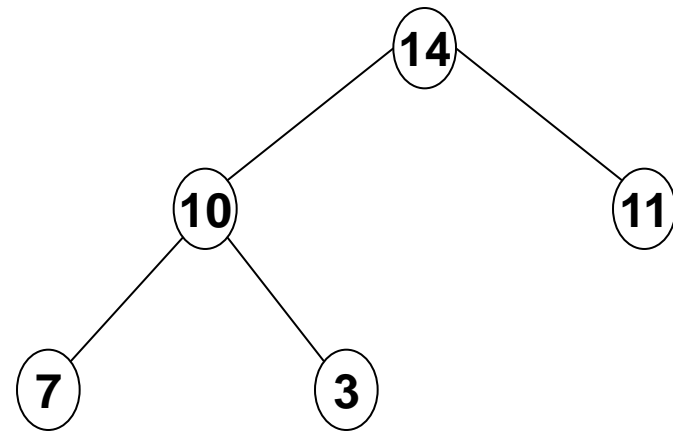
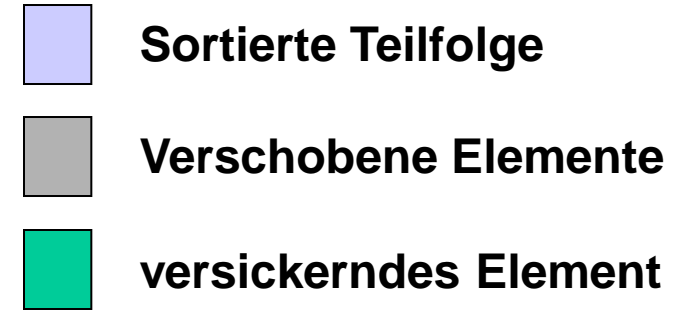
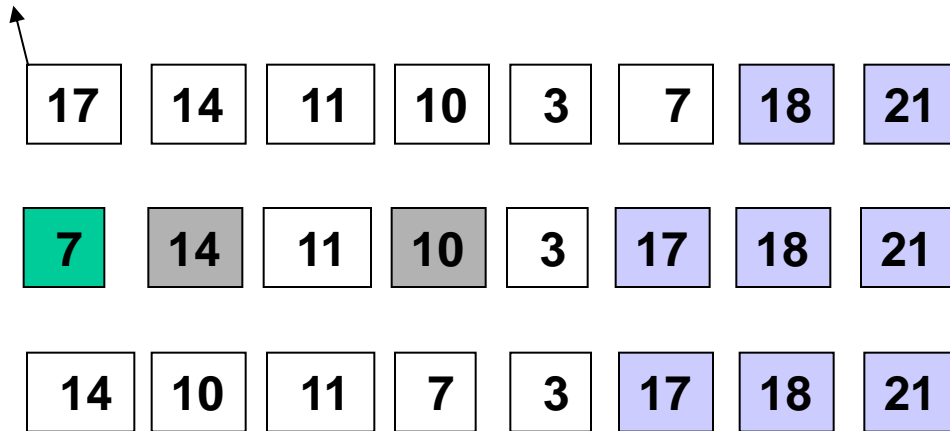




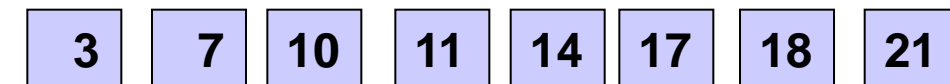
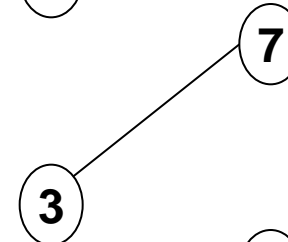
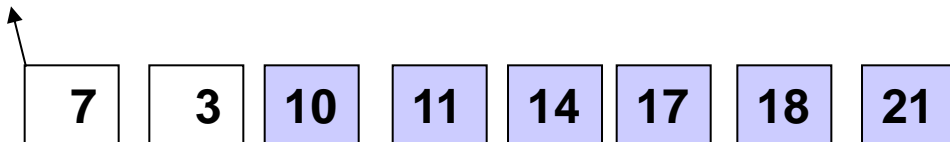
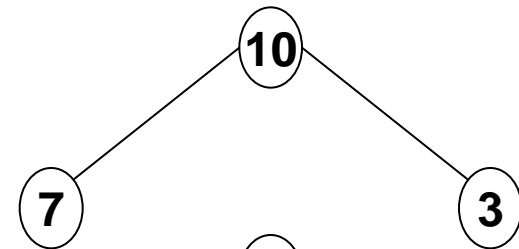
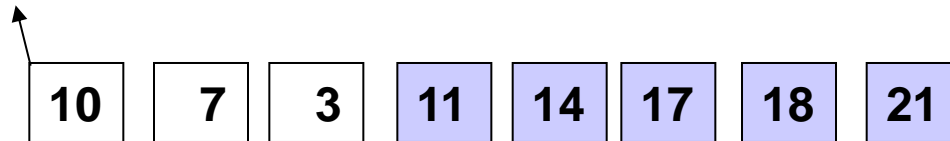
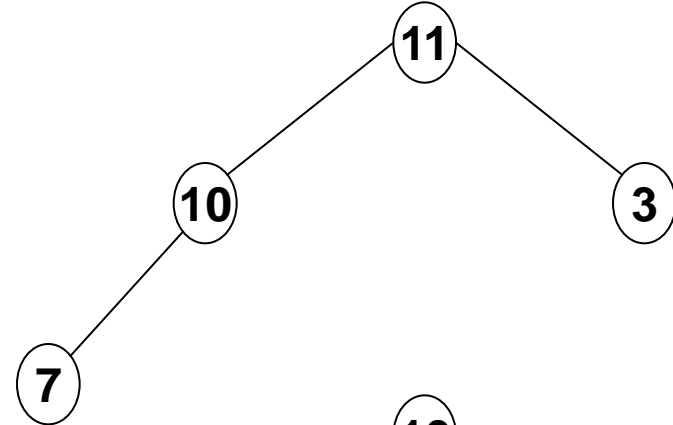
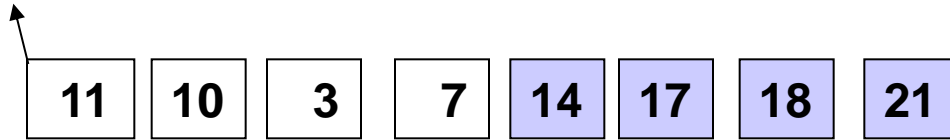
# Beispiel für Phase 2 von HeapSort

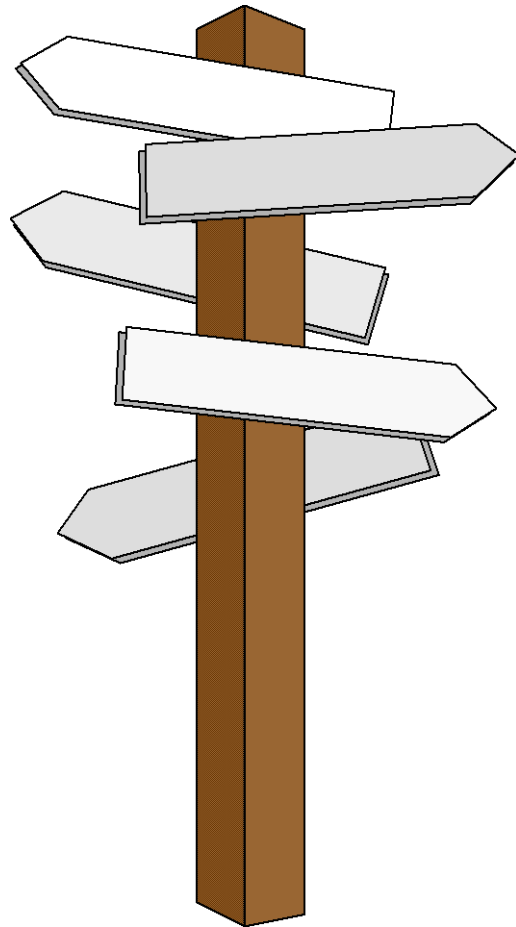


## Beispiel für Phase 2 von HeapSort



# Beispiel für Phase 2 von HeapSort

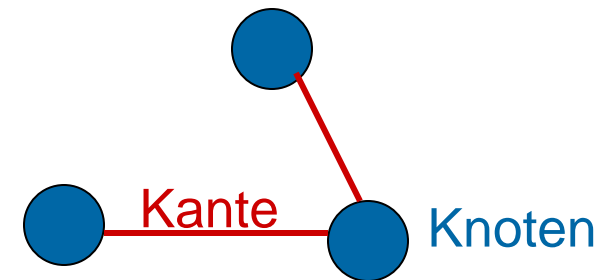




## Vorlesung 10

**Graphen**  
**Breitensuche**  
**Bipartite Graphen**

- sehr allgemeine Datenstruktur
- bestehen aus Knoten und Kanten, die die Knoten verbinden
- Modellierung vieler praktischer Probleme mit Graphen
  - **Z.B. Telefonnetze, Flugpläne, ...**
- viele Algorithmen für Graphen bekannt



- Ein *gerichteter Graph*  $G = (V, E)$  besteht
  - aus einer endlichen, nicht leeren Menge  $V = \{v_1, \dots, v_n\}$  von **Knoten** (Engl.: „vertices“) und
  - einer Menge  $E \subseteq V \times V$  von geordneten Paaren  $e = (u, v)$ , den **Kanten** (Engl.: „edges“).  
Jede Kante  $(u, v) \in E$  hat einen Anfangsknoten  $u$  und einen Endknoten  $v$  und damit eine Richtung von  $u$  nach  $v$  ( $u = v$  ist möglich).
- Graphische Darstellung:

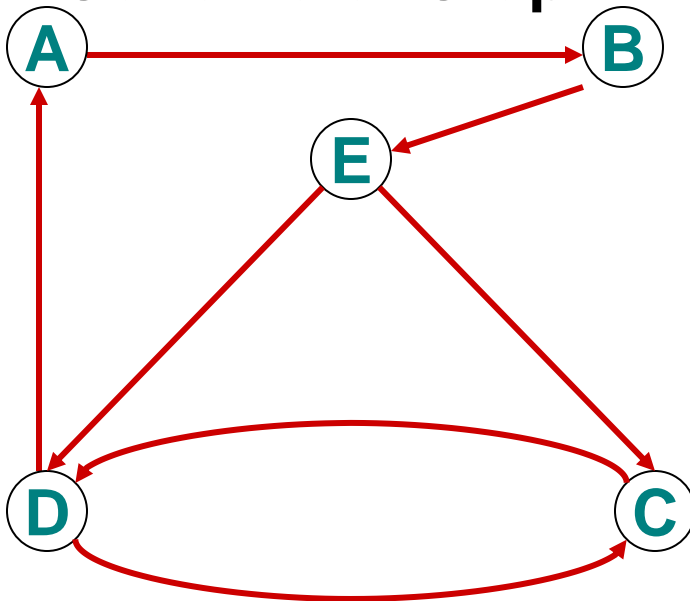


- Ein *ungerichteter Graph*  $G = (V, E)$  ist ein gerichteter Graph, bei dem die Relation  $E$  symmetrisch ist:

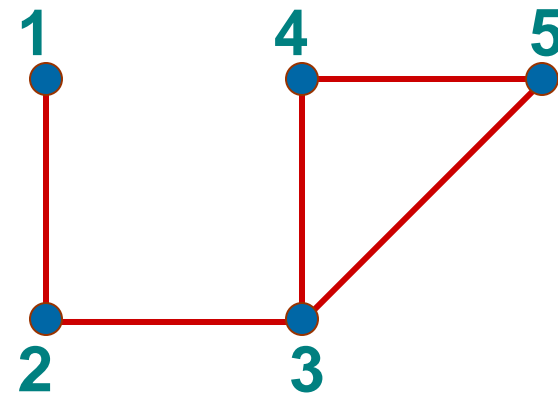
$$(u, v) \in E \Leftrightarrow (v, u) \in E$$

- Graphische Darstellung: 

## Gerichteter Graph

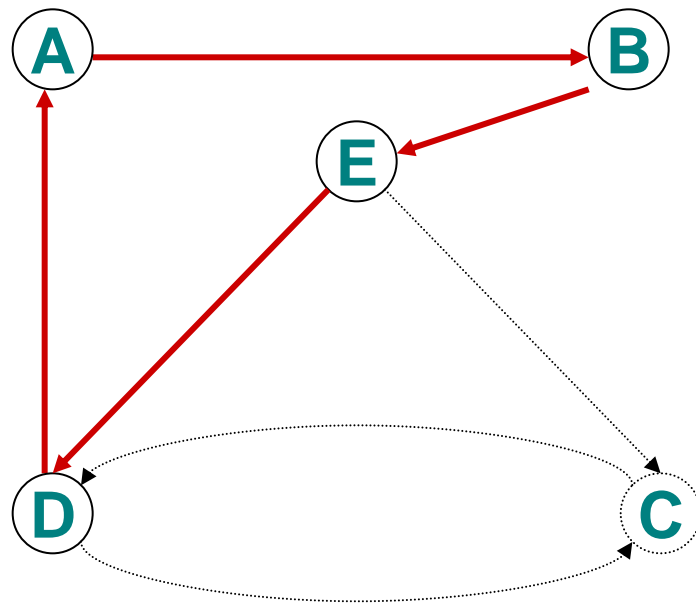


## Ungerichteter Graph





- $G' = (V', E')$  heißt **Teilgraph** von  $G = (V, E)$ , wenn gilt:  
 $V' \subseteq V$  und  $E' \subseteq E$ .



- Zwei Knoten heißen **adjazent** (benachbart), wenn eine Kante sie verbindet.
- Bei einem gerichteten Graphen mit der Kante  $u \rightarrow v$  ist **u** **Vorgänger** von **v**; **v** ist **Nachfolger** von **u**.

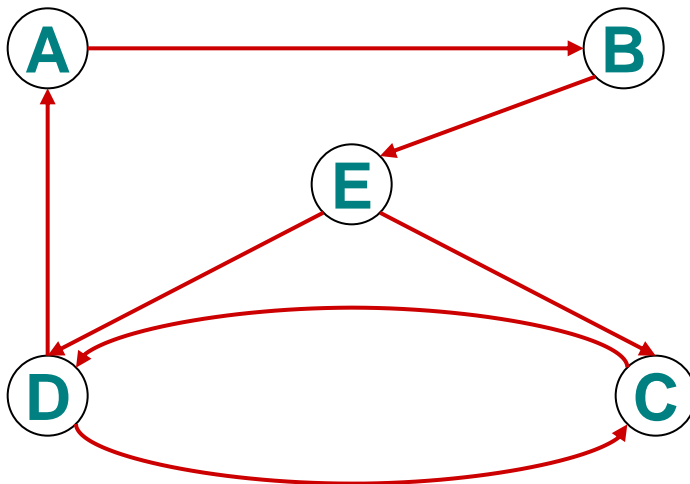


Sei  $G = (V, E)$  ein (gerichteter oder ungerichteter) Graph.

- Eine Folge von Knoten  $W := (v_1, v_2, v_3, \dots, v_n)$  heißt **Weg** (oder **Pfad**) in  $G$ , falls gilt:

$$\forall i=1, \dots, n-1 : (v_i, v_{i+1}) \in E$$

(also eine Folge von „zusammenhängenden“ Kanten)



Beispielwege:

(A, B)

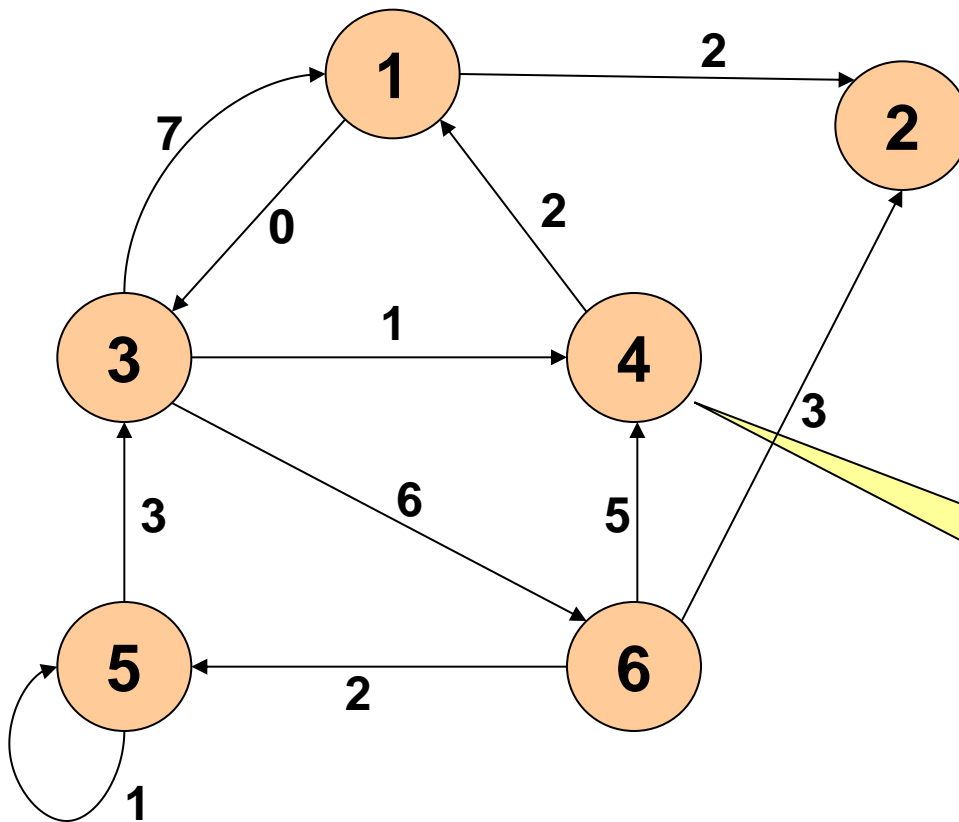
(B, E, D, C)

(D, C, D, C, D, C)

(A, B, E, D, A)

- $v_1 =: \alpha(W)$  heißt **Anfangsknoten** des Weges  $W$
- $v_n =: \omega(W)$  heißt **Endknoten** des Weges  $W$
- $\forall v_i \in V: (v_i)$  heißt **trivialer** Weg und ist stets ein Weg in  $G$ .
- Die **Länge eines Weges** ist  $l(W) := n-1$ , falls  $n$  Knoten auf diesem Weg besucht werden.
- Ein Weg heißt **einfacher Weg**, wenn kein Knoten (außer eventuell dem ersten/letzten – siehe Zykel) mehr als einmal vorkommt.
- Ein **Zykel** (Kreis) ist ein (nicht-trivialer) einfacher Weg mit der Bedingung  $\alpha(W) = \omega(W)$ .

- Ein Graph heißt **gewichtet** (bewertet), wenn jeder Kante ein Wert als Gewicht zugeordnet ist (z.B. Transportkosten, Entfernung, etc.).



Um von 3 nach 1 zu kommen, ist der Weg über 4 billiger als der direkte Weg.

## Je nach Zielsetzung:

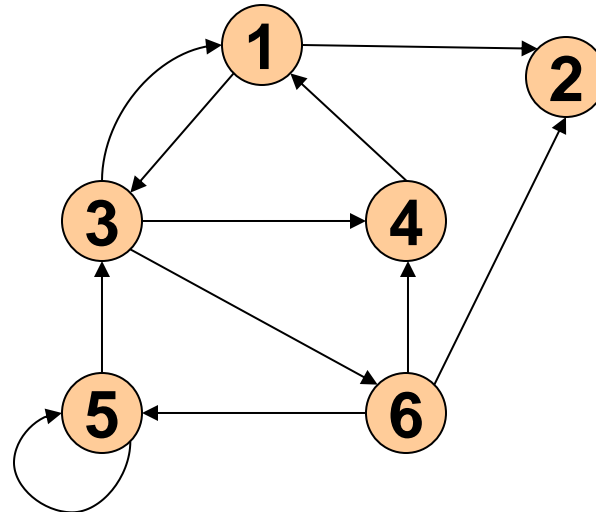
- **Kantenorientiert**

- Prinzip:
  - Index für Kanten
  - Für jede Kante speichern: Vorgängerknoten, Nachfolgerknoten, ggfs.: Markierung, Gewicht
- Meist statische Darstellung, z.B. Kantenliste

- **Knotenorientiert**

- gebräuchlicher als kantenorientiert
- in vielen Ausprägungen, z.B.:
  - Knotenliste
  - Adjazenzmatrix
  - Adjazenzliste

# Kantenliste für gerichteten Graphen



Platzbedarf:  $2+2 \cdot |E|$

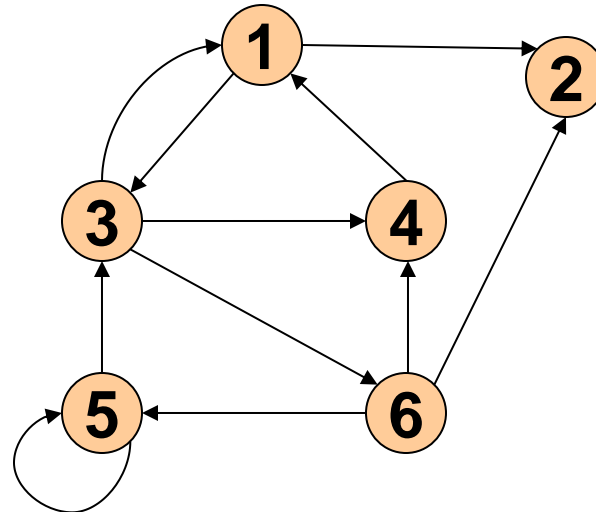
Anzahl der  
Knoten

Anzahl der  
Kanten

Für jede Kante:  
Vorgänger und Nachfolger

6, 11, 1, 2, 1, 3, 3, 1, 4, 1, 3, 4, 3, 6, 5, 3, 5, 5, 6, 5, 6, 2, 6, 4

# Knotenliste für gerichteten Graphen



Platzbedarf:  $2+|V|+|E|$   
(normalerweise weniger  
als Kantenliste)

Anzahl der  
Knoten

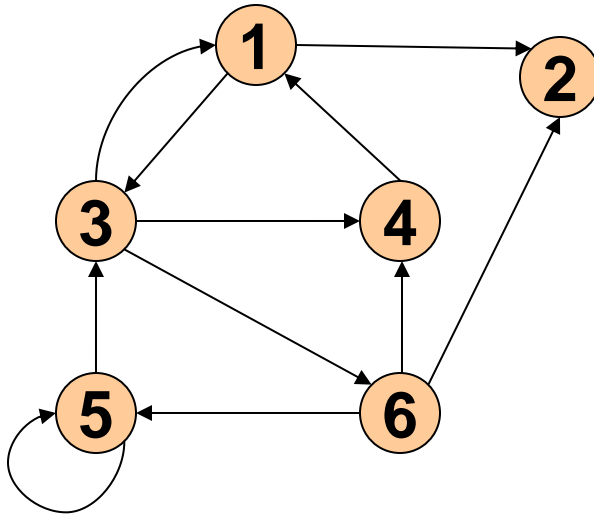
Anzahl der  
Kanten

Für jeden Knoten:  
**Ausgangsgrad** und  
und Liste der Nachfolger

6, 11, 2, 2, 3, 0, 3, 1, 4, 6, 1, 1, 2, 3, 5, 3, 2, 4, 5



# Adjazenzmatrix für gerichteten Graphen



Platzbedarf:  $|V|^2$

Adjazenzmatrix für ungerichtete Graphen ist symmetrisch

Gewichte können anstelle boolescher Werte gespeichert werden

$$A_{ij} = \begin{cases} true, & falls (i, j) \in E \\ false, & andernfalls \end{cases}$$

nach

$$\begin{matrix} \text{v} \\ \text{o} \\ \text{n} \end{matrix} \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

## Vorteile:

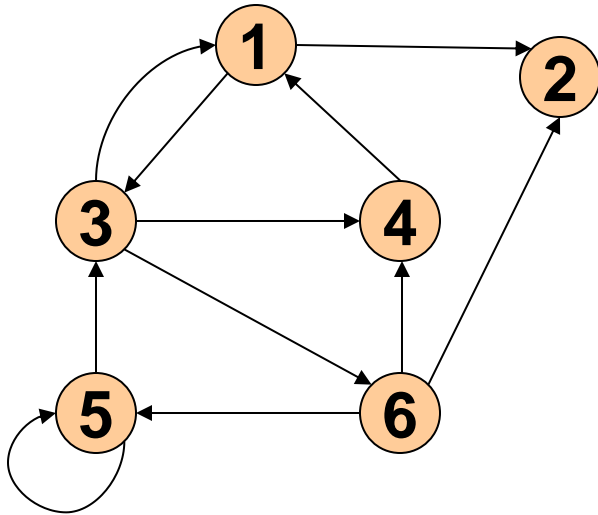
- + Entscheidung, ob  $(i,j) \in E$ , in Zeit  $O(1)$
- + erweiterbar für Kantenmarkierung und Gewichte

## Nachteile:

- Platzbedarf stets  $O(|V|^2) \Rightarrow$  ineffizient falls  $|E| \ll |V|^2$
- Initialisierung benötigt Zeit  $O(|V|^2)$

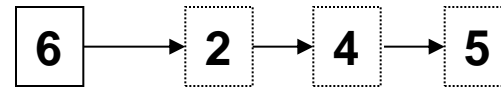
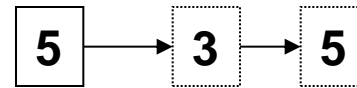
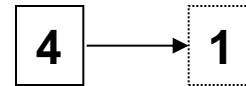
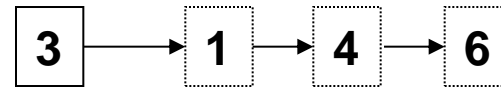
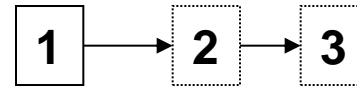
„viel  
kleiner  
als“

# Adjazenzliste für gerichteten Graphen



Platzbedarf:

- $|V|$  Listen
- $|E|$  Listenelemente



## Vorteil:

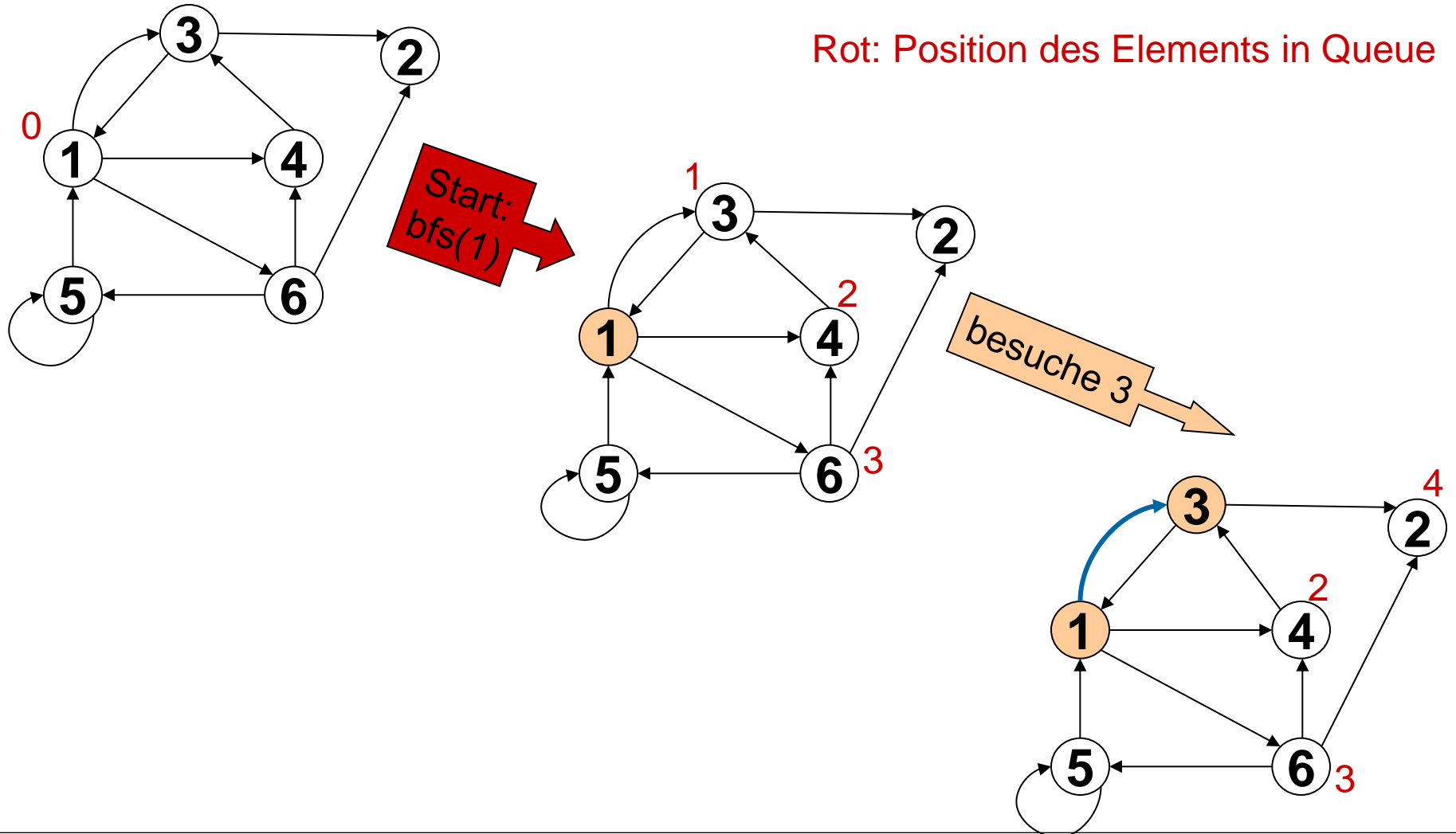
+ Geringer Platzbedarf von  $O(|V| + |E|)$

## Nachteil:

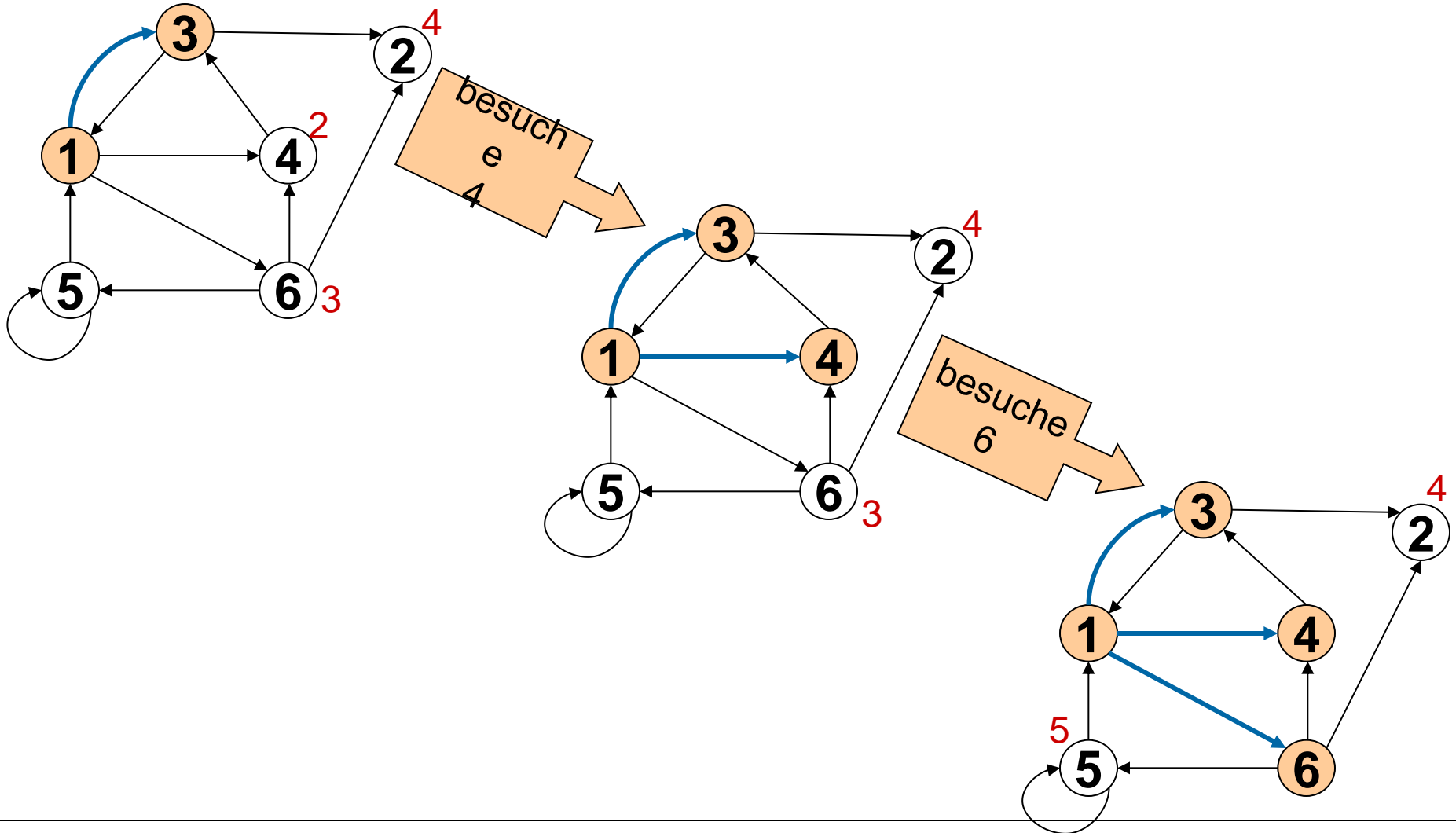
– Entscheidung, ob  $(i,j) \in E$  in Zeit  $O(|E|)$  im schlimmsten Fall

1. zunächst alle Knoten als „noch nicht besucht“ markieren
2. **Startpunkt**  $v$  wählen (und als „besucht“ markieren)
3. Jetzt:
  - i. alle von  $v$  aus **direkt** erreichbaren (noch nicht besuchten) Knoten „besuchen“
  - ii. alle von  $v$  aus **über zwei Kanten** erreichbaren Knoten „besuchen“
  - iii. ... **über  $n$  Kanten** ...Implementierung mit einer Queue.
4. wenn dann noch nicht alle Knoten besucht worden sind:  
einen unbesuchten Knoten als **neuen Startpunkt** wählen  
und von 3. beginnen

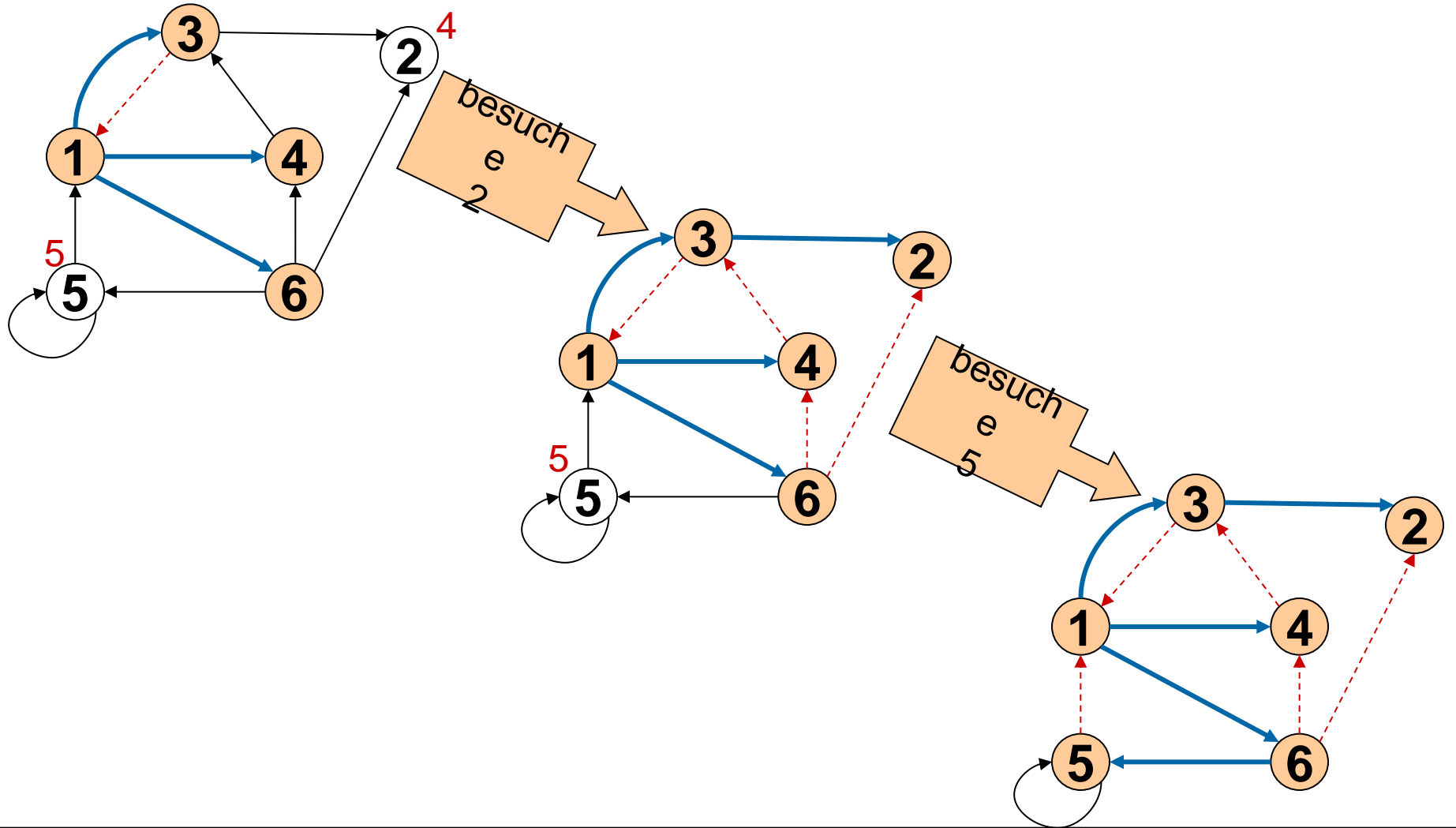
# Beispiel zu BFS (1)



# Beispiel zu BFS (2)



# Beispiel zu BFS (3)



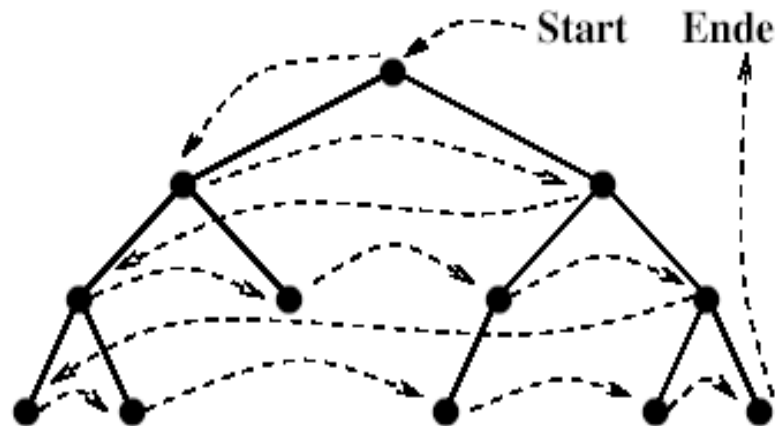


bfs(v)

|                                                                       |                         |
|-----------------------------------------------------------------------|-------------------------|
| Initialisiere leere queue                                             |                         |
| queue.put(v)                                                          |                         |
| Markiere Knoten 'v' als besucht                                       |                         |
| Solange queue nicht leer                                              |                         |
| u = queue.get()                                                       |                         |
| Für jeden direkten Nachfolger w von u<br>(es gibt Kante von u nach w) |                         |
| Falsch                                                                | 'w' schon besucht? Wahr |
| queue.put(w)                                                          |                         |
| Markiere Knoten 'w' als besucht                                       |                         |

## Baumdurchläufe: Level-Order

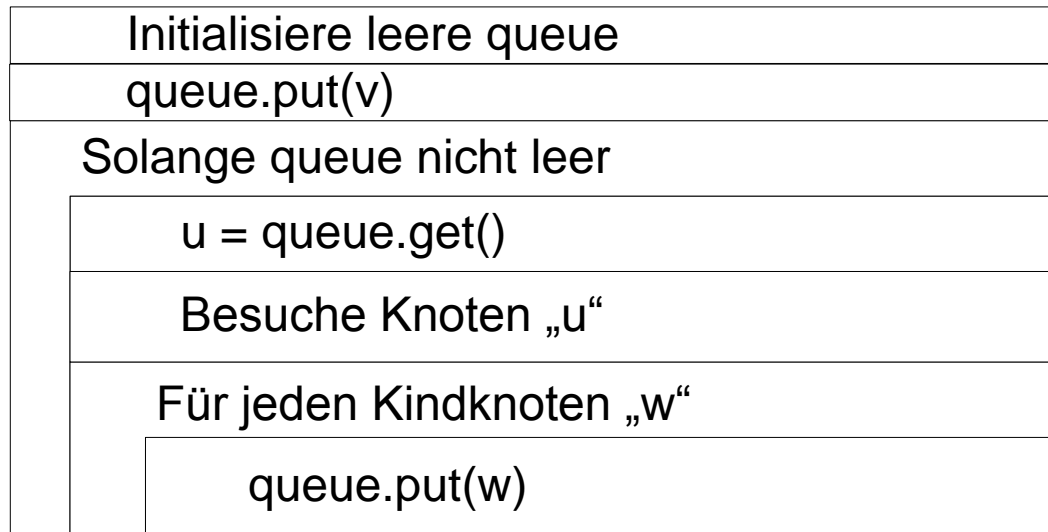
- Führt man die Breitensuche bei Bäumen aus, stellt man fest:
  - Es ist nicht nötig, die Knoten zu markieren
  - Der Baum wird Ebene für Ebene durchlaufen:
    - Angefangen bei der Wurzel
    - Wenn man die Kindknoten jeweils von links nach rechts besucht, wird jede Ebene von links nach rechts durchlaufen
- Diesen Durchlauf nennt man **Level-Order-Durchlauf**

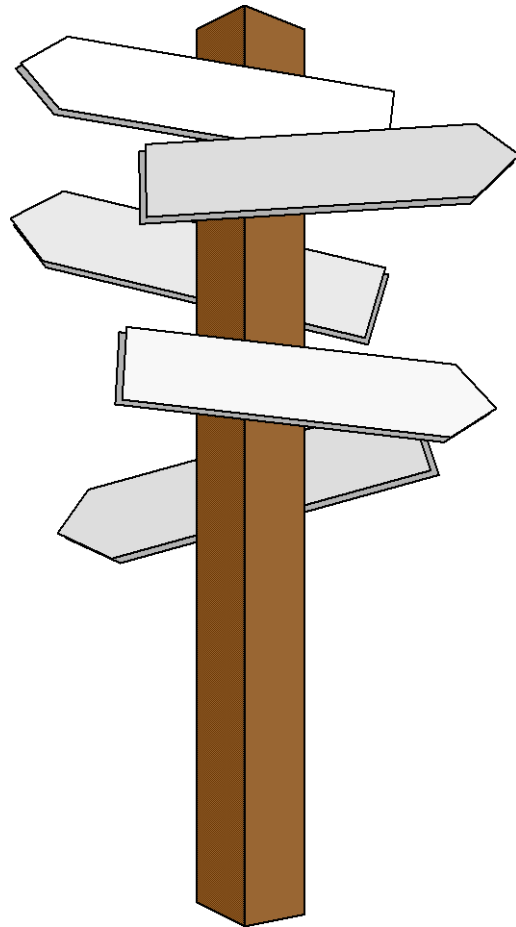


Zu Beginn muss dafür gesorgt sein, dass alle Knoten als „noch nicht besucht“ markiert sind:

|                                     |
|-------------------------------------|
| Für alle Knoten 'v'                 |
| markiere Knoten 'v' als 'unbesucht' |

Durchlaufe Levelorder(v)



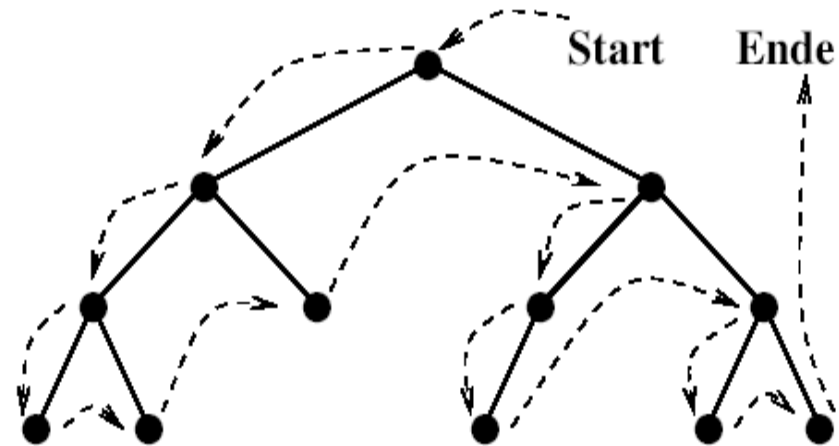


## Vorlesung 11

### Tiefensuche Baumdurchlauf (2)

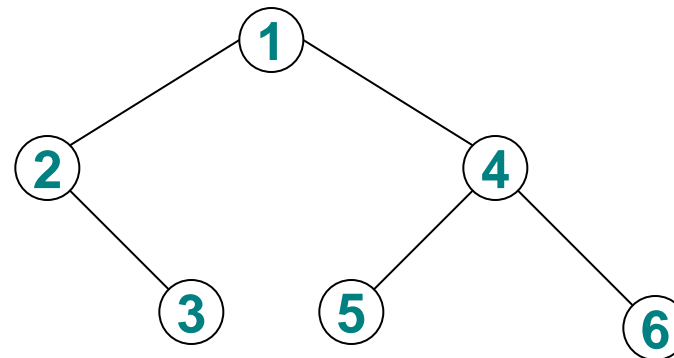
## Baumdurchläufe: Preorder

1. Betrachte die Wurzel (und führe eine Operation auf ihr aus)
2. Durchlaufe linken Teilbaum
3. Durchlaufe rechten Teilbaum



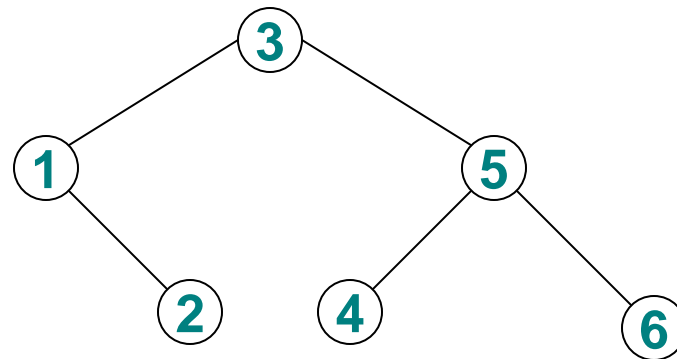
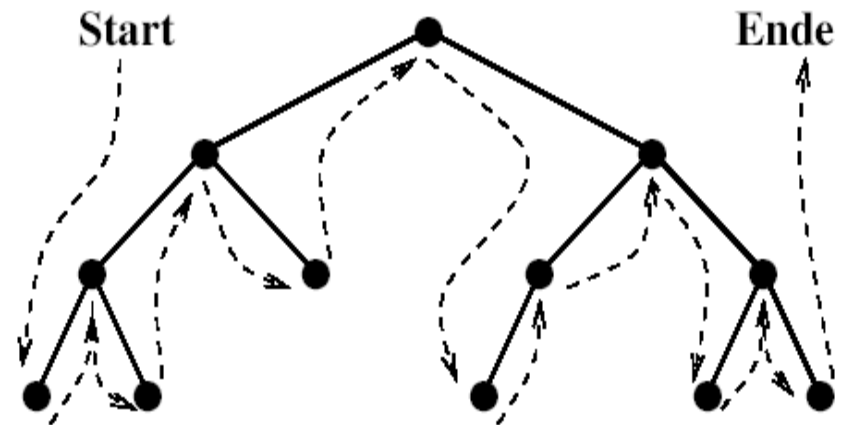
Preorder (wurzel)

|                         |
|-------------------------|
| Besuche Wurzel          |
| Preorder (linker Sohn)  |
| Preorder (rechter Sohn) |



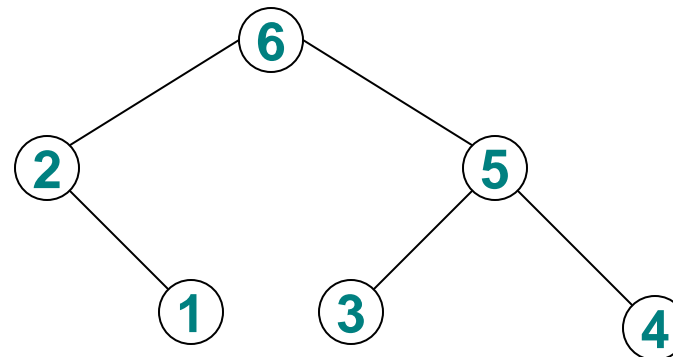
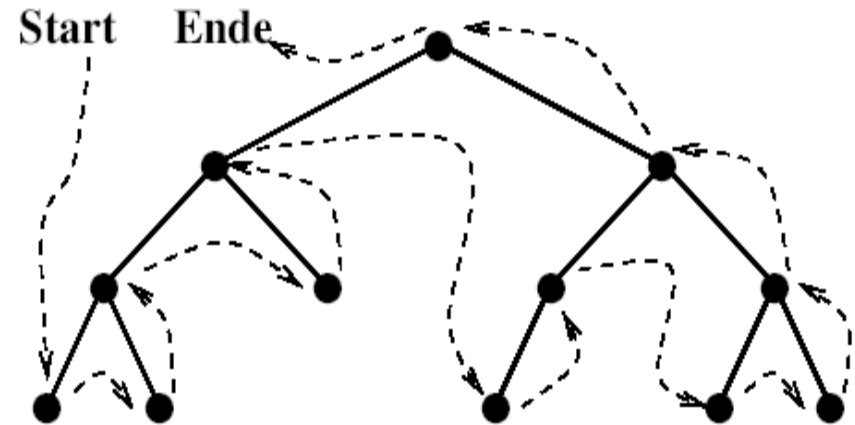
## Baumdurchläufe: Inorder

1. Durchlaufe linken Teilbaum
2. Betrachte die Wurzel (und führe eine Operation auf ihr aus)
3. Durchlaufe rechten Teilbaum

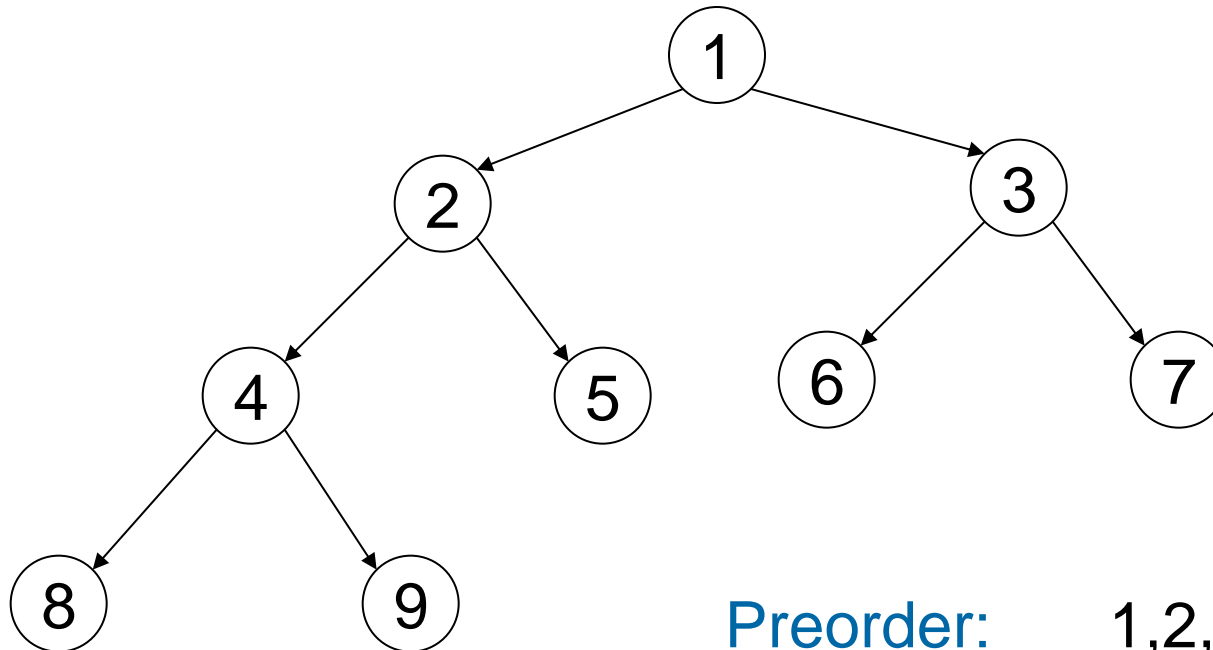


## Baumdurchläufe: Postorder

1. Durchlaufe linken Teilbaum
2. Durchlaufe rechten Teilbaum
3. Betrachte die Wurzel (und führe eine Operation auf ihr aus)





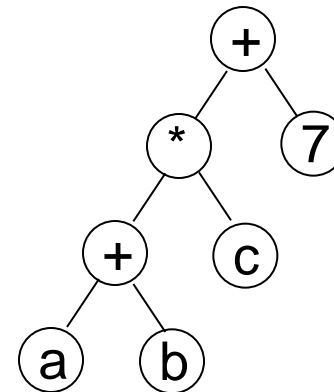


**Preorder:** 1,2,4,8,9,5,3,6,7  
**Inorder:** 8,4,9,2,5,1,6,3,7  
**Postorder:** 8,9,4,5,2,6,7,3,1  
**Levelorder:** 1,2,3,4,5,6,7,8,9

- $(a + b) * c + 7$

Prefix-Notation:  $+ * + a b c 7$

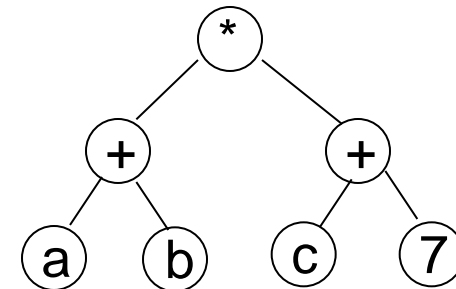
Postfix-Notation:  $a b + c * 7 +$



- $(a + b) * (c + 7)$

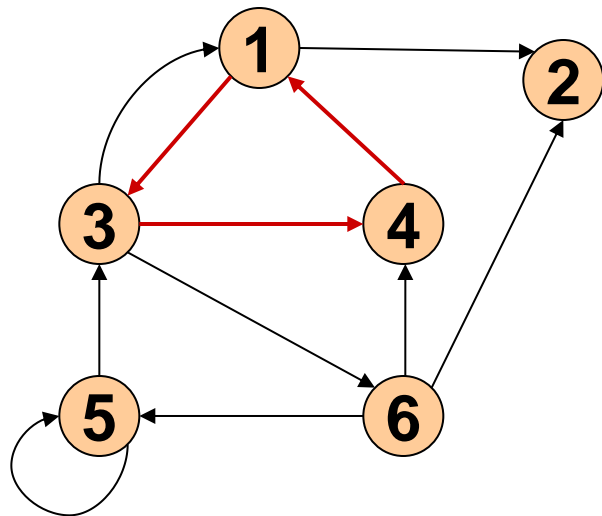
Prefix-Notation:  $* + a b + c 7$

Postfix-Notation:  $a b + c 7 + *$



**Bemerkung:** Bei der Infix-Notation sind i.allg. Klammern notwendig, um die Baumstruktur eindeutig wiedergeben zu können.

- Graphdurchläufe sind prinzipiell ähnlich wie Baumdurchläufe.
- Allerdings müssen die durchlaufenen Knoten markiert werden, damit man auf zyklischen Wegen nicht in Endlosschleifen gerät.

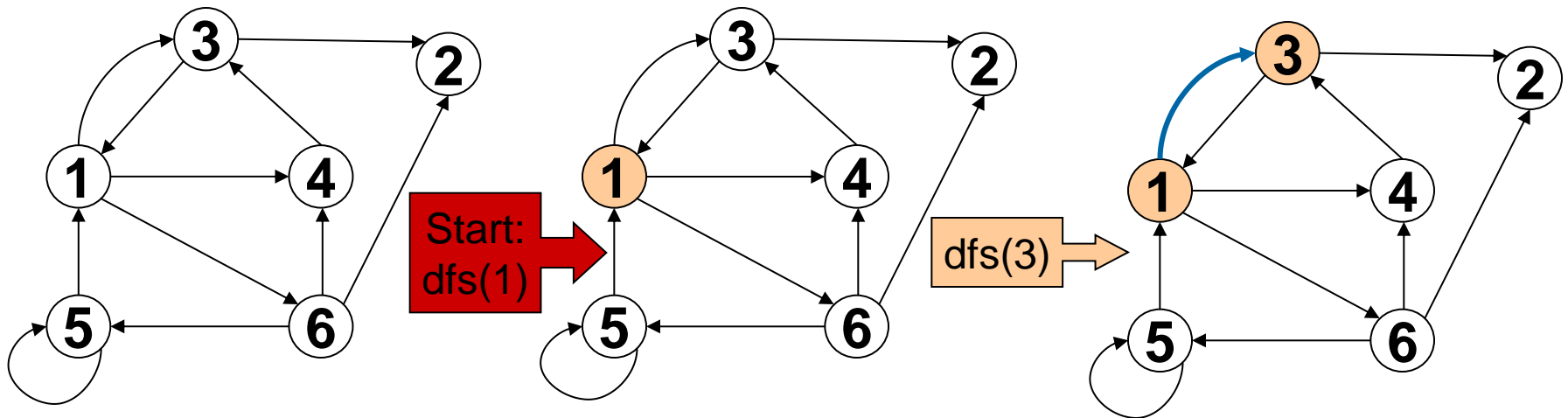


Pre-Order → **Tiefensuche**

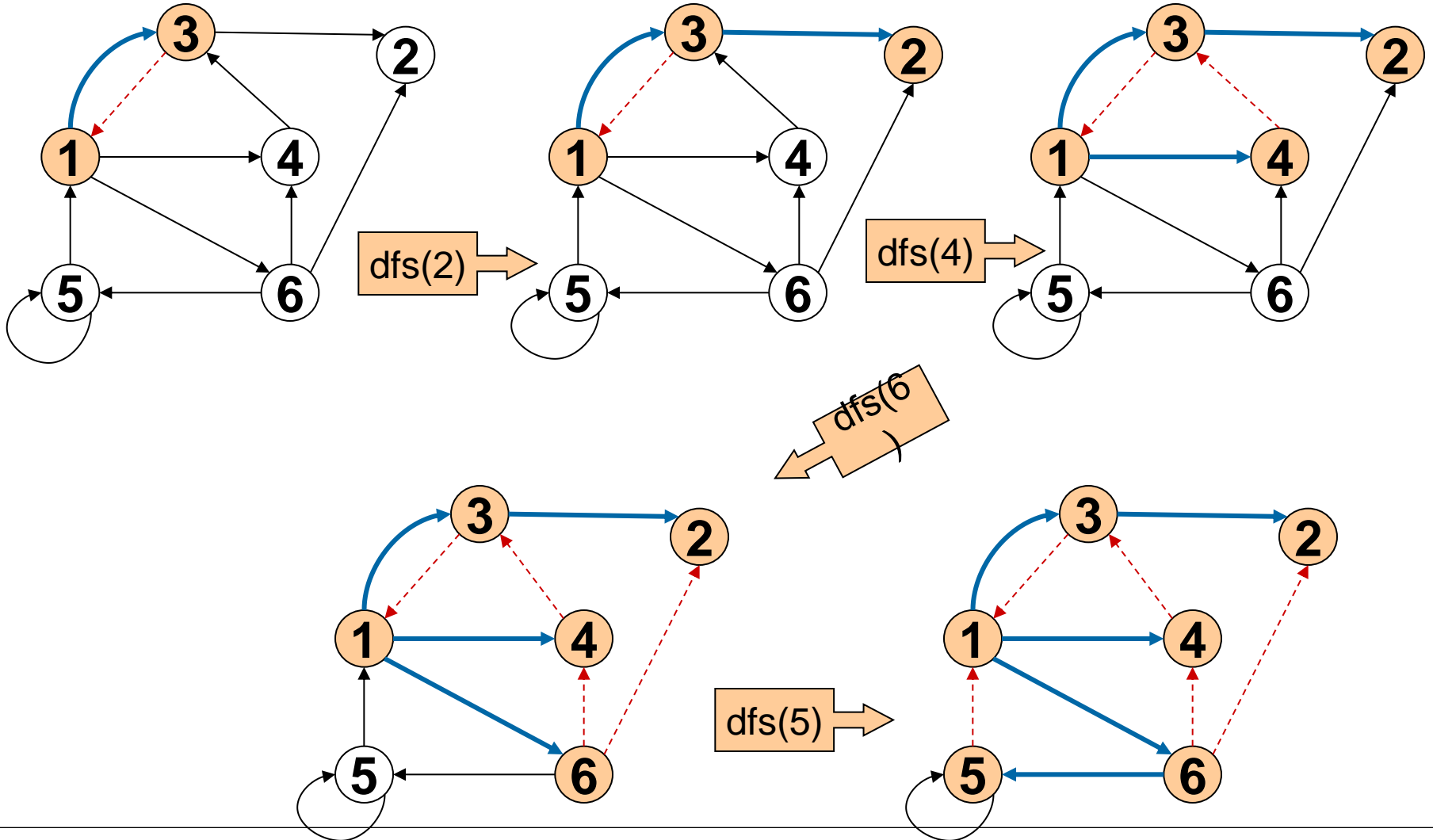
Entspricht **Preorder-Durchlauf** bei Baum:

1. zunächst alle Knoten als „**noch nicht besucht**“ markieren
2. **Startpunkt** wählen (und als „besucht“ markieren)
3. von dort aus **möglichst langen Pfad entlang gehen**; dabei nur bisher „unbesuchte“ Knoten „besuchen“ (Implementierung rekursiv oder mit explizitem Stack).
4. wenn dann noch nicht alle Knoten besucht worden sind: einen unbesuchten Knoten als **neuen Startpunkt wählen** und von 3. beginnen

# Beispiel zu DFS (1)



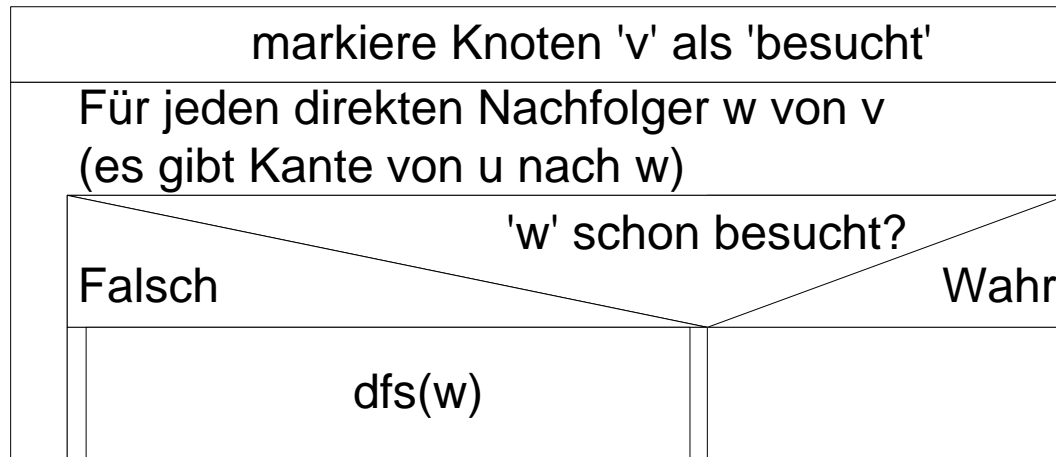
# Beispiel zu DFS (2)



Zu Beginn muss dafür gesorgt sein, dass alle Knoten als „noch nicht besucht“ markiert sind:

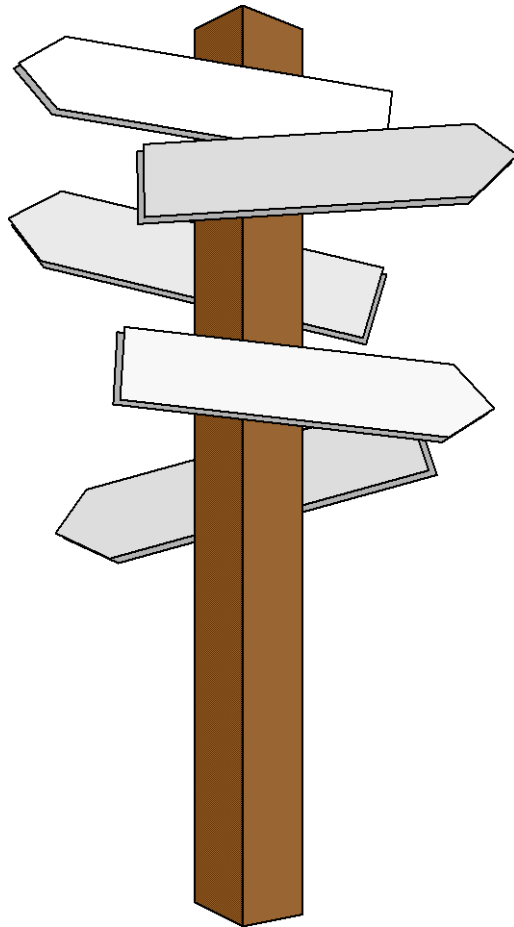
|                                     |
|-------------------------------------|
| Für alle Knoten 'v'                 |
| markiere Knoten 'v' als 'unbesucht' |

$dfs(v)$





- **Zeitkomplexität** von DFS und BFS:
  - Bei Speicherung des Graphen mit **Adjazenzlisten**:
    - DFS:  $O(|V| + |E|)$
    - BFS:  $O(|V| + |E|)$
  - Bei Speicherung des Graphen mit **Adjazenzmatrix**:
    - DFS:  $O(|V|^2)$
    - BFS:  $O(|V|^2)$
- Warum Bezeichnungen Tiefen/Breiten-**Suche**?
  - Eine Anwendung ist die **Suche** im Graphen nach einem bestimmten Knoten:
    - Vergleich des Knoteninhalts mit gesuchtem Inhalt und Abbruch bei Gleichheit.
    - Worst-Case-Komplexität bleibt gleich.



## Vorlesung 12

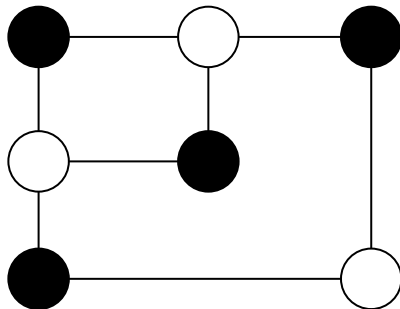
### Einfache Graphalgorithmen

In dieser Vorlesung werden behandelt:

- Bipartite Graphen
- Matching-Probleme
- Minimaler Spannbaum
- Algorithmus von Prim

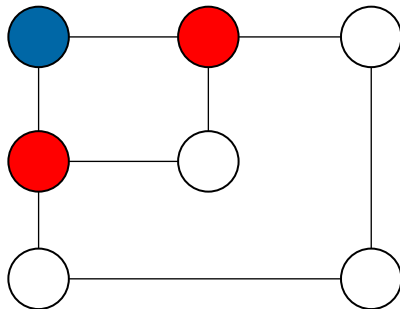
## Bipartite Graphen

- Ein Graph  $G=(V,E)$  heißt **bipartit**, wenn man  $V$  in disjunkte Mengen  $U$  und  $W$  zerlegen kann, so dass alle Kanten zwischen  $U$  und  $W$  verlaufen.
- Die Knoten eines bipartiten Graphen lassen sich mit zwei Farben (schwarz/weiß) so einfärben, dass zwei Nachbarknoten immer unterschiedlich eingefärbt sind.



Test: Ist ein Graph bipartit (hier: blau/rot)

- Einfärben des 1. Knotens
- Durchlauf mit Breitensuche
- Für jeden durchlaufenen Knoten K
  - Für jeden Nachfolger L
    - Knoten L noch nicht eingefärbt -> Knoten entgegengesetzt zu K einfärben.
    - Knoten K und L gleich eingefärbt -> Graph ist nicht bipartit.



Eine Gruppe Kinder soll in zwei Fußballmannschaften aufgeteilt werden.

Um Prügeleien innerhalb einer Mannschaft zu vermeiden, sollen Kinder, die sich nicht leiden können, in unterschiedlichen Mannschaften spielen.

Modellieren Sie das Problem und beschreiben Sie einen Lösungsalgorithmus, der ermittelt:

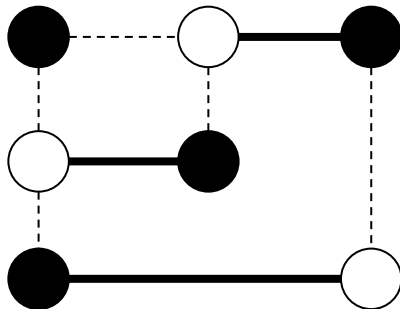
- Gibt es eine entsprechende Aufteilung?
- Wenn ja, wie sieht die Aufteilung aus?

- Studierende einer Hochschule wollen ein Benefiz-Rugby-Spiel organisieren. Aus Sorge vor Repressalien, sollen Studenten nur gegen Professoren spielen, bei denen Sie keine Prüfung mehr ablegen müssen. Umgekehrt sollen, im Sinne der Teambildung, in einem Team nur Professoren sein, die allen Studenten ihres Teams noch mindestens eine Prüfung abnehmen müssen.
- Entwickeln Sie einen Algorithmus, der entscheidet ob die Teilnehmer unter dieser Voraussetzungen grundsätzlich auf zwei Teams aufgeteilt werden können.
- Ihr Algorithmus sollte eine Komplexität von  $O(n + r)$  haben, wobei  $n$  die Zahl der Teilnehmer ist und  $r$  die Anzahl der Paarungen Professor, Student ist, die sich in keiner Prüfung mehr begegnen werden.

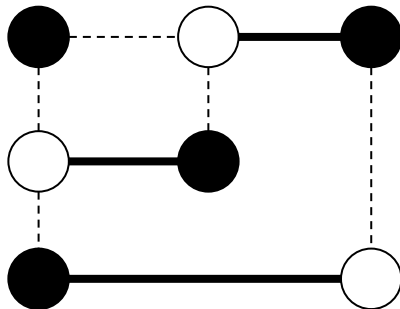


## Matching-Probleme

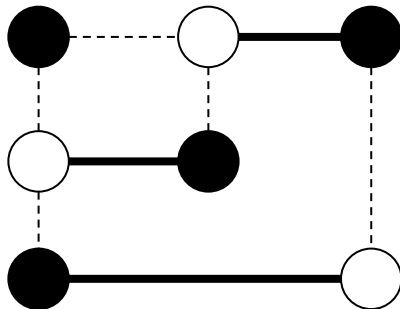
- Matching-Probleme: Finden einer Teilmenge von Kanten, so dass möglichst viele weiße und schwarze Elemente einander eindeutig zugeordnet sind.
- Einfacher Spezialfall: Matching bei bipartiten Graphen



- Kann dem Bild unten eine weiteres Zuordnung hinzugefügt werden (ist das Matching **erweiterbar**)?
- Ordne möglichst viele schwarze Punkte entsprechenden weißen Punkten zu (finde das **größtmögliche Matching**).

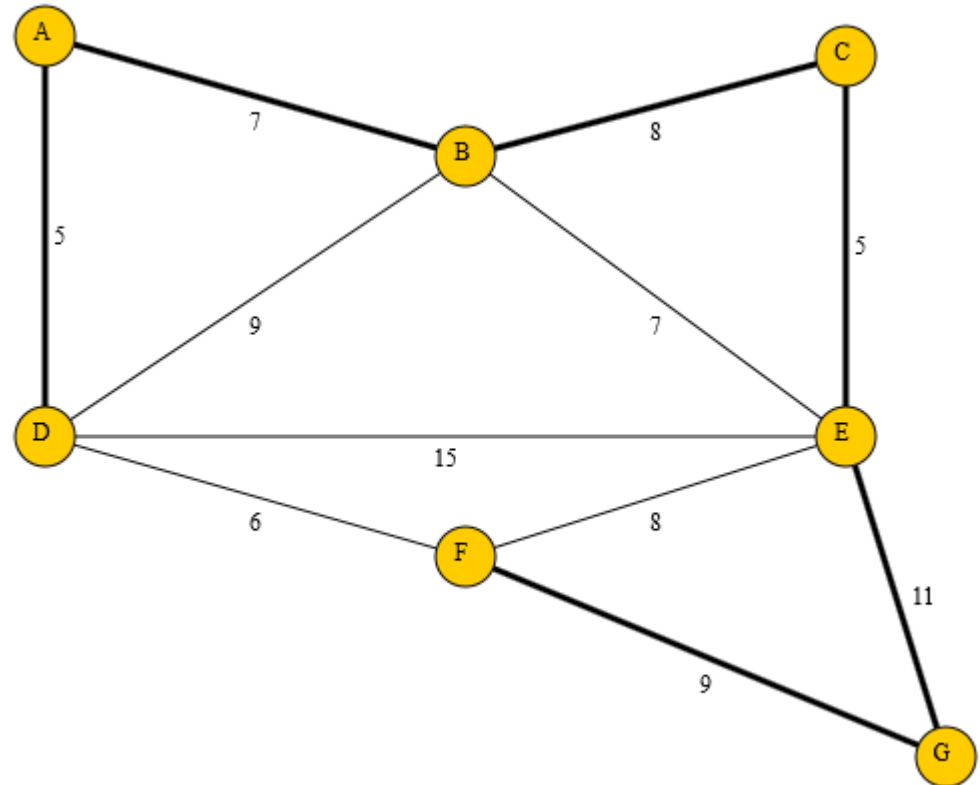


- Kann man alle Punkte einander zuordnen (gibt es ein **perfektes Matching**)?
- Die Lösungsalgorithmen sind nicht Bestandteil der Vorlesung.



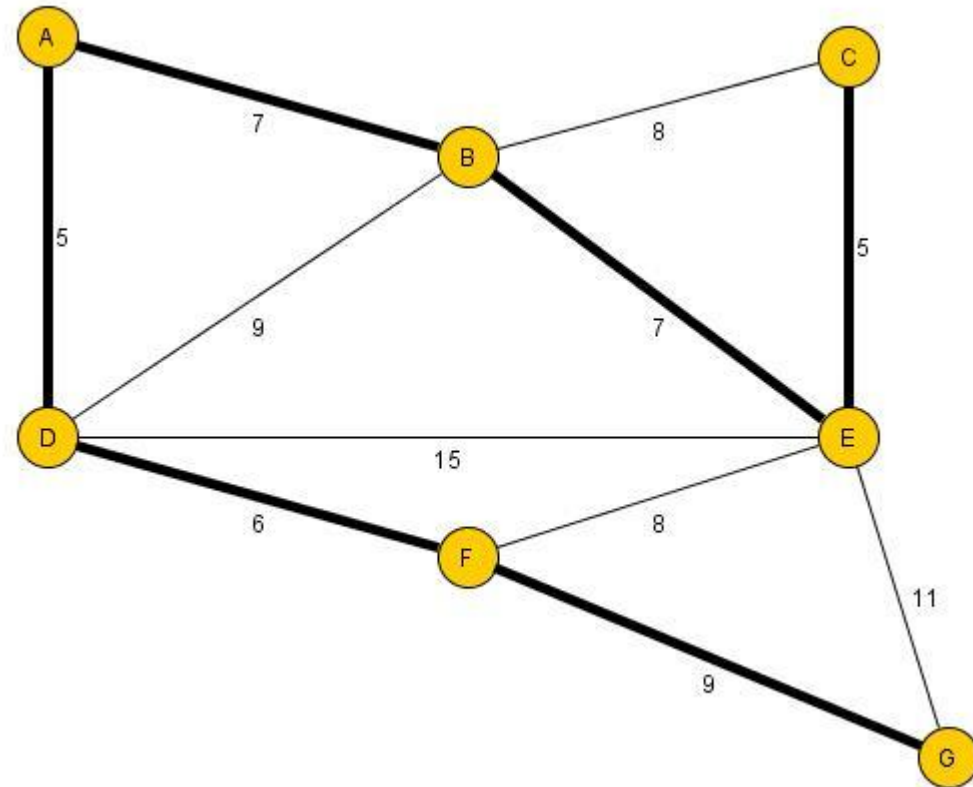
## Minimaler Spannbaum / Algorithmus von Prim

Ein **Spannbaum** verbindet alle Knoten eines ungerichteten Graphen miteinander, hat jedoch keine Zirkel.

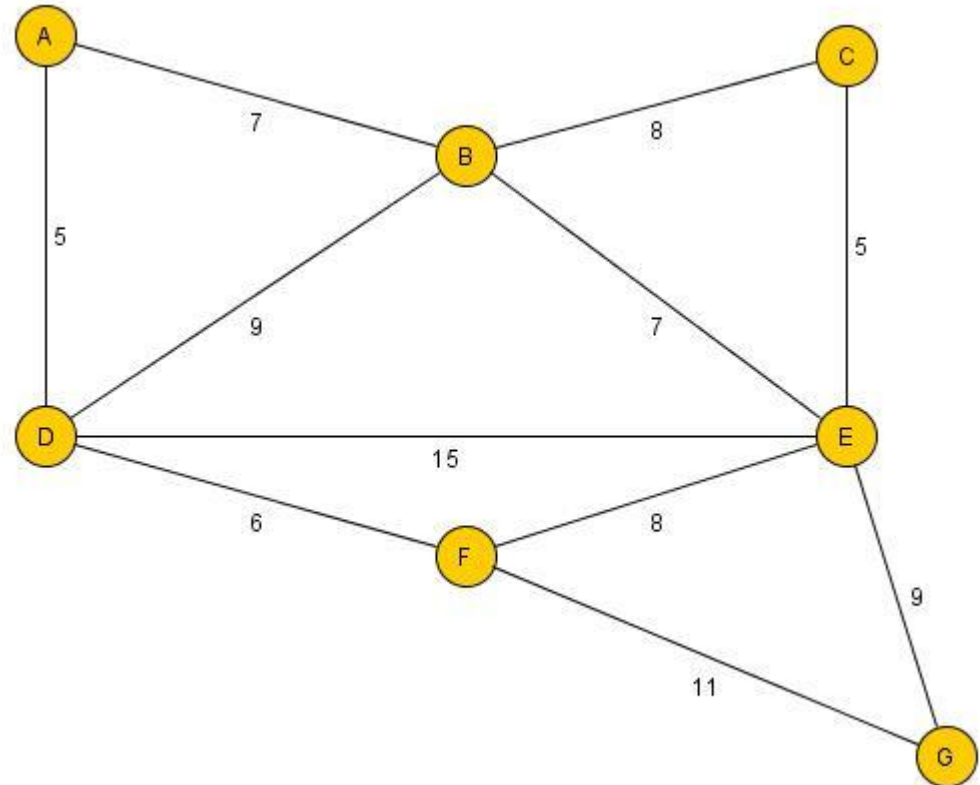


Der minimale Spannbaum ist der Spannbaum, dessen Kanten das kleinste summierte Gewicht haben.

Einen minimalen Spannbaum kann man nach dem **Algorithmus von Prim** (Vorlesung) oder dem Algorithm. von Kruskal (nicht Vorl.) konstruieren.



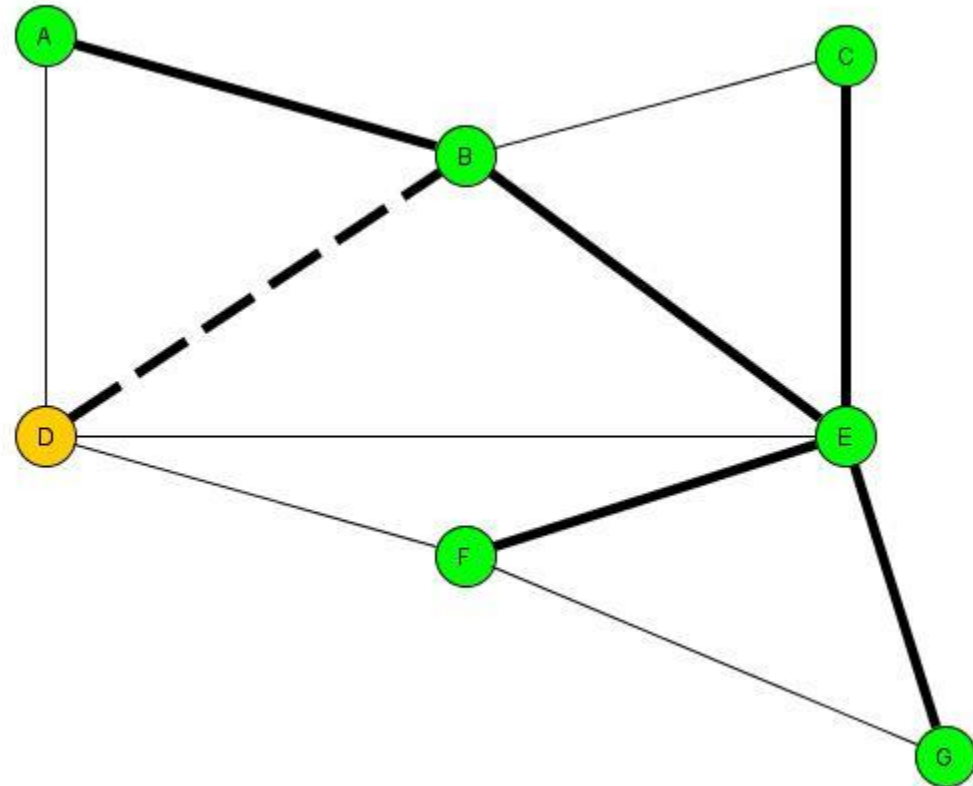
- Beispiel:



Zunächst sucht man sich einen beliebigen Knoten (hier Knoten D).

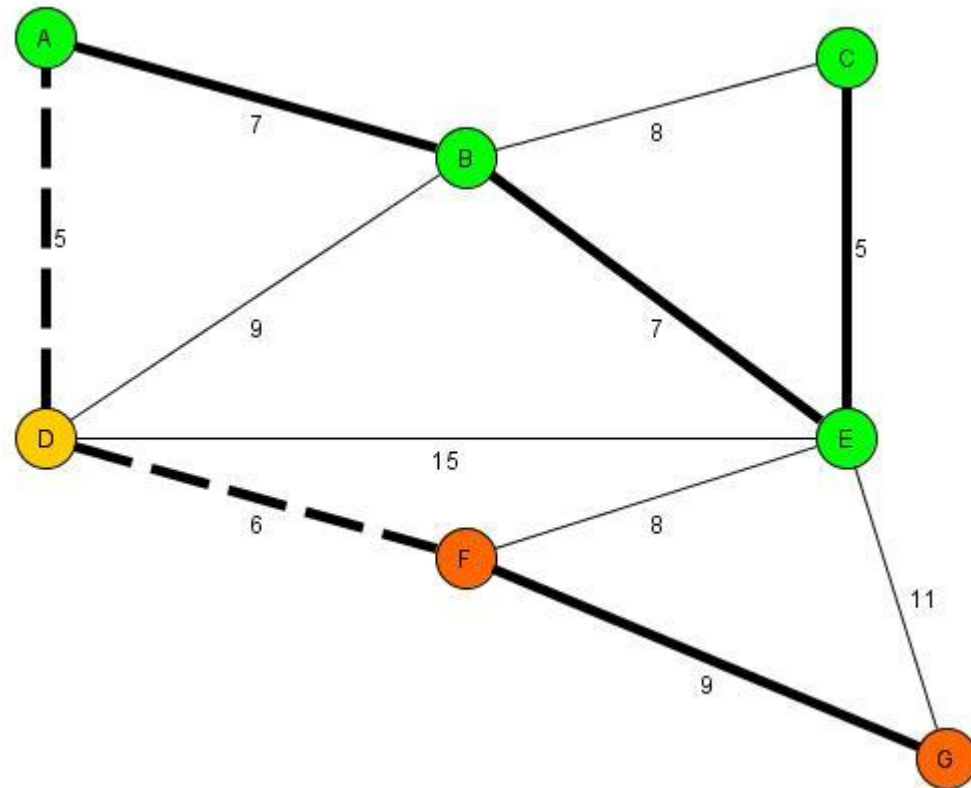


- Dieser Knoten ist entweder über eine einzige Kante mit dem Rest des (noch unbekanntem) Spannbaums verbunden...
- (Hier ist ein Spannbaum zur Verdeutlichung schon dick eingezeichnet).



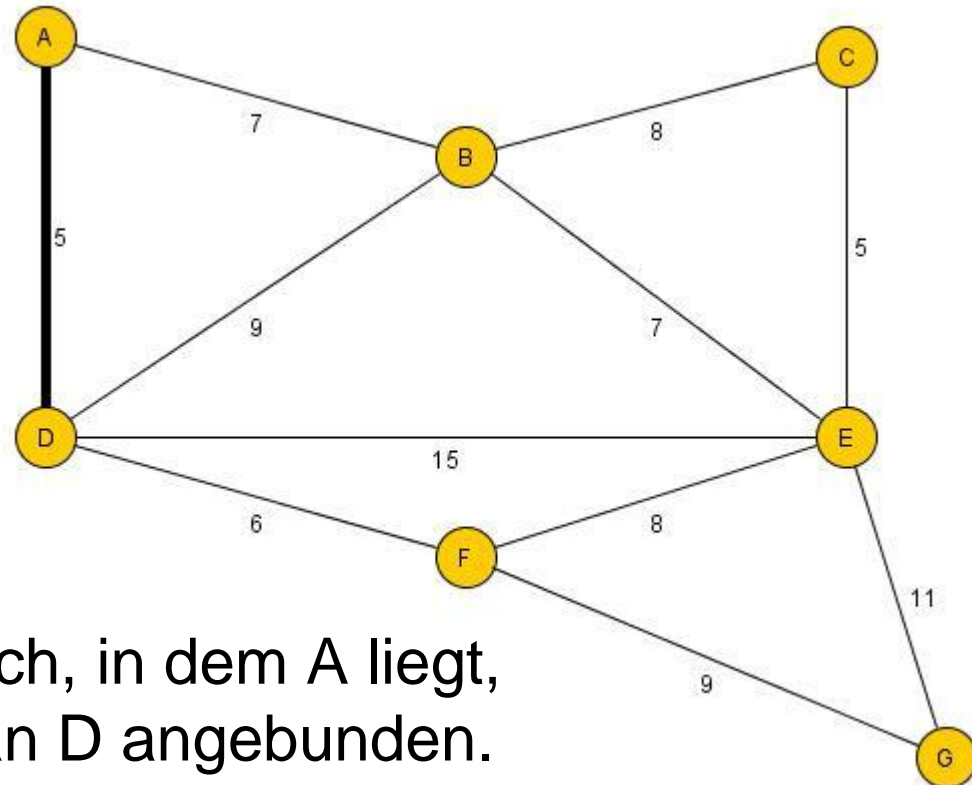
## Konstruktion nach Prim

- .. oder der Knoten ist über mehrere Kanten mit einzelnen Teilen des Spannbaums verbunden (rot und grün).
- Die Teile dürfen untereinander nicht verbunden sein -> sonst Zirkel

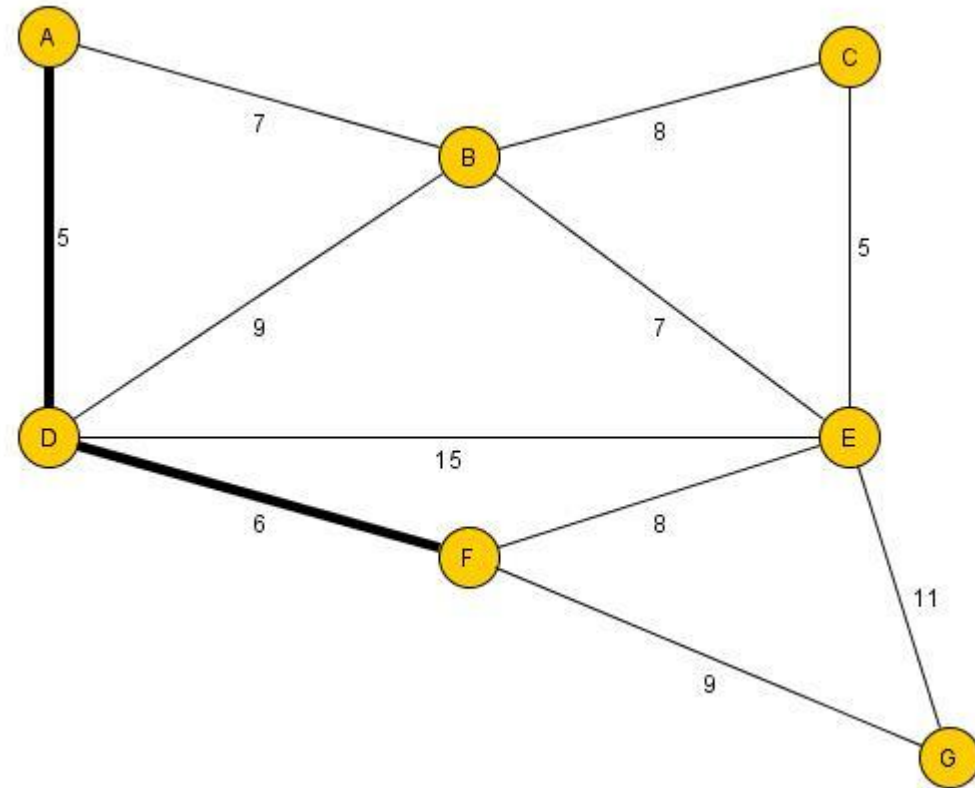


In beiden Fällen gilt:

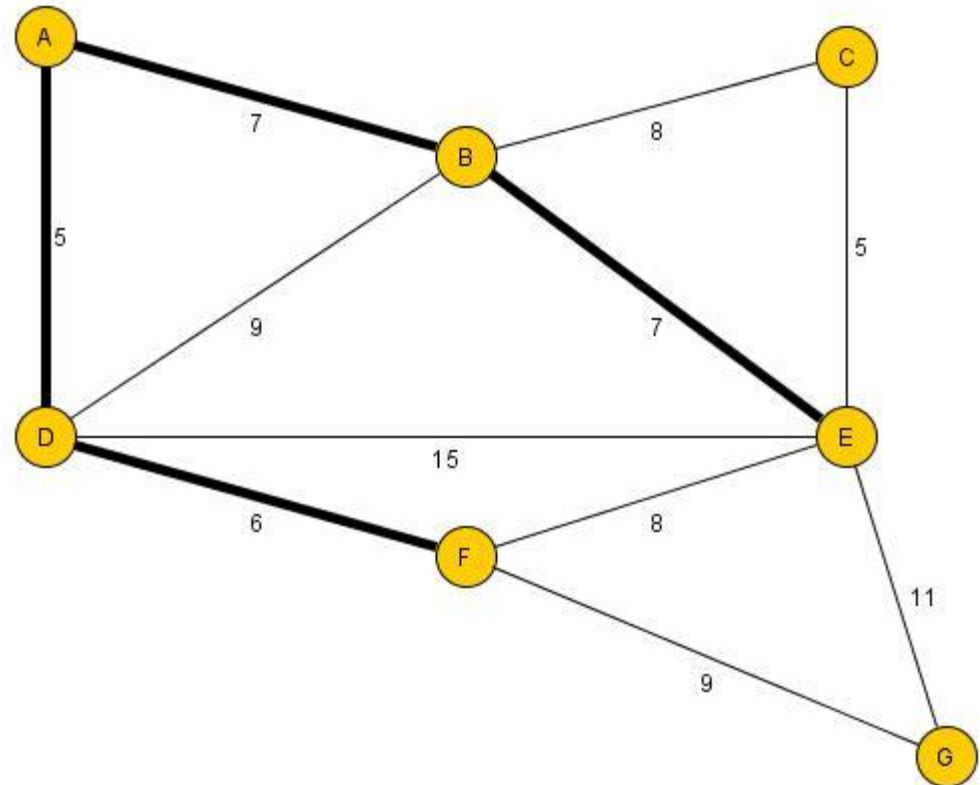
- Die niedrigste Kante von D aus (D- $\rightarrow$ A; Wert 5). **muss** zum minimalen Spannbaum gehören.
- Ansonsten wäre der Bereich, in dem A liegt, über eine größere Kante an D angebunden. Diese Kante könnte gegen die Kante A- $\rightarrow$ D ausgetauscht und der Spannbaum damit verkleinert werden.



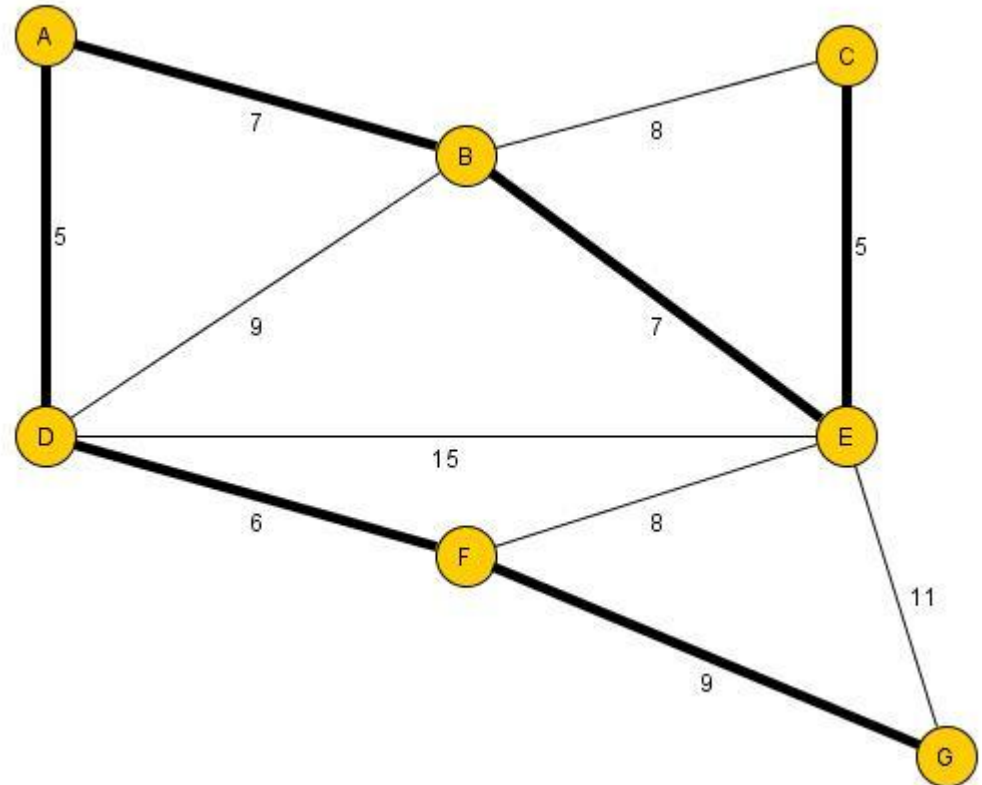
- Nun wiederholt man die gleiche Argumentation:
- Die Kante mit dem kleinsten Wert, die von den verbundenen Knoten zu nicht verbundenen Knoten führt, muss zum minimalen Spannbaum gehören.

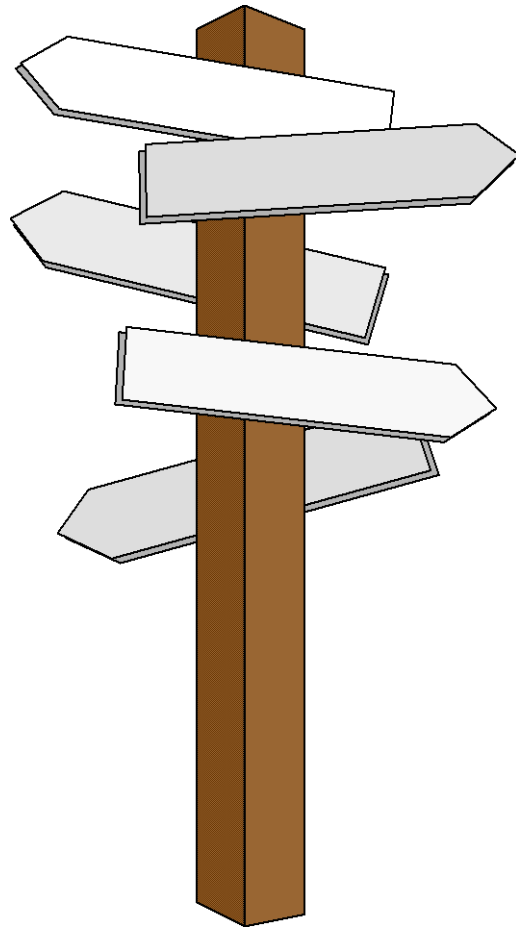


- Man fügt also solange die kleinste Kante zu einem noch freien Knoten hinzu, bis alle Knoten verbunden sind.



- Der fertige Spannbaum:





## Vorlesung 13

### Entwurfsprinzipien Traveling Salesman Problem

## Themen der Vorlesung

- **Entwurfsprinzipien**  
Algorithmen lassen sich in Gruppen sortieren, die einem ähnlichen Aufbau folgen.
- **Problem des Handlungsreisenden  
(Traveling Salesman Problem /TSP)**  
Bekanntes Problem der Informatik, von dem Näherungslösungen mit unterschiedlichen Entwurfsprinzipien existieren.



Diese Entwurfsprinzipien werden in der Vorlesung betrachtet:

heute

Greedy

Backtracking

Branch and Bound

Divide and Conquer

Dynamische Programmierung

Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

Nearest Neighbour (Greedy)

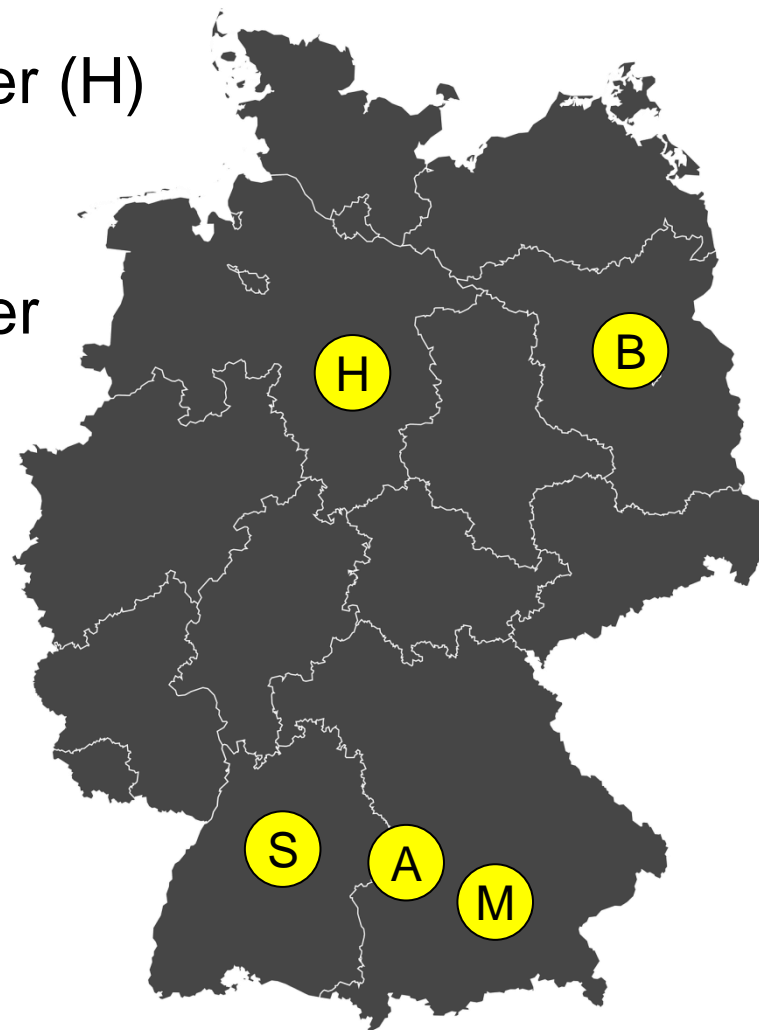
Random Insertion

Minimaler Spannbaum

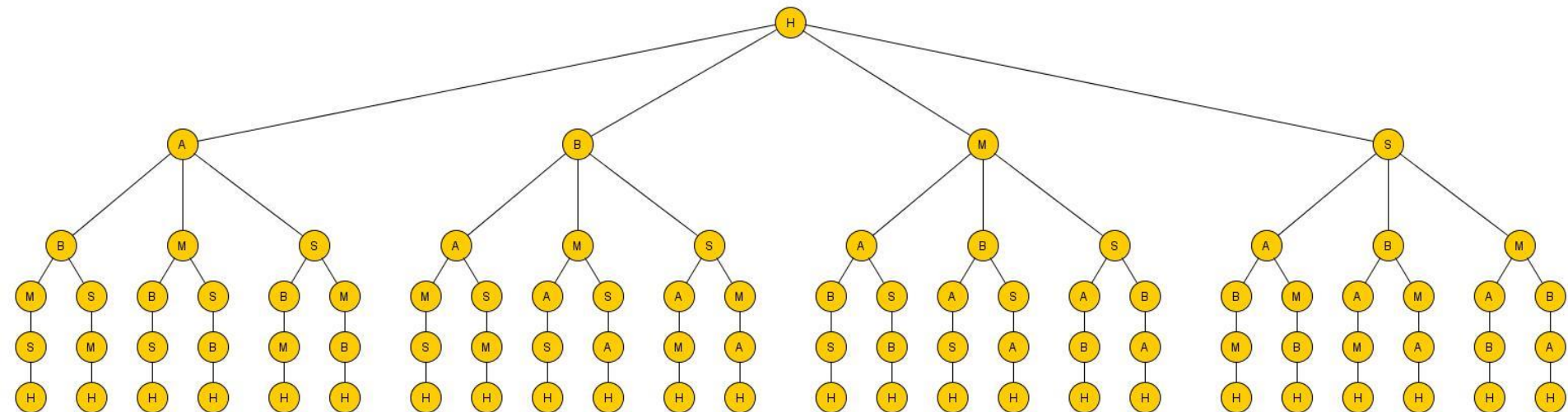
Brute Force (Backtracking)

Branch and Bound

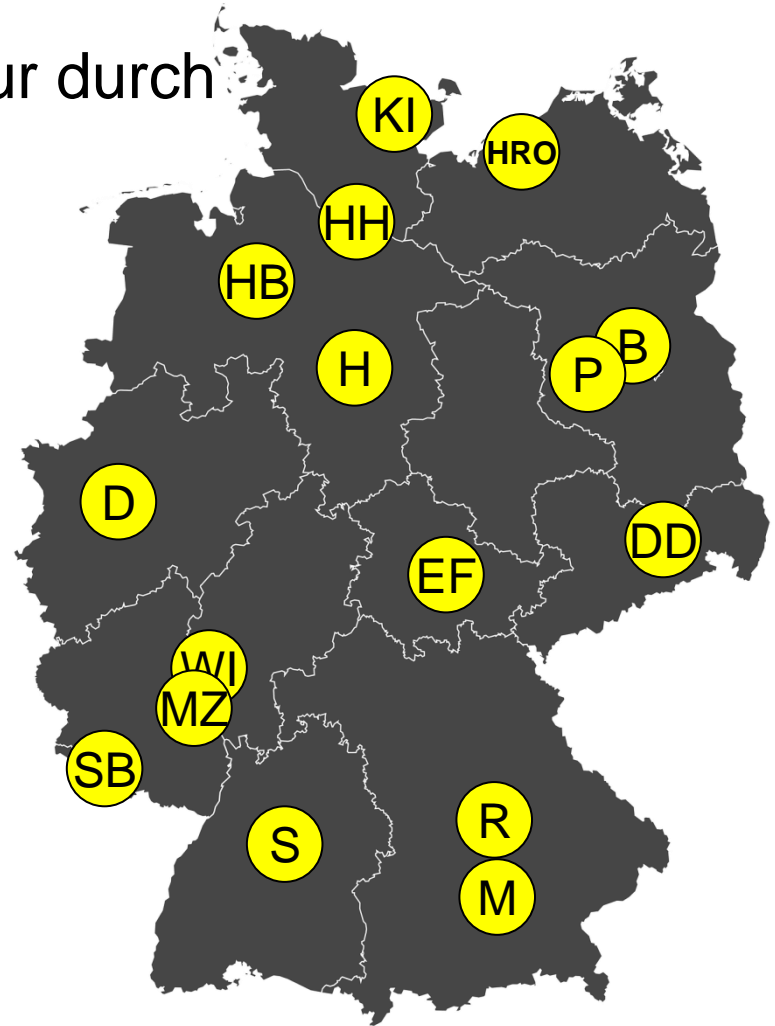
- Ein Vertreter muss von Hannover (H) aus die Städte Berlin (B), Augsburg (A) und München (M) besuchen, ehe er nach Hannover zurückkehrt.
- Er möchte gerne die kürzeste Route nehmen.



- Dieses Problem erscheint trivial.
- Alle Möglichkeiten zusammengenommen ergeben aber bereits einen recht umfangreichen Baum.



- Ein größeres Beispiel ist die Tour durch die 16 Landeshauptstädte.
- Dieses Problem ist nicht mehr trivial.
- Es gibt  $(n-1)!$  mögliche Touren.
- Immer jeweils 2 Touren sind symmetrisch.
- Alle Touren abprüfen:  $O(n!)$ .  
Faustregel: Probleme bei  $n > 10$ .



Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

**Nearest Neighbour (Greedy)**

Random Insertion

Minimaler Spannbaum

Christofides

Brute Force (Backtracking)

Branch and Bound

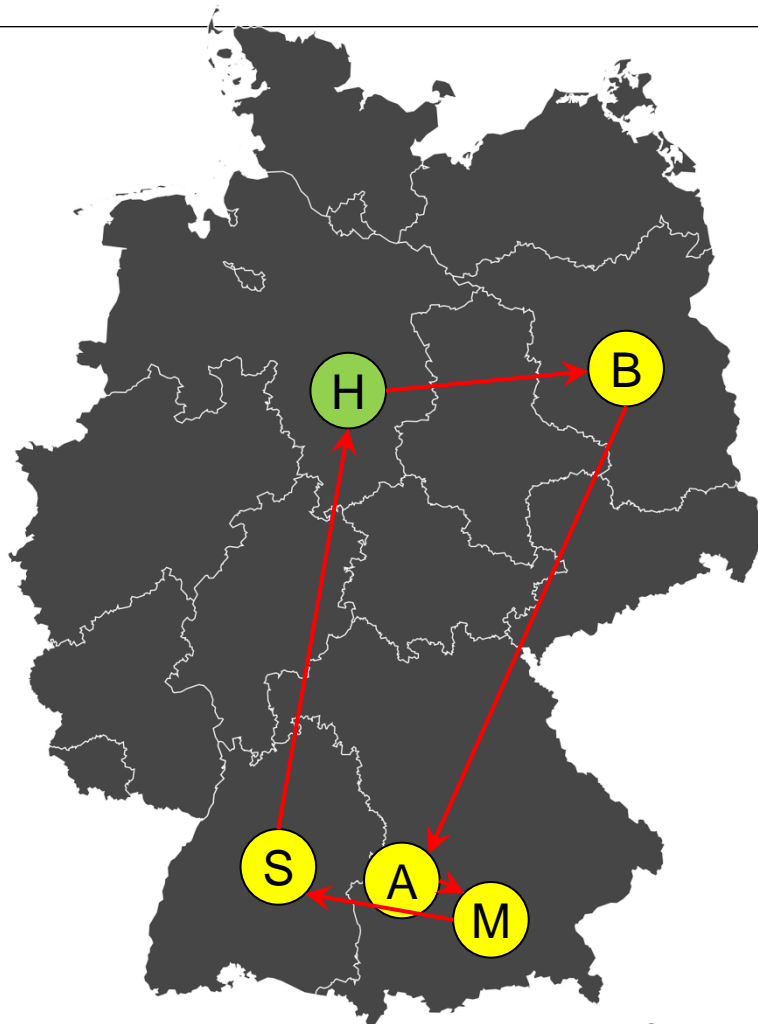
### Entwurfsprinzip Greedy:

Wähle immer den Schritt, dessen unmittelbarer Folgezustand das beste Ergebnis liefert. Denke nicht an die weiteren Schritte (greedy = gierig).

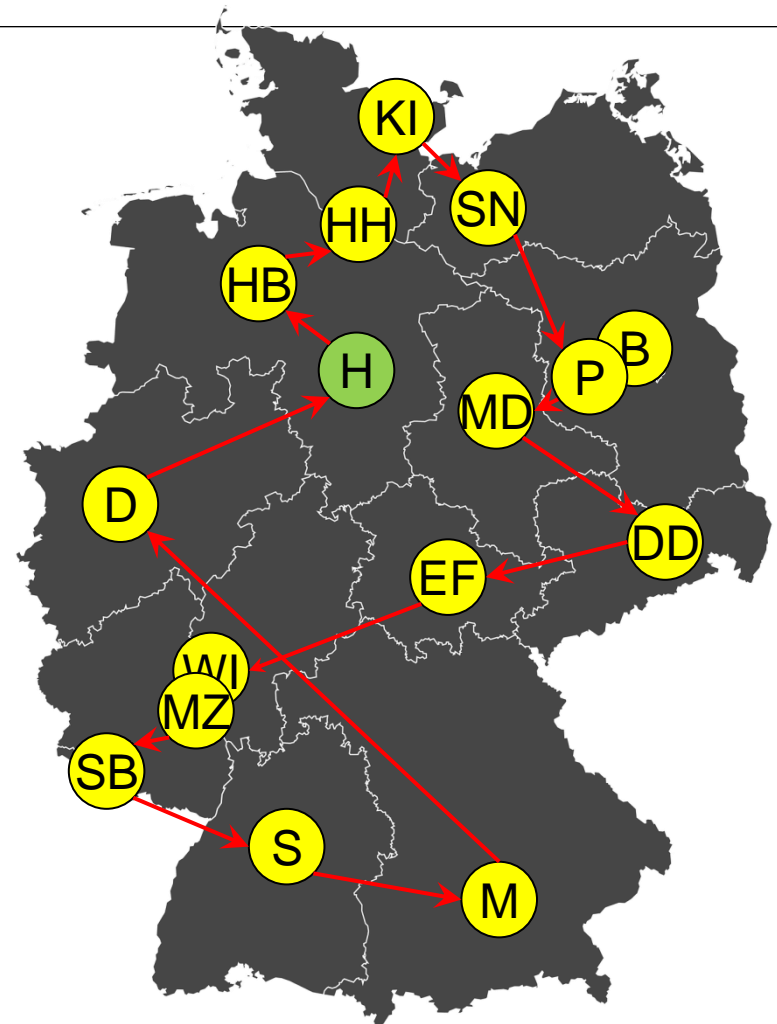
Eine Greedy-Lösung ist die **Nearest-Neighbour-Lösung**:

- Gehe in jedem Schritt zur nächstliegenden Stadt, die noch nicht besucht ist.
- O-Klasse:  $O(n^2)$ .

# Nearest-Neighbour-Lösung



1639 km – optimal?



3183 km – nicht optimal



Ob ein Greedy-Algorithmus die optimale Lösung liefert, oder nicht, ist vom konkreten Problem abhängig.

Bei TSP gilt: Greedy liefert nicht die optimale Lösung.

Ganz im Gegenteil: Der Quotient  $\frac{\text{Greedy-Wert}}{\text{optimaler Wert}}$  kann beliebig schlecht werden.

Nearest-Neighbour arbeitet aber meistens einigermaßen gut.

Diese Entwurfsprinzipien werden in der Vorlesung betrachtet:

**Greedy:** Mache in jedem Teilschritt den größtmöglichen Schritt.

Backtracking

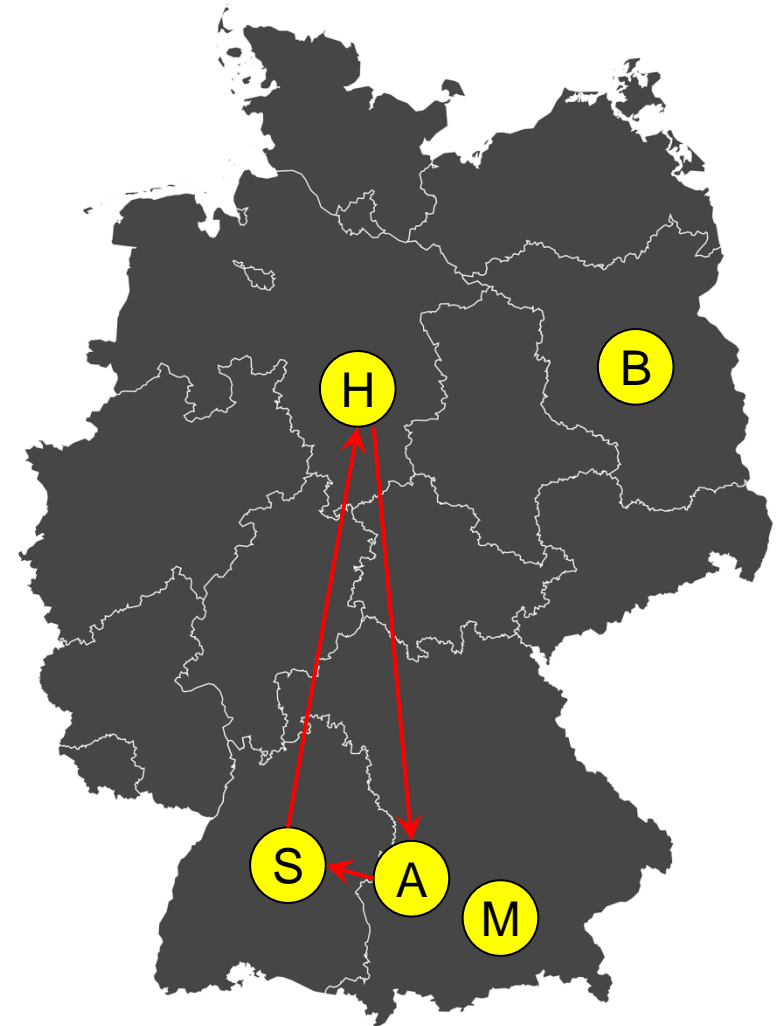
Branch and Bound

Divide and Conquer

Dynamische Programmierung

Wähle zufällig 2 Städte aus und verbinde sie mit der Start-/Zielstadt.

Diese Verbindung ist (abgesehen von der Richtung) eindeutig.



Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

**Nearest Neighbour (Greedy)**

**Random Insertion**

**Minimaler Spannbaum**

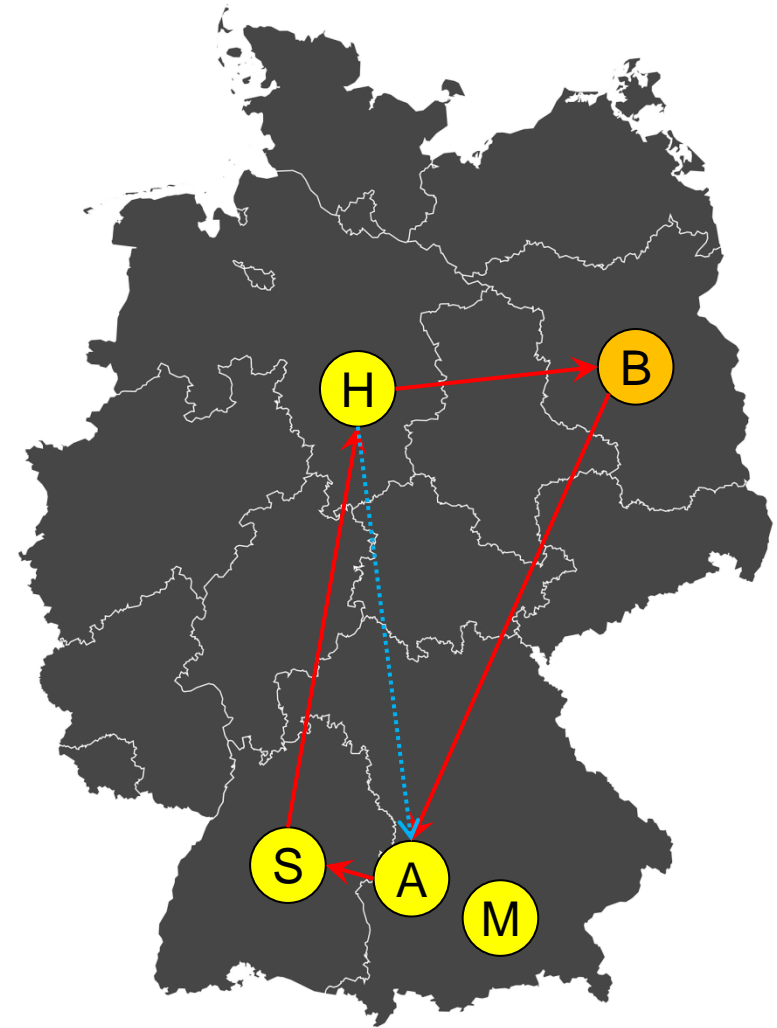
**Christofides**

**Brute Force (Backtracking)**

**Branch and Bound**

Wähle zufällig eine weitere Stadt.

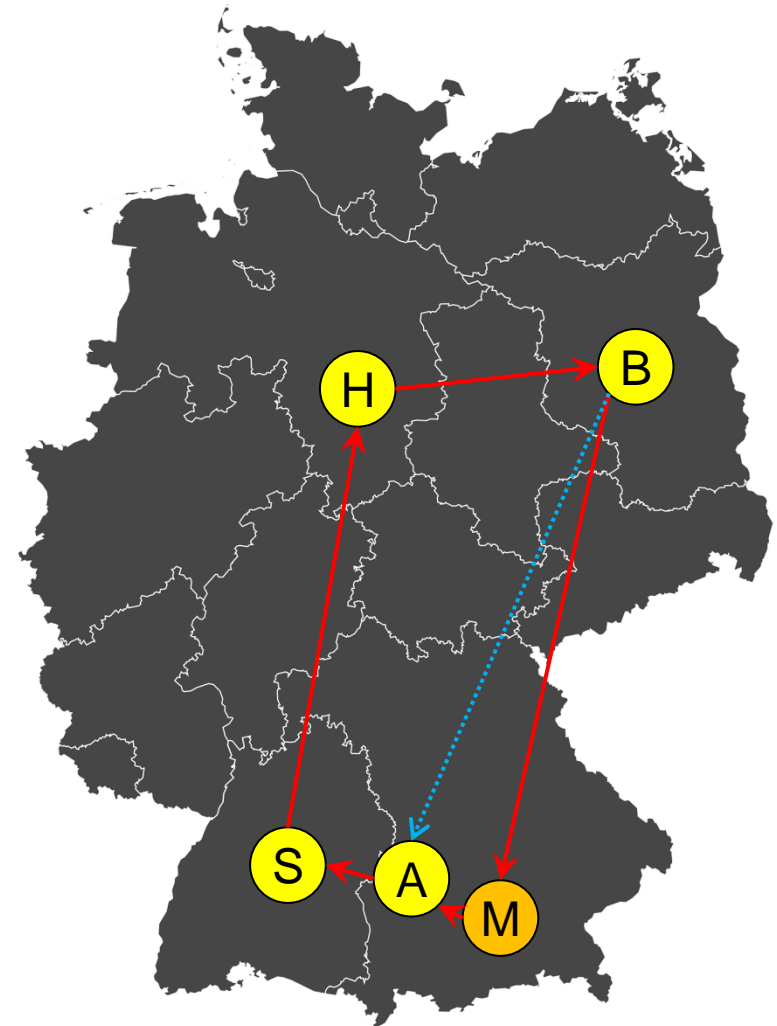
Füge die Stadt an der Stelle ein,  
an der die Route am wenigsten  
vergrößert wird.



Wiederhole den Vorgang, bis alle Städte eingefügt sind.

In diesem Beispiel erhält man den optimalen Weg (1591 km).

Da die Reihenfolge der Städte zufällig ist, kann die Lösung mit weniger Glück auch nicht optimal sein.



O-Klasse:  $O(n^2)$ .

Auch für Random-Insertion gibt es keine Garantie, dass die Lösung irgend einem Gütekriterium genügt.

Im Allgemeinen ist die Lösung aber ziemlich gut.

Weiterer Vorteil: Durch den Zufallsfaktor kann man das Verfahren beliebig oft wiederholen und sich die beste Lösung heraussuchen.

Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

**Nearest Neighbour (Greedy)**

**Random Insertion**

**Minimaler Spannbaum**

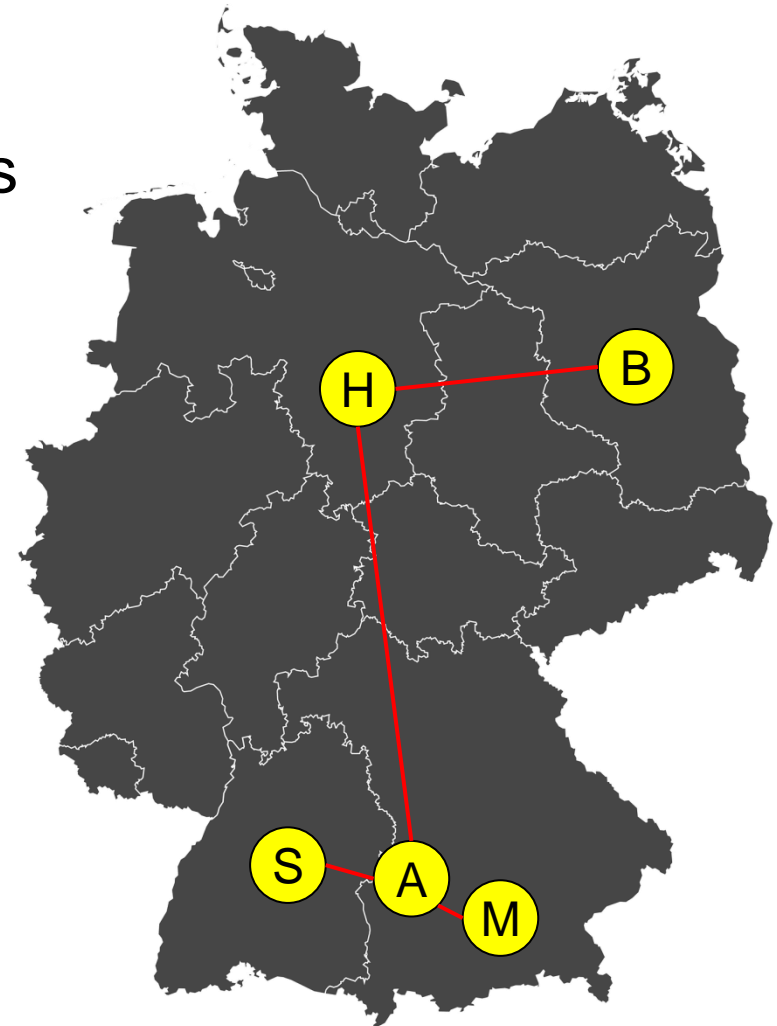
**Christofides**

**Brute Force (Backtracking)**

**Branch and Bound**

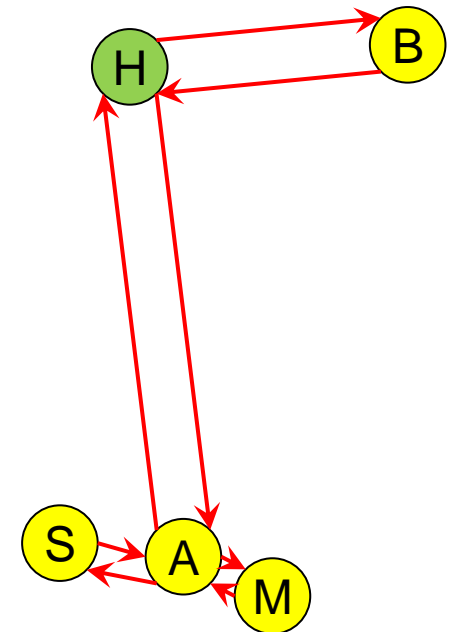


Konstruktion eines minimalen  
Spannbaums nach dem Algorithmus  
von Prim.



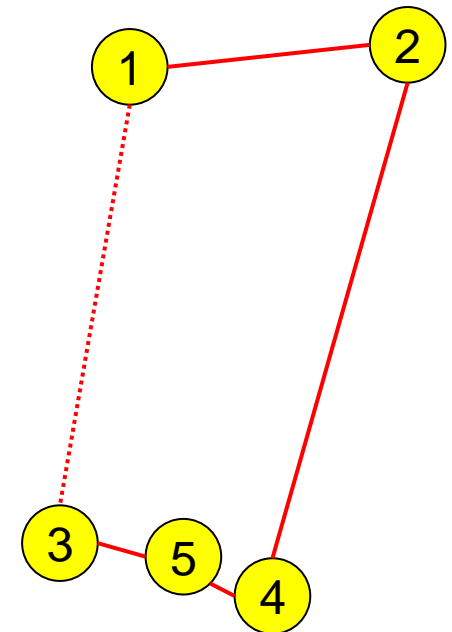
Hat man den minimalen Spannbaum gefunden, dann durchläuft man ihn mit Tiefensuche.

- Jede Kante wird dabei 2x durchlaufen.



Die summierten Gewichte des minimalen Spannbaums müssen kleiner als das optimale TSP-Ergebnis sein.

Ansonsten könnte man vom TSP-Ergebnis eine Kante auflösen und hätte einen noch besseren Spannbaum.

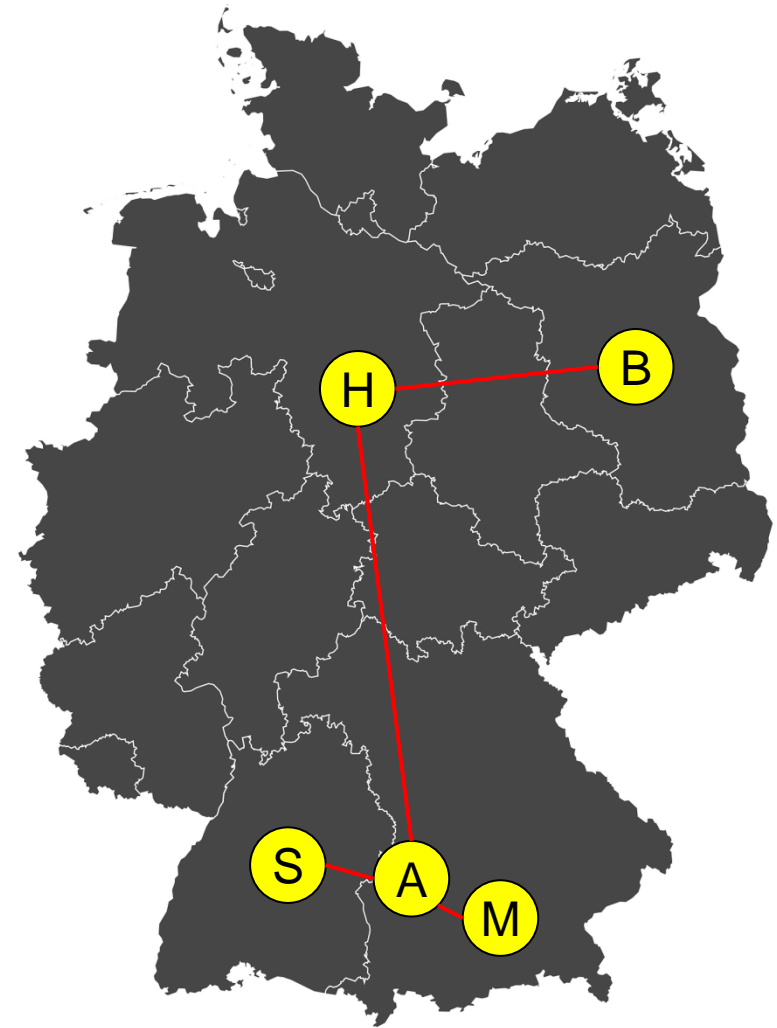


Der doppelte Durchlauf muss also **besser sein als das Doppelte des optimalen Wertes.**

Hier:

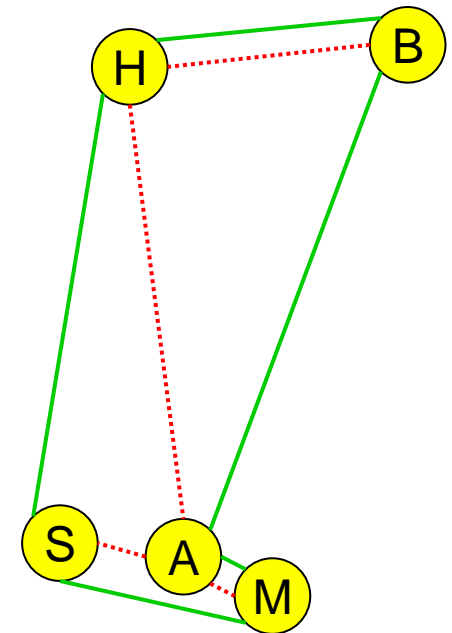
Spannbaum: 1083 km

Doppelter Durchlauf: 2166 km

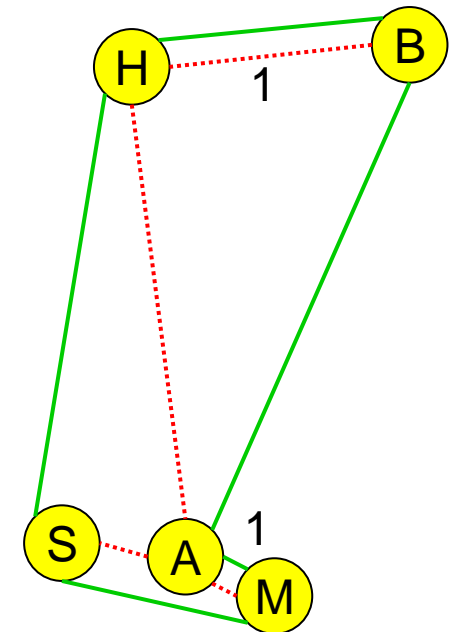


Optimierung: Hat man einen Knoten schon einmal besucht, überspringt man ihn beim nächsten Mal.

Pythagoras: Die Lösung verbessert sich (1639 km).

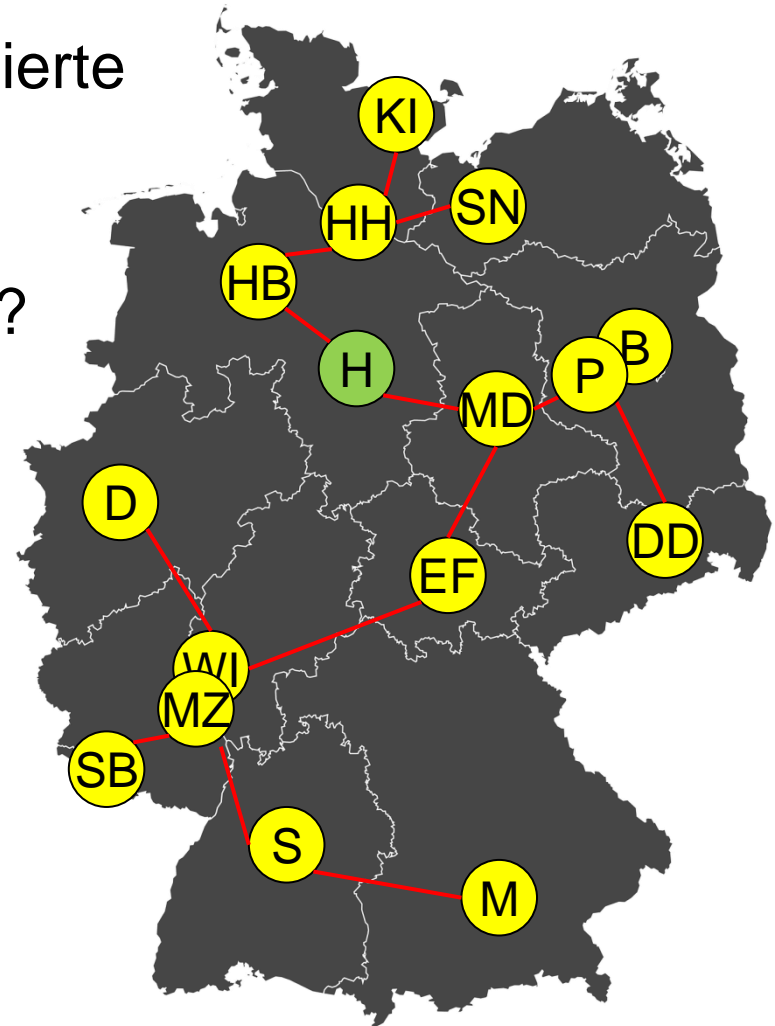


Hat man bei einem Schritt mehrere Möglichkeiten, kann man z.B. zuerst den kürzeren Weg wählen.



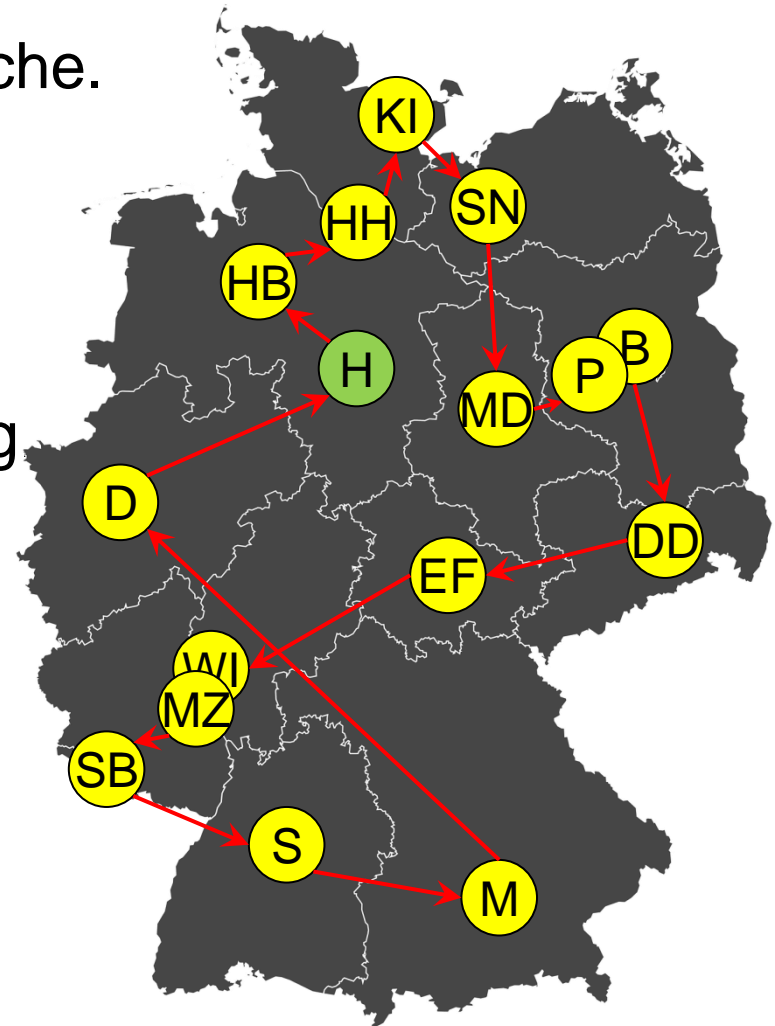
# Minimaler Spannbaum – 16 Städte

- Die Lösung hat also eine garantierte Güte.
- Reicht uns das im Problem mit den 16 Landeshauptstädten?
- Hier der minimale Spannbaum.



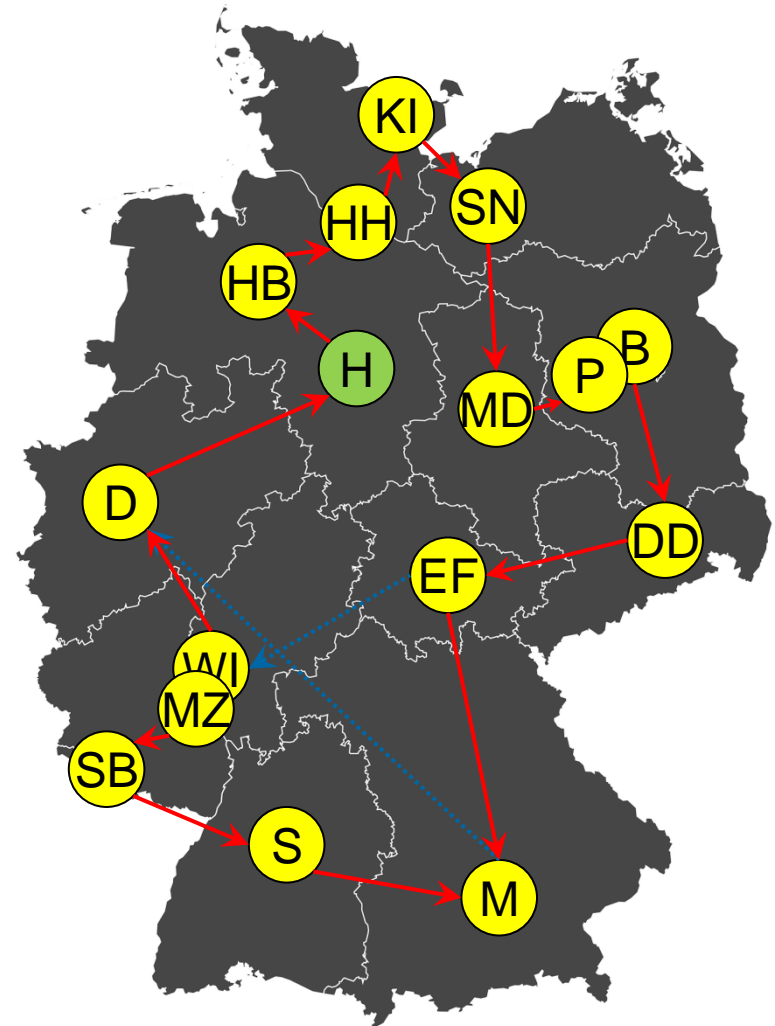
# Minimaler Spannbaum – 16 Städte

- Hier der Durchlauf mit Tiefensuche.
- Ergebnis: 3215 km
- Das Ergebnis hat eine Kreuzung
- Von Hand wäre es sicher besser geworden.





- Einfachste Verbesserung: Kreuzungen auflösen.
- Das Ergebnis muss sich verbessern
- Ergebnis: 2982 km



Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

**Nearest Neighbour (Greedy)**

**Random Insertion**

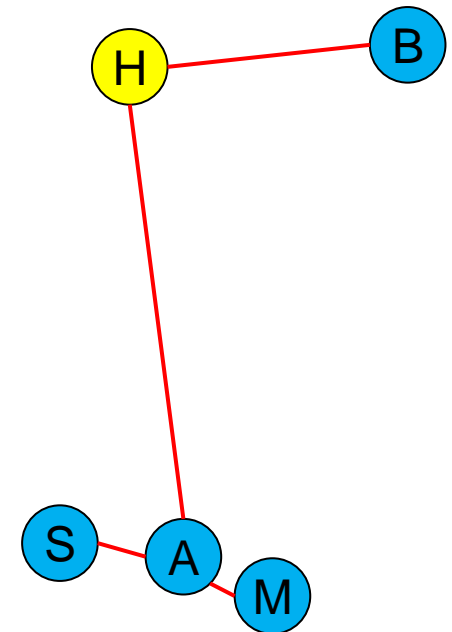
**Minimaler Spannbaum**

**Christofides**

**Brute Force (Backtracking)**

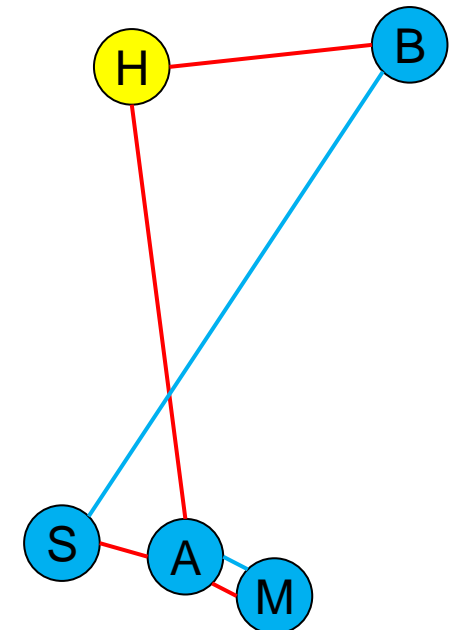
**Branch and Bound**

- Schafft eine bessere Garantie:  
Im schlechtesten Fall  
1,5x optimale Lösung.
- Ausgangspunkt ist der  
minimale Spannbaum.
- Man betrachtet die Knoten mit  
einer ungeraden Anzahl von  
Kanten (blau).



- Für diese Knoten sucht man ein perfektes Matching mit minimalem Gewicht (blau). Diese Operation hat  $O(n^3)$ .
- Diese Kanten fügt man hinzu.

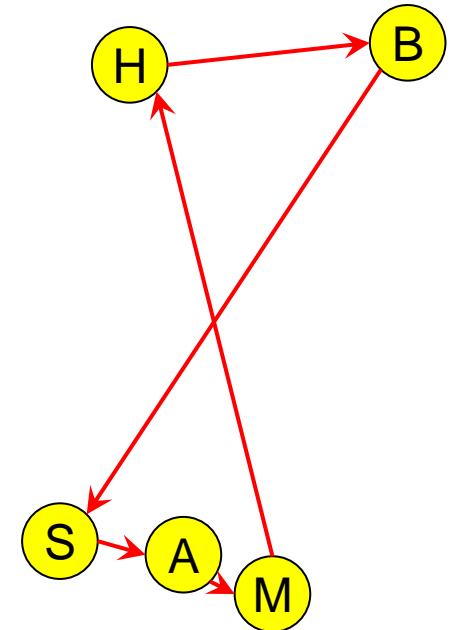
|   | A         | B          | H   | M   | S |
|---|-----------|------------|-----|-----|---|
| A | -         |            |     |     |   |
| B | 588       | -          |     |     |   |
| H | 609       | 251        | -   |     |   |
| M | <b>78</b> | 597        | 646 | -   |   |
| S | 145       | <b>660</b> | 520 | 202 | - |



Den daraus entstehenden Zirkel durchläuft man und überspringt wieder doppelte Punkte.

Schon an diesem kleinen Beispiel ist zu erkennen:

- Die Garantie ist gut.
- Das Ergebnis sieht trotzdem enttäuschend aus.



Für unser Problem mit den Landeshauptstädten haben wir bisher betrachtet:

Backtracking:  $O(n-1)!$ , Performanceprobleme

Nearest Neighbour:  $O(n^2)$ , Ergebnis enttäuschend

Random Insertion:  $O(n^2)$ , Ergebnis zufallsabhängig

Minimaler Spannbaum:  $O(n^2)$ , Garantie (opt x 2), Ergebnis enttäuschend

Christofides:  $O(n^3)$ : Garantie (opt x 1,5), Ergebnis enttäuschend

Aber:

Die Algorithmen funktionieren (nach Daumenregel) bis  $n=100.000$  (Christofides bis  $n=1000$ ).

Für so viele Städte sind die Ergebnisse vielleicht doch nicht so schlecht.

Bis  $n=20$  hat man (nach Daumenregel) auch eine Chance, alle Möglichkeiten durchzuprobieren.

Das simple Durchprobieren vieler Möglichkeiten nennt man **Brute Force**. Versuchen wir es einfach einmal.

Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

**Nearest Neighbour (Greedy)**

**Random Insertion**

**Minimaler Spannbaum**

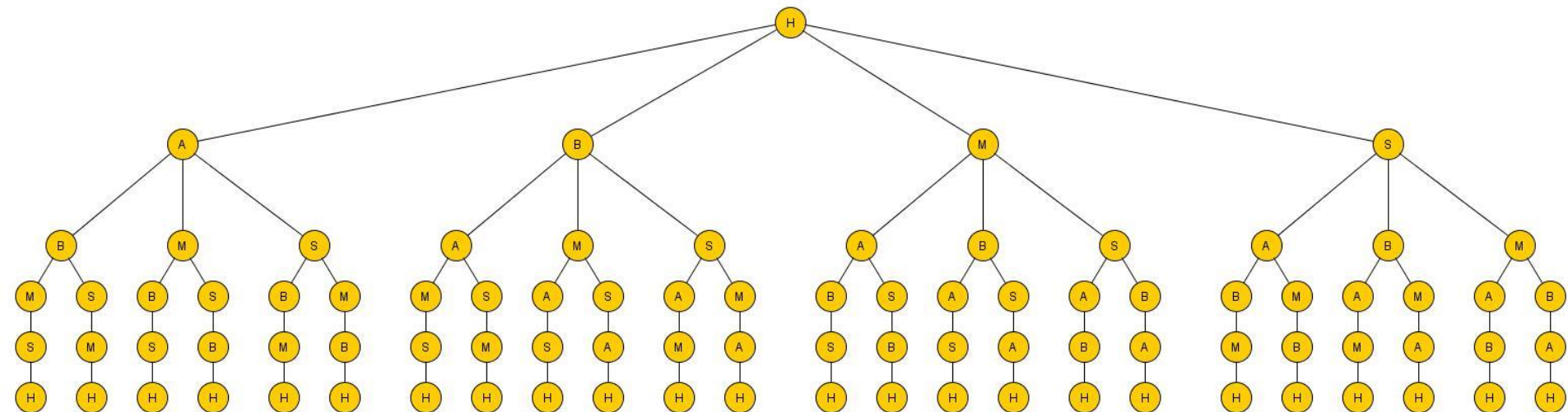
**Christofides**

**Brute Force (Backtracking)**

**Branch and Bound**



- Alle möglichen Wege lassen sich als Baum darstellen. Dieser Baum umfasst alle Möglichkeiten für  $n=5$ .
- **Backtracking**: Man durchläuft diesen Baum in Pre-Order.
- Jeder Weg ist zweimal vertreten (Hin- und Rückweg).





Diese Entwurfsprinzipien werden in der Vorlesung betrachtet:

**Greedy:** Mache in jedem Teilschritt den größtmöglichen Schritt.

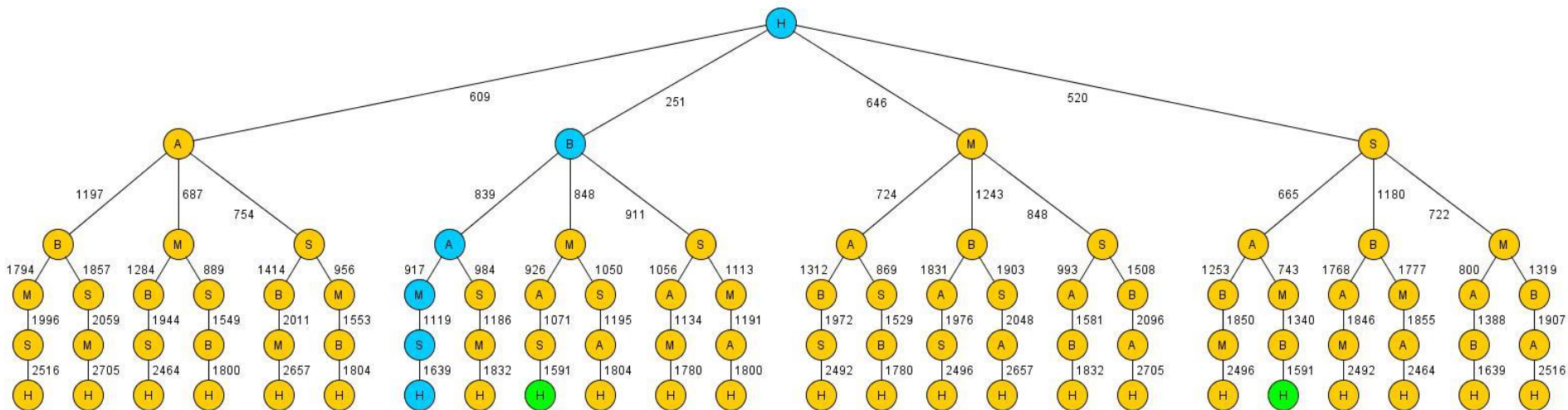
**Backtracking:** Durchlauf eines Lösungsbaums in Pre-Order-Reihenfolge.

Branch and Bound

Divide and Conquer

Dynamische Programmierung

- Hier ist gut zu sehen, dass man mit Greedy die optimale Lösung nicht trifft.



| n  | Backtracking  |
|----|---------------|
| 12 | 12 s          |
| 13 | 2 min 30 s    |
| 14 | 36 min        |
| 16 | 5 Tage        |
| 18 | 4 Jahre       |
| 20 | 1300 Jahre    |
| 22 | 500.000 Jahre |

Diese Lösungsansätze für das TSP werden in der Vorlesung betrachtet:

**Nearest Neighbour (Greedy)**

**Random Insertion**

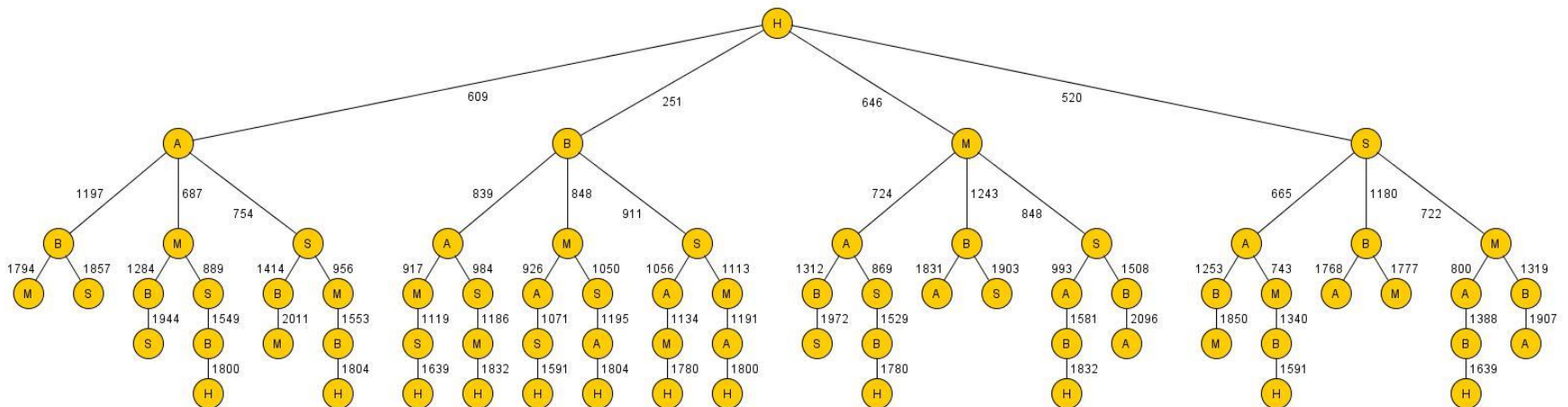
**Minimaler Spannbaum**

**Christofides**

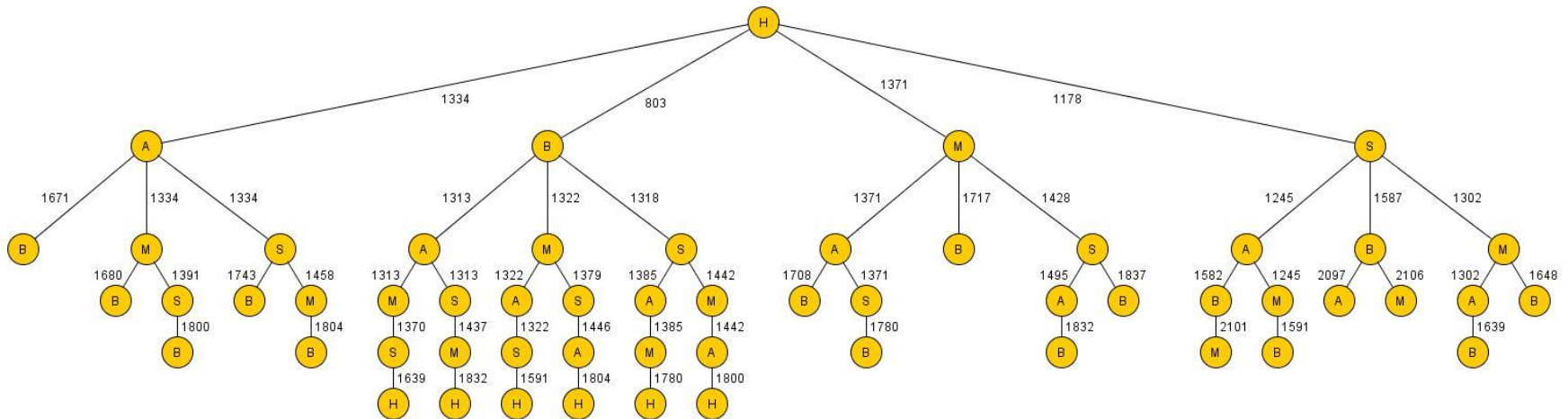
**Brute Force (Backtracking)**

**Branch and Bound**

- Ermitteln einer oberen Schranke: 1639 (Greedy)
- Falls dieser Wert überschritten wird, kann man die Suche auf diesem Pfad abbrechen.
- Bessere Lösungen nutzt man, um die Schranke zu verbessern.
- Folgender Baum muss noch durchlaufen werden:

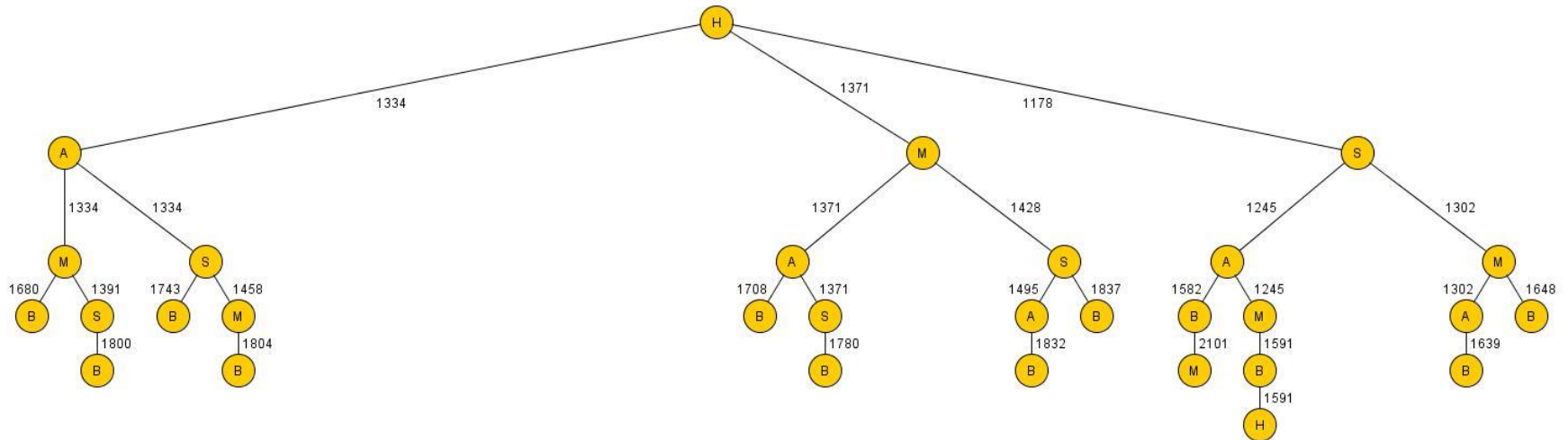


- Bei noch nicht besuchten Städten und der Zielstadt kann man die kleinste Anfahrtstrecke hinzuaddieren.
- Ist die Lösung dann oberhalb der Schranke, kann man bereits abbrechen.
- Folgender Baum muss noch durchlaufen werden:





- Da jede Lösung zweimal vorkommt, kann man weitere Zweige weglassen.
- Z.B.: Man sucht sich die Stadt, die dem Start am nächsten liegt. Diese darf nur in der zweiten Hälfte der Route vorkommen.
- Folgender Baum muss noch durchlaufen werden:



| n  | Backtracking  | B&B 1      | B&B 2      | B&B 2 + Sym |
|----|---------------|------------|------------|-------------|
| 12 | 12 s          | 0,27 s     | 40 ms      | 30 ms       |
| 13 | 2 min 30 s    | 0,8 s      | 0,1 s      | 65 ms       |
| 14 | 36 min        | 4,5 s      | 0,18 s     | 0,12 s      |
| 16 | 5 Tage        | 35 s       | 0,6 s      | 0,32 s      |
| 18 | 4 Jahre       | 8 min      | 6 s        | 3 s         |
| 20 | 1300 Jahre    | 2 h 40 min | 53 s       | 28 s        |
| 22 | 500.000 Jahre | 3 Tage     | 9 min 30 s | 6 min 50 s  |

Diese Entwurfsprinzipien werden in der Vorlesung betrachtet:

**Greedy:** Mache in jedem Teilschritt den größtmöglichen Schritt.

**Backtracking:** Durchlauf eines Lösungsbaums in Pre-Order-Reihenfolge.

**Branch and Bound:** Zweige des Lösungsbaums, die oberhalb einer Schranke liegen, werden abgebrochen.

Divide and Conquer

Dynamische Programmierung

- Das Verfahren „Branch and Bound“ liefert uns das optimale Ergebnis für die Landeshauptstädte: 2922 km.
- Damit können wir folgendes überprüfen:  
Wie viel Glück braucht man, damit Random Insertion das richtige Ergebnis liefert?
- Anders formuliert: Wie viele Versuche benötigt man, um mit 99%er Wahrscheinlichkeit das richtige Ergebnis zu erhalten?

- Bei meiner Testdatei reichten für  $n=22$

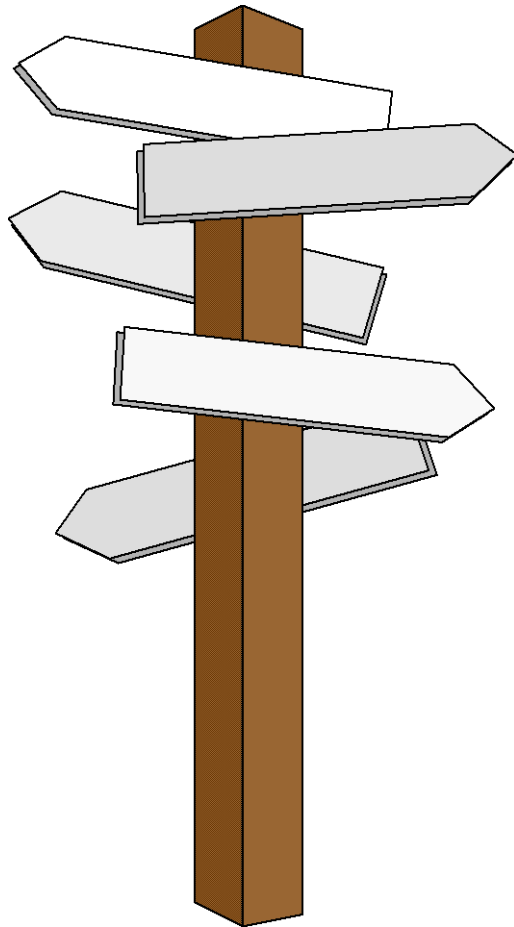
200 Versuche für 99% Wahrscheinlichkeit  $\Rightarrow$  21 ms

400 Versuche für 99,9% Wahrscheinlichkeit  $\Rightarrow$  25 ms

Selbst 1.000.000 Versuche benötigen nur 4 s.

- Random Insertion gibt keine Garantie, ist aber sehr effizient.

(Bei 100.000 Testdurchläufen. Die Testergebnisse geben nur einen groben Richtwert wieder.)



## Vorlesung 14

### Backtracking 1

- **Schreiben Sie eine Funktion**  
`public void printKombinationen(int n)`  
**die alle Zeichenketten auf dem Bildschirm ausgibt, die**
  - n Zeichen lang sind
  - Ausschließlich aus den Buchstaben ‚a‘ bis ‚d‘ bestehen.
- **n ineinandergeschachtelte for-Schleifen? **Geht nicht!****
- **Abbildung auf Baumstruktur**
  - n Ebenen (da n Zeichen)
  - Für jeden Knoten Verzweigungsgrad=4 (nächstes Zeichen a,b,c,d)
- **Pre-Order-Durchlauf des (gedachten) Baums: Backtracking**

- **Die Aufgabe wird erweitert:**
  - Die Ergebnisse sollen in einer ArrayList zusammengefasst und zurückgegeben werden.
- **5 mögliche Verfahren**
  - Klassisches Backtracking
  - Backtracking mit Korb
  - Level-Order mit Queue
  - Codierung
  - evtl. dynamische Programmierung
- **Ziel:**
  - Sie können mindestens „Backtracking mit Korb“ und „Level-Order mit Queue“ anwenden.

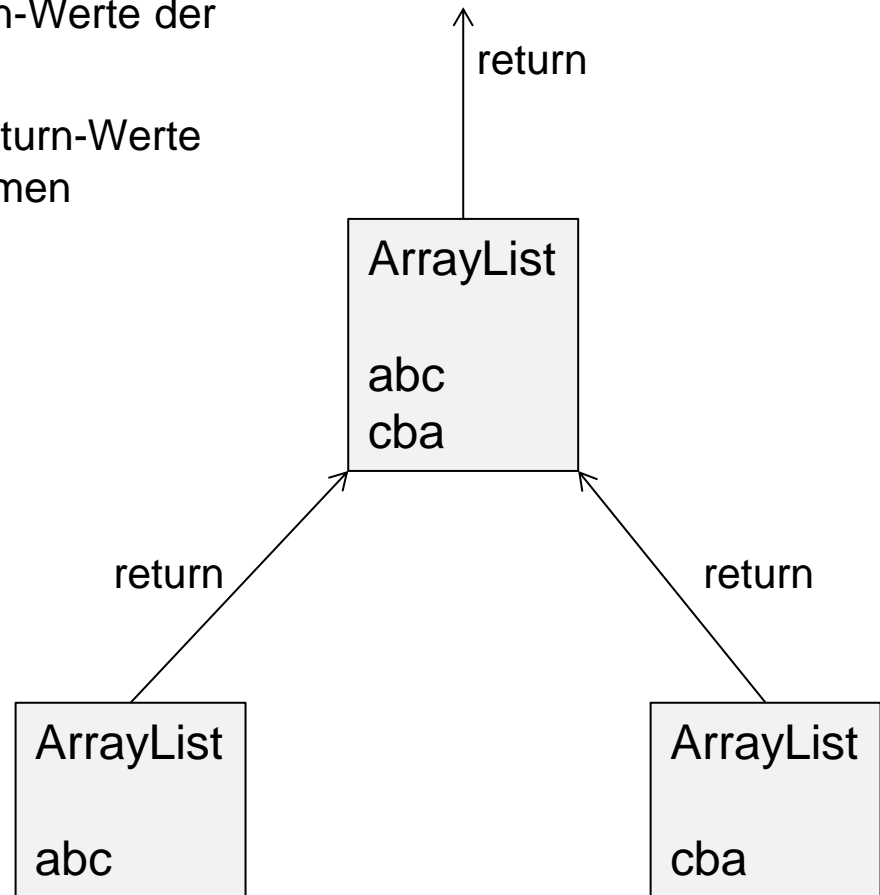


## Klassisches Backtracking

Die Ergebnisse werden über die return-Werte der Rekursionsaufrufe zurückgegeben.

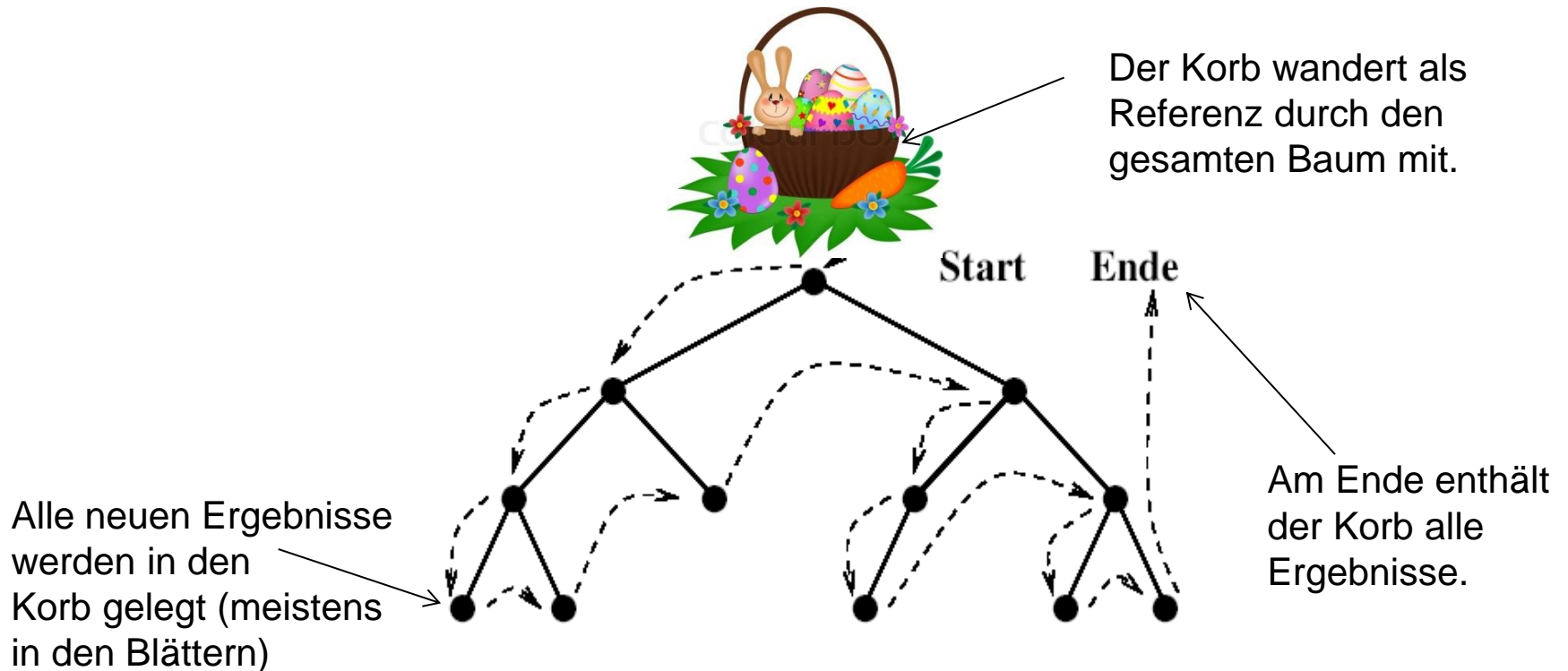
Jeder Rekursionsschritt sammelt die return-Werte der Unterschritte und fasst sie zusammen

(in Java z.B. mit `addAll`).



## Backtracking mit Korb

Eine Korb (z.B. ArrayList) wird per Übergabeparameter (als Referenz) mitgegeben. Die Ergebnisse werden in dieser Liste gesammelt.



## Level-Order mit Queue

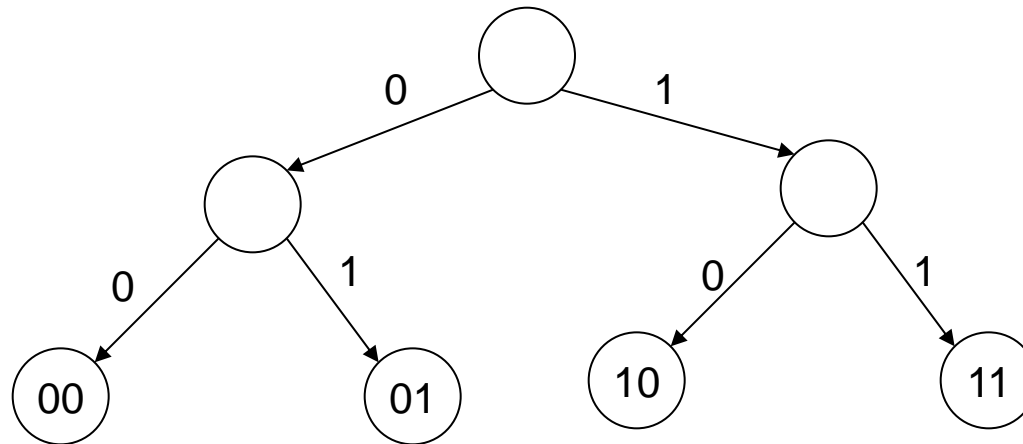
Kein Backtracking, keine Rekursion, nur Schleifen.

- Der Backtracking-Baum wird in Level-Order durchlaufen → Benutzung einer Queue.
- Beispiel für `printKombinationen`:
- 1. Initialisierung: Leerstring wird in die Queue gesteckt.
- 2. Durchlauf (Beispiel): „ac“ wird aus der Queue geholt
  - Ergänzte Strings „aca“ bis „acd“ in Queue stecken.
- 3. Sammeln der Ergebnisse
  - String mit Länge n wird aus Queue geholt → String ist Ergebnis. Ergebnis in Ergebnis-Liste aufnehmen (kann lokale Variable sein, da keine Rekursion vorhanden).
- 4. Abbruch
  - Wenn die Queue leer ist, wird die Ergebnisliste zurückgegeben.

## Codierung

Kein Backtracking, keine Rekursion, nur Schleifen.

- Jedem Blatt eines Baums mit Verzweigungsgrad  $n$  kann man eine Codenummer im Zahlensystem zur Basis  $n$  zuordnen.



- Beispiel: Alle Blätter sind mit einem Code zwischen 00 (dez. 0) und 11 (dez. 3) versehen.

## Codierung

Beispiel für `printKombinationen` ( $n=3$ ):

Alle Blätter sollen eine Codezahl erhalten.

- Die Codezahl ist dreistellig, denn die Blätter liegen in der dritten Ebene
- Die Codezahl steht im 4er-System, denn der Verzweigungsgrad ist 4.

Schleife von  $0_{\text{dez}}$  (Code 000) bis  $4^3 - 1 = 63_{\text{dez}}$  (Code 333)

Dezimalzahl ins 4er-System umrechnen und in String wandeln: ( $0 \rightarrow a$ ,  $1 \rightarrow b$ ,  $2 \rightarrow c$ ,  $3 \rightarrow d$ )

String in Ergebnis-Liste aufnehmen.

## Dynamische Programmierung

Definition aus Wikipedia:

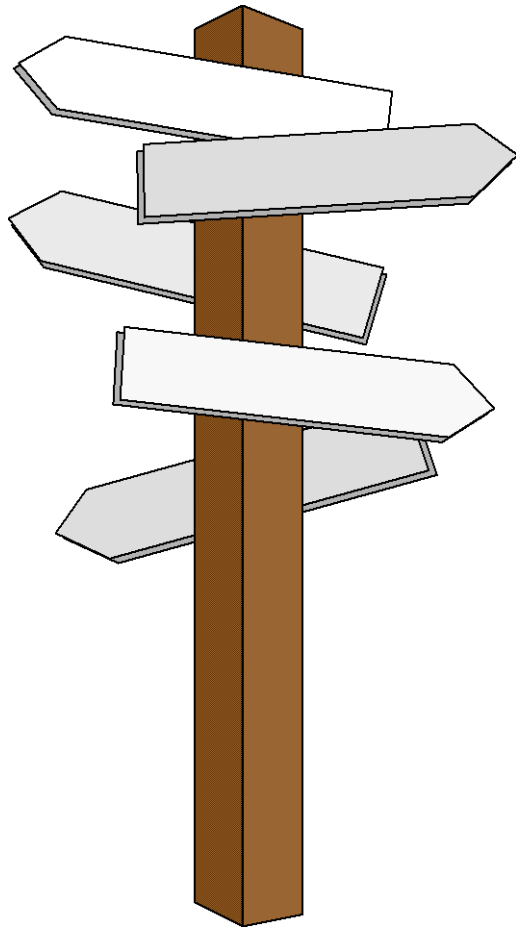
**In der dynamischen Programmierung werden zuerst die optimalen Lösungen der kleinsten Teilprobleme direkt berechnet und dann geeignet zu einer Lösung eines nächstgrößeren Teilproblems zusammengesetzt. Dieses Verfahren setzt man fort...** (Die Definition ist nicht klausurrelevant. Schauen Sie den Rest bei Interesse in Wikipedia nach.)

Dynamische Programmierung führt oft zu sehr schnellen Algorithmen

Es ist sehr schwierig, selbst Algorithmen zu entwickeln, die dynamische Programmierung benutzen.

Dies ist auch nicht Bestandteil der Vorlesung. Wir werden jedoch zwei Algorithmen betrachten, die dynamische Programmierung verwenden.

- den Bellman-Ford-Algorithmus (Vorlesung 15)
- den Floyd-Warshall-Algorithmus (Vorlesung 16)



## Vorlesung 15

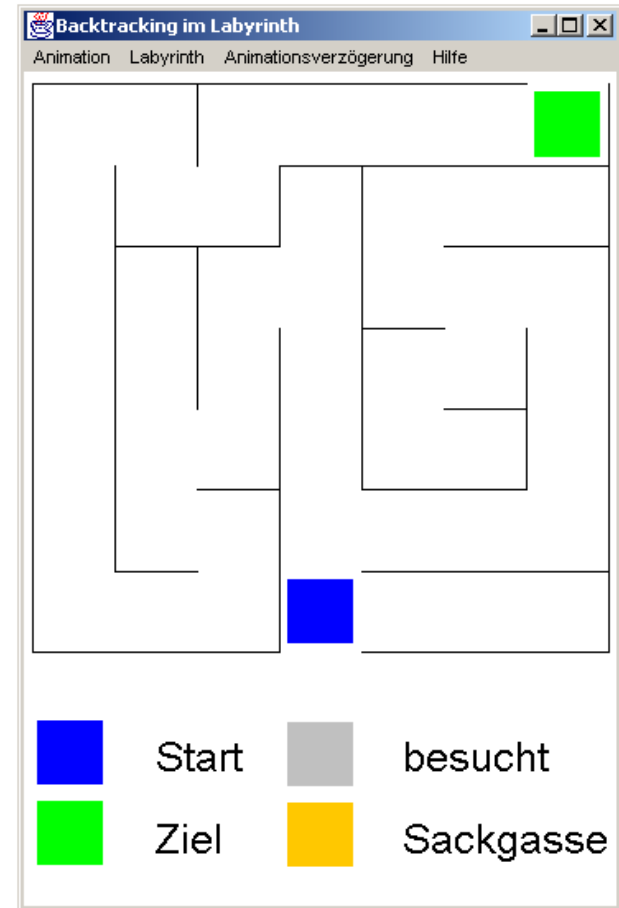
### Backtracking 2

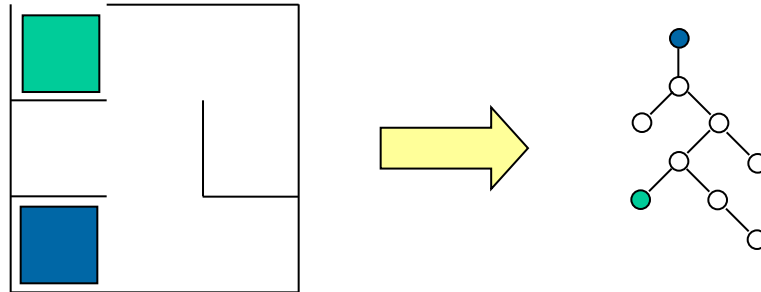
- Situation: Mehrere Alternativen sind in bestimmten Schritten des Algorithmus möglich
  - z.B. 4 Richtungen bei Suche in einem Labyrinth:  $\rightarrow, \downarrow, \uparrow, \leftarrow$
- Lösung mit Backtracking:
  - Wähle eine Alternative und verfolge diesen Weg weiter.
  - Falls man so eine Lösung des Problems findet, ist man fertig.
  - Ansonsten gehe einen **Schritt zurück** und verfolge **rekursiv eine** andere (bisher noch nicht probierte) Alternative in diesem Schritt.
  - Falls alle Alternativen erfolglos probiert wurden: Einen Schritt zurückgehen ...



# Beispiel: Labyrinthsuche

- „Wie kommt die (virtuelle) Maus zum Käse?“
- Bei Abzweigungen durchsucht sie erst die eine Richtung. Führt sie in eine Sackgasse, kehrt sie an die Abzweigung zurück (Backtracking) und wählt eine andere Richtung.
- Rekursiver Algorithmus.
- Bild aus:  
<http://www.swisseduc.ch/informatik/vortraege/backtracking/index.html>



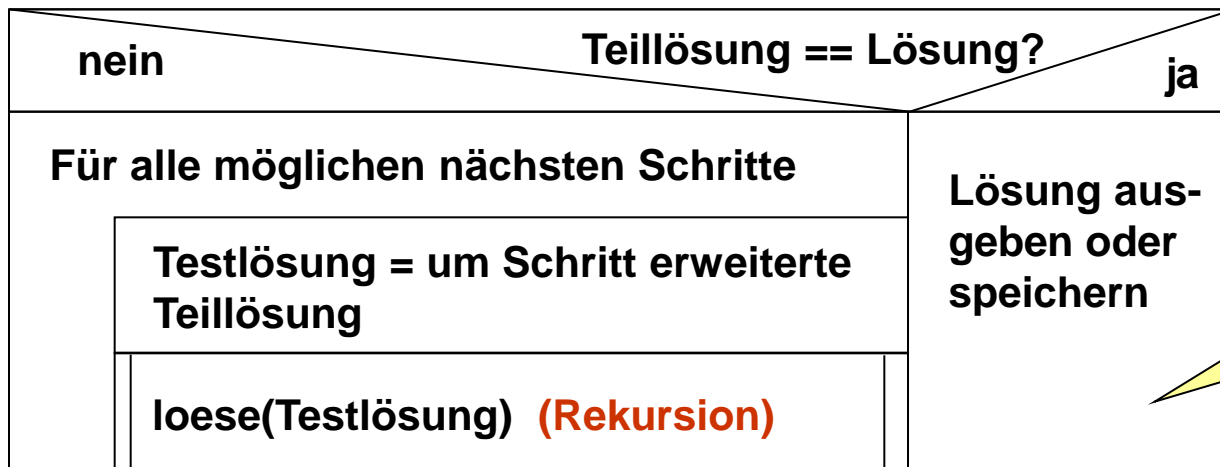


Baumdurchlauf (mit Pre-Order) und Suchen eines bestimmten Knotens im Baum.

Das „Hochlaufen“ des Baumes (nach Beendigung eines Rekursionsaufrufs) entspricht dem Zurückgehen eines Schritts im Labyrinth.

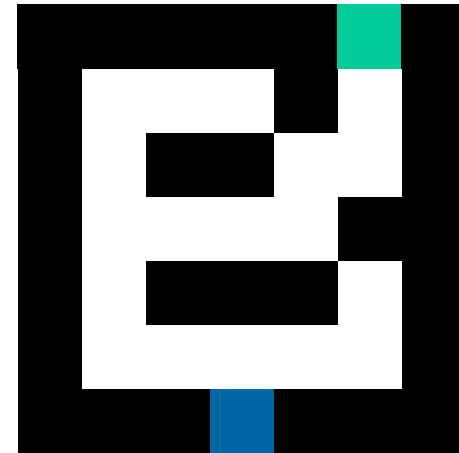
Daher hat diese Suche den Namen **Backtracking**.

**loese(Teillösung)**



**Sucht anschließend weiter. Abbruch??**

- Wände bestehen aus einem Block
- Ränder sind Wände (außer Start und Ziel)  
⇒ Erster Schritt immer eindeutig
- Abbruch der Suche nach Erreichen des Ziels
  
- Klarheit des Codes geht vor Performance
- Objektorientiertheit



## Location

```
- x, y: int
+Location(x: int, y: int)
+add(loc: Location): Location
+equals(loc: Location): boolean
```

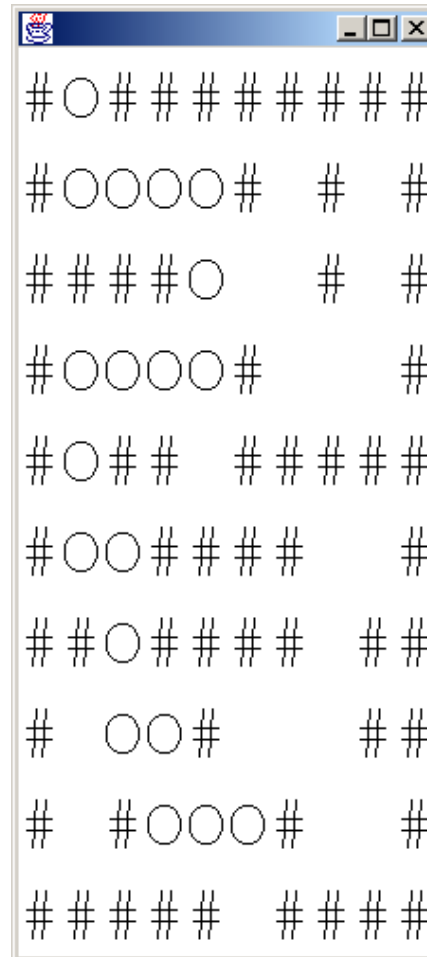
## Backtracking

```
+Backtracking(m: Maze)
+solve(loc: Location,
 lastLoc: Location): boolean
+static main(String[] args): int
```

## Maze

```
+Maze(int width, int height)
+isFree(loc: Location): boolean
+isEnd(loc: Location): boolean
+getStart(): Location
+getFirstStep(): Location
+markTrack(loc: Location,
 vis: boolean): void
+print(): void
```

(Code ist auf Kursseite verfügbar.)

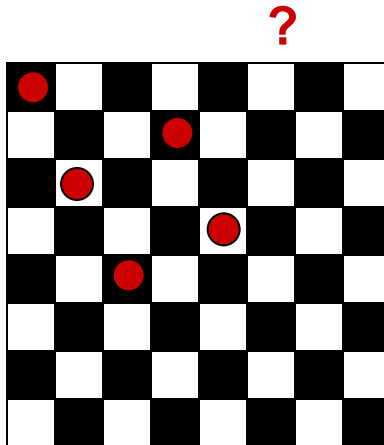




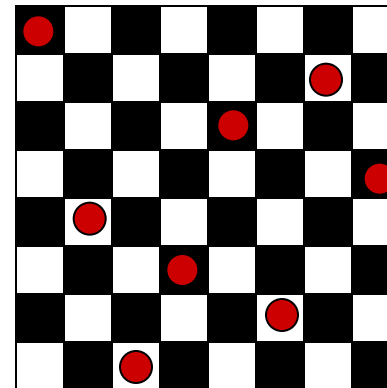
**Aufgabenstellung:** 8 Damen auf Schachbrett so aufstellen, dass sie sich gegenseitig nicht schlagen können (In keiner Diagonalen, Waagerechten oder Senkrechten darf mehr als eine Dame stehen)

Eng verwandt:  
"Springerproblem"

Sackgasse



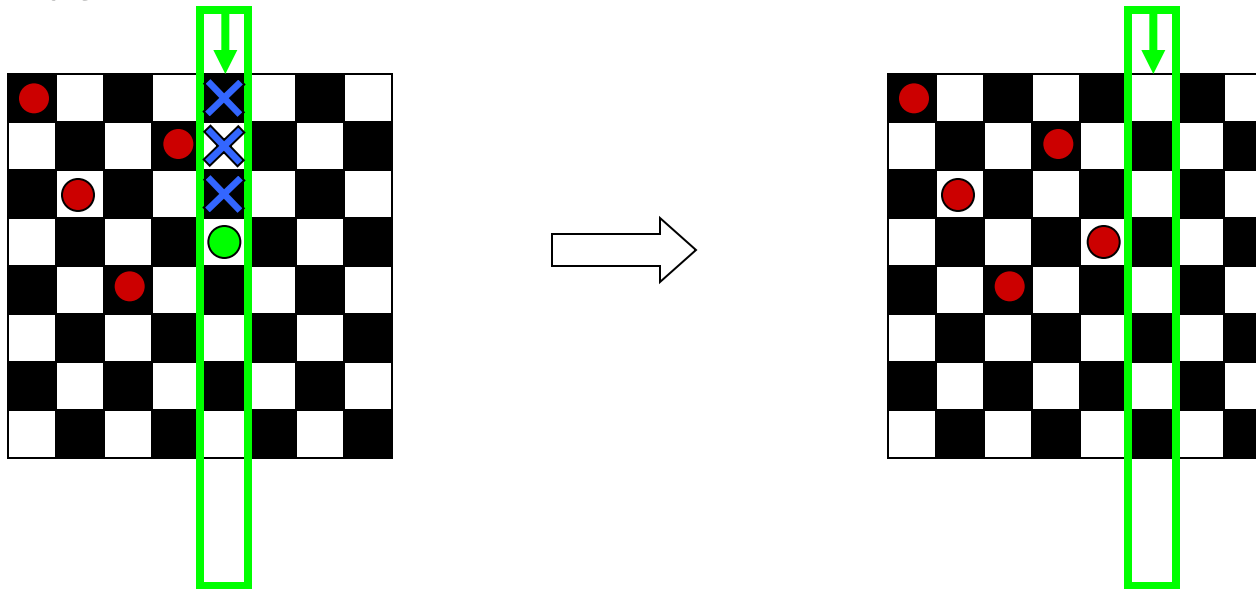
eine Lösung



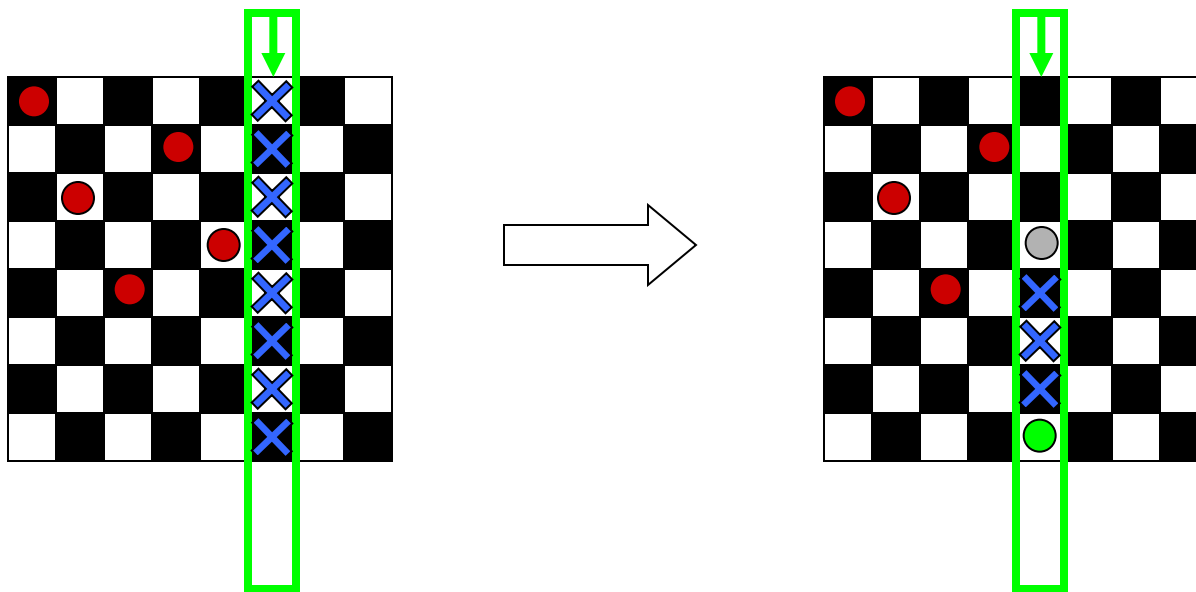


Teile Problem in 8 Schritte auf:

- In Schritt  $i$ : eine Dame in  $i$ -ter Spalte platzieren unter Berücksichtigung der bereits platzierten  $i-1$  Damen. Dazu: Platzierung in Zeilen  $j=1-8$  ausprobieren: Erste Zeile wählen, in der die  $i$ -te Dame nicht „bedroht“ ist.
- Falls eine sichere Position gefunden:  $(i+1)$ -te Dame genauso behandeln.



- Falls sichere Platzierung in **keiner Zeile möglich**: einen **Schritt zurück** gehen und alternative Position für  $(i-1)$ -te Dame wählen.
- Verfahren funktioniert rekursiv



Array `queenpos[8]`: Positionen der Damen, d.h.

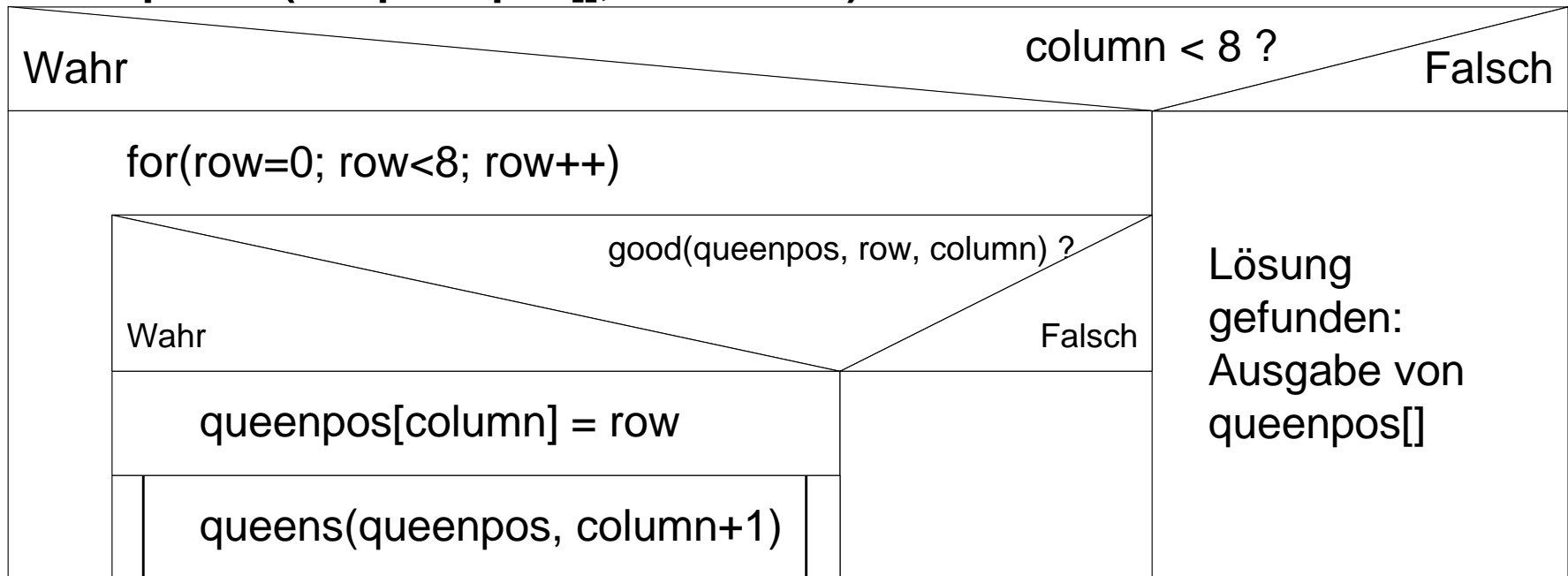
`queenpos[i]` = Nummer der Zeile, in der  $i$ -te Dame in Spalte  $i$  platziert wird.

# Realisierung: 8-Damen-Problem

Aufruf:

```
queens(queenpos,0)
```

**queens(int queenpos[], int column):**



Bemerkung: Verfahren findet alle Lösungen

**boolean good(int queenpos[], int row, int column):**

isGood = true

for(i=0; (i<column) && isGood; i++)

Dame in Spalte i kann Dame an Position row/  
column schlagen

Wahr

Falsch

isGood = false

return isGood

- Backtracking-Algorithmen können **exponentiellen Aufwand** haben!
- Durch Einführung von Zusatzbedingungen möglichst viele Sackgassen ausschließen.
- Symmetriebedingungen ausnutzen.

**Gegeben:** Es gibt  $n$  Gegenstände. Jeder Gegenstand  $i$  hat ein Gewicht  $g_i$  und einen Wert  $w_i$ .

In einen Rucksack passen Gegenstände bis zum Gesamtgewicht  $G$ .

**Aufgabenstellung:** Packe Gegenstände so in den Rucksack, dass der Gesamtwert maximal wird.

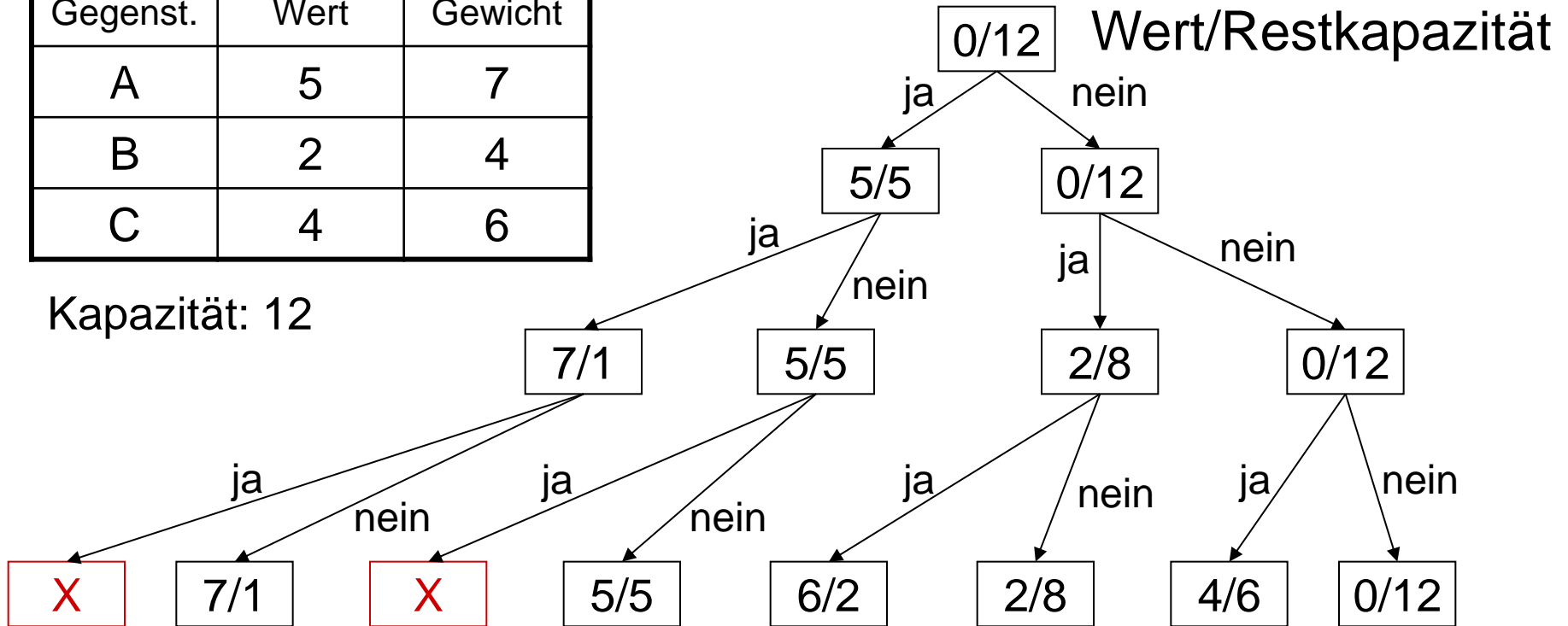
$$\left( \sum_{c_i \in \{0,1\}} c_i g_i < G \right) \wedge \left( \sum_{c_i \in \{0,1\}} c_i w_i \text{ maximal} \right)$$

**Komplexität:** mit Binärbedingung:  $2^n$

# Lösung mit Baumdurchlauf

| Gegenst. | Wert | Gewicht |
|----------|------|---------|
| A        | 5    | 7       |
| B        | 2    | 4       |
| C        | 4    | 6       |

Kapazität: 12



Suche des Blattes mit dem größten Wert

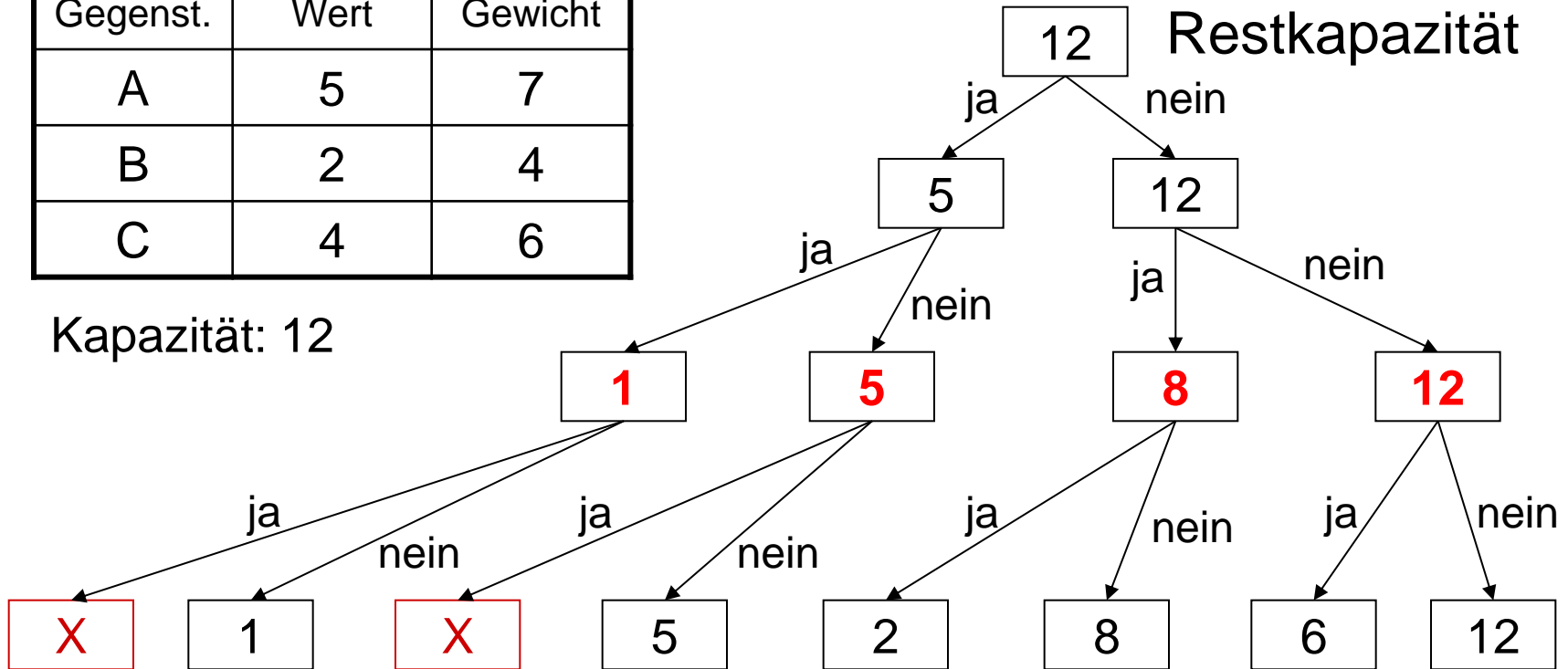


- Der Baumdurchlauf hat die Komplexität  $O(2^n)$ .
  - Ab  **$n=10$**  ist mit **Performance-Problemen** zu rechnen.
  
- Alternative Lösung: Dynamische Programmierung
  - **Laufzeitkomplexität:  $O(n^2)$ .**

# Lösung mit dynamischer Programmierung (2)

| Gegenst. | Wert | Gewicht |
|----------|------|---------|
| A        | 5    | 7       |
| B        | 2    | 4       |
| C        | 4    | 6       |

Kapazität: 12



Achte in der vorletzten Zeile auf die Restkapazitäten (rot). Sie müssen zwischen 0 und der Maximalkapazität des Rucksacks liegen.

1

5

8

12

Die roten Einträge sind die Restkapazität vor der Entscheidung über den letzten Gegenstand. An ihnen erkennt man, wie die letzte Entscheidung ausfällt. Wenn für den letzten Gegenstand Platz ist, wird er mitgenommen, ansonsten nicht.

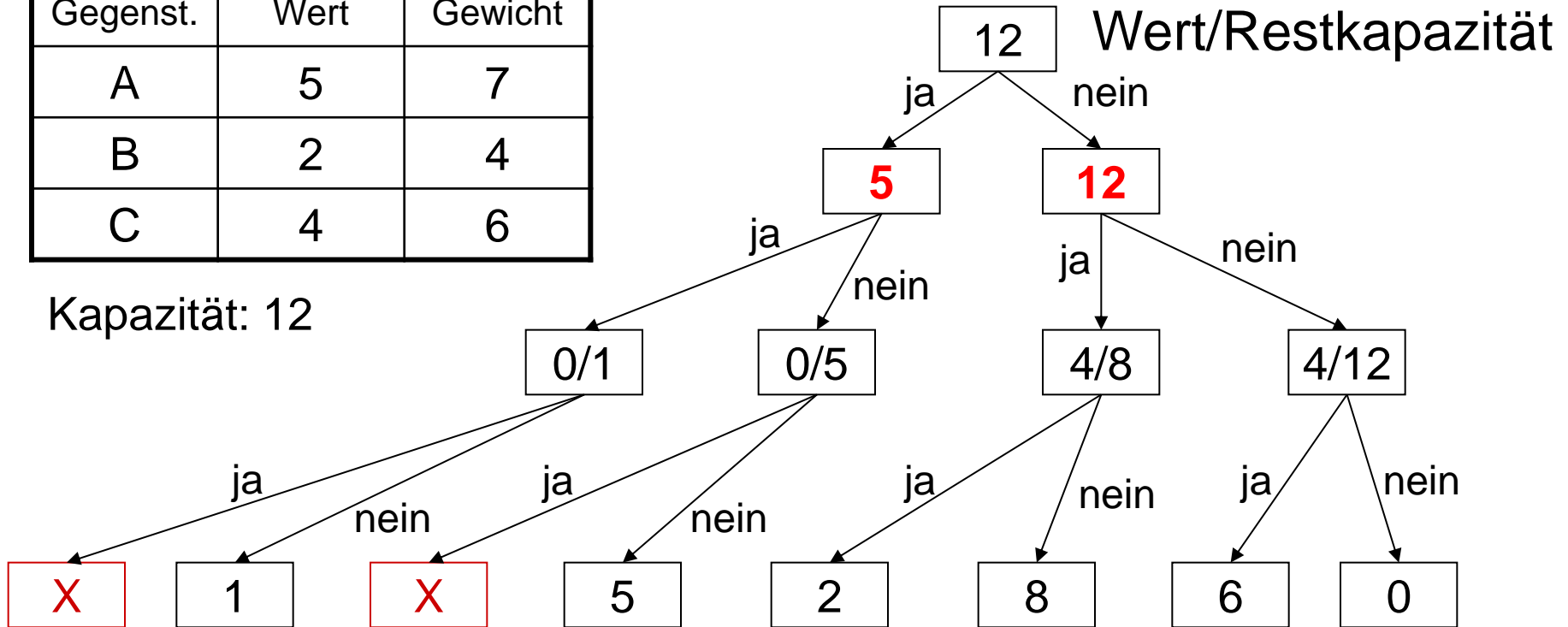
Man kann das in Tabellenform ausdrücken: Hat der letzte Gegenstand Gewicht 6 und Wert 4, so gilt:

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 4        | 4        | 4        | 4        | 4         | 4         | 4         |
| Gegenstand        | -        | -        | -        | -        | -        | -        | C        | C        | C        | C        | C         | C         | C         |

# Lösung mit dynamischer Programmierung (4)

| Gegenst. | Wert | Gewicht |
|----------|------|---------|
| A        | 5    | 7       |
| B        | 2    | 4       |
| C        | 4    | 6       |

Kapazität: 12



Nun arbeiten wir uns Zeile um Zeile nach oben vor. Welchen Wert kann man abhängig von den Restkapazitäten in der 2. Zeile (rot) mitnehmen?

5

12

Gegenstand B hat Gewicht 4 und Wert 2. Wenn man ihn nicht einpackt, gilt für den letzten Gegenstand exakt die Tabelle aus dem letzten Schritt.

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 4        | 4        | 4        | 4        | 4         | 4         | 4         |
| Gegenstand        | -        | -        | -        | -        | -        | -        | C        | C        | C        | C        | C         | C         | C         |

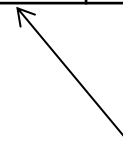
Wenn man ihn einpackt, erhöht sich der einpackbare Wert um 2 und die Restkapazität verringert sich um 4.  
 -> Vorgehen auf der nächsten Folie.

*Gegenstand B hat Gewicht 4 und Wert 2.*

In eine neue Tabelle trägt man den Wert 2 ein, falls B Platz hat. Die alte Tabelle verschiebt man 4 Positionen (Gewicht von B) nach rechts.

|                   |          |          |          |          |          |          |          |          |          |          |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 4        | 4        | 4        | 4        | 4         |
| Gegenstand        | -        | -        | -        | -        | -        | -        | C        | C        | C        | C        | C         |

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 2        | 2        | 2        | 2        | 2         | 2         | 2         |
| Gegenstand        | -        | -        | -        | -        | B        | B        | B        | B        | B        | B        | B         | B         | B         |



Neue Tabelle

Nun kann man in der alten Tabelle (oben) ablesen, was nach dem Einpacken von B noch Platz hat. Dann addiert man die obere zur unteren Tabelle. Jetzt hat man den erreichbaren Wert, wenn man Gegenstand B einpackt.

|                   |          |          |          |          |          |          |          |          |          |          |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 4        | 4        | 4        | 4        | 4         |
| Gegenstand        | -        | -        | -        | -        | -        | -        | C        | C        | C        | C        | C         |

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 2        | 2        | 2        | 2        | 2         | 2         | 2         |
| Gegenstand        | -        | -        | -        | -        | B        | B        | B        | B        | B        | B        | B         | B         | B         |

---

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 2        | 2        | 2        | 2        | 6         | 6         | 6         |
| Gegenstand        | -        | -        | -        | -        | B        | B        | B        | B        | B        | B        | BC        | BC        | BC        |

Nun rückt man beide Tabellen wieder untereinander und vergleicht: Ist es besser, den zweiten Gegenstand einzupacken oder nicht?

|   |                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|---|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| n | Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
|   | Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 4        | 4        | 4        | 4        | 4         | 4         | 4         |
|   | Gegenstand        | -        | -        | -        | -        | -        | -        | C        | C        | C        | C        | C         | C         | C         |

|   |                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|---|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| j | Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
|   | Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 2        | 2        | 2        | 2        | 6         | 6         | 6         |
|   | Gegenstand        | -        | -        | -        | -        | B        | B        | B        | B        | B        | B        | BC        | BC        | BC        |

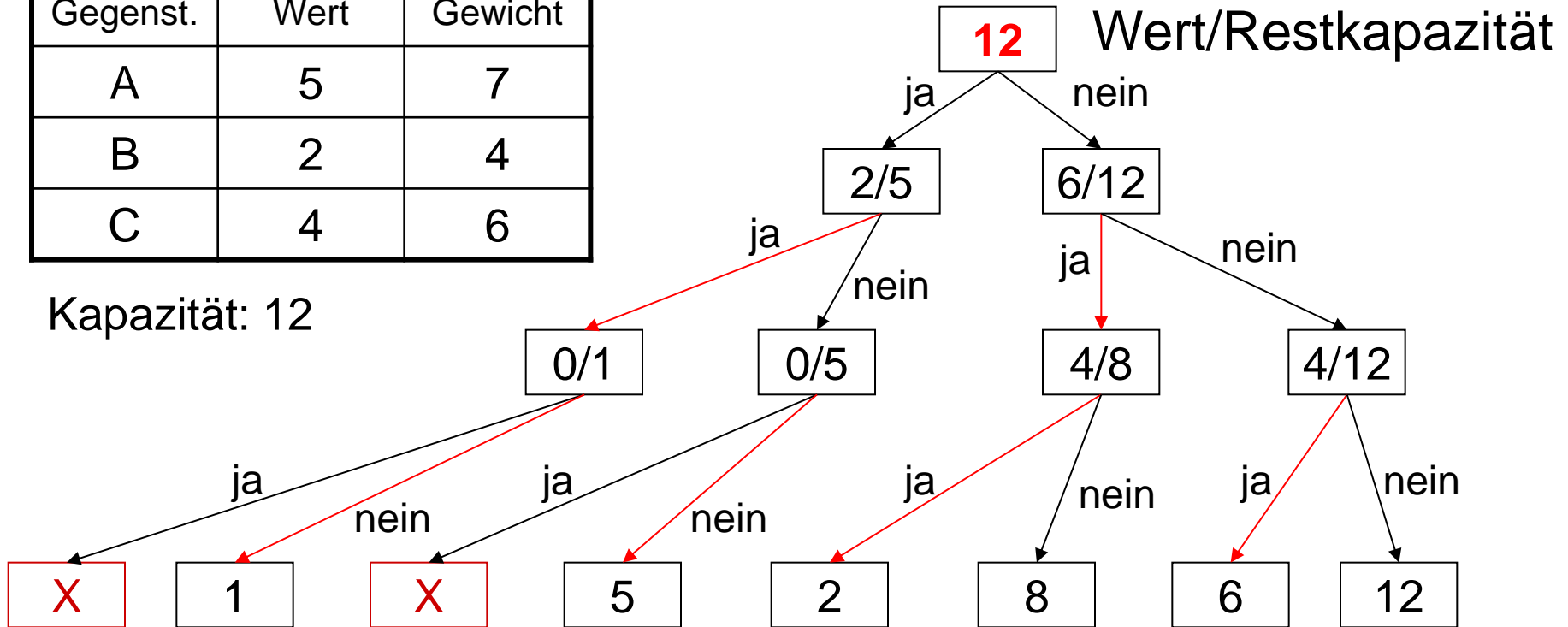
|  |                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|--|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
|  | Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
|  | Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 4        | 4        | 4        | 4        | 6         | 6         | 6         |
|  | Gegenstand        | -        | -        | -        | -        | B        | B        | C        | C        | C        | C        | BC        | BC        | BC        |



# Lösung mit dynamischer Programmierung (9)

| Gegenst. | Wert | Gewicht |
|----------|------|---------|
| A        | 5    | 7       |
| B        | 2    | 4       |
| C        | 4    | 6       |

Kapazität: 12



In der nächsten Zeile sind wir oben angelangt. Eigentlich interessiert nur noch der Eintrag Nr. 12. Aus Übungsgründen berechnen wir trotzdem noch einmal die ganze Zeile.

Gegenstand A: Wert 5, Gewicht 7.

Nicht einpacken: Letzte Tabelle übernehmen:

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 4        | 4        | 4        | 4        | 6         | 6         | 6         |
| Gegenstand        | -        | -        | -        | -        | B        | B        | C        | C        | C        | C        | BC        | BC        | BC        |

*Gegenstand A: Wert 5, Gewicht 7.*

Einpacken: Neue Tabelle erstellen, alte nach rechts rücken und addieren.

|                   |          |          |          |          |          |          |          |          |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 4        | 4        |
| Gegenstand        | -        | -        | -        | -        | B        | B        | C        | C        |

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 5        | 5        | 5        | 5         | 5         | 5         |
| Gegenstand        | -        | -        | -        | -        | -        | -        | -        | A        | A        | A        | A         | A         | A         |

|                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
| Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 5        | 5        | 5        | 5         | 7         | 7         |
| Gegenstand        | -        | -        | -        | -        | -        | -        | -        | A        | A        | A        | A         | AB        | AB        |

Gegenstand A: Wert 5, Gewicht 7.

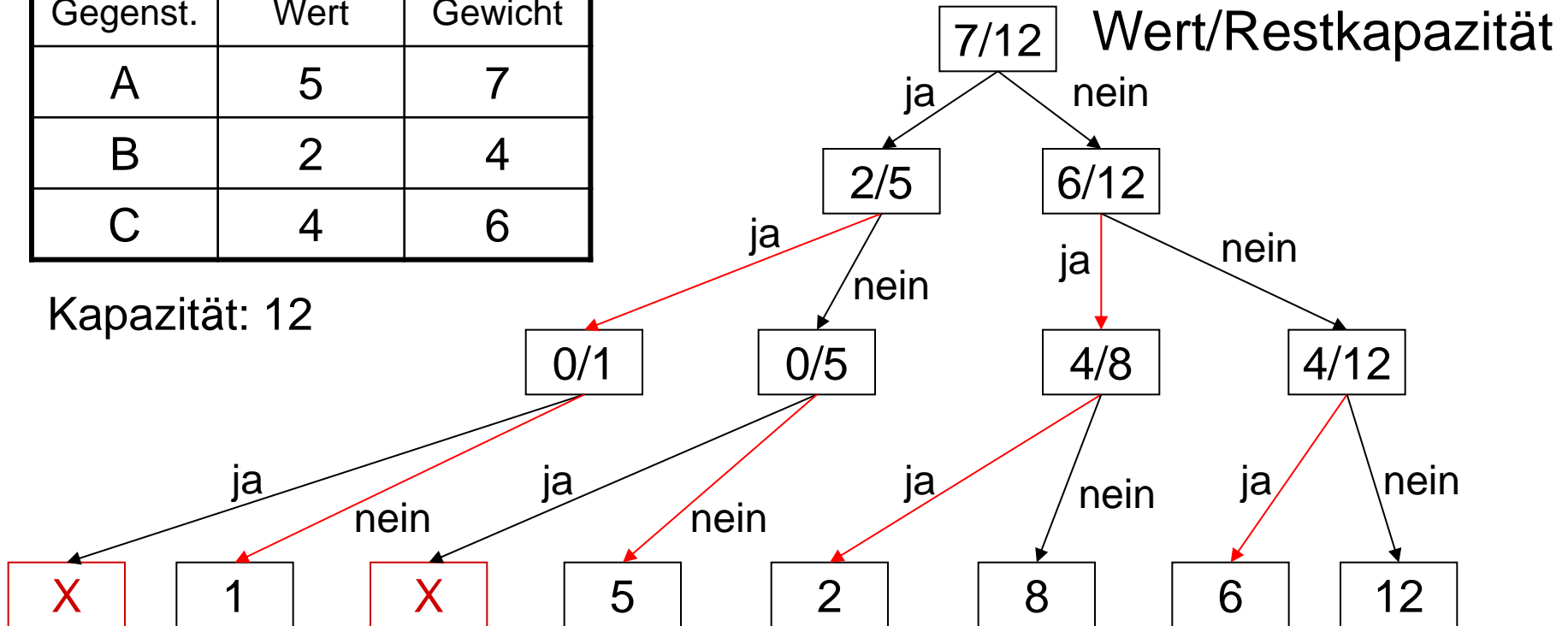
Zurückrücken und vergleichen. Am besten ist es, A und B einzupacken.

|   |                   |          |          |          |          |          |          |          |          |          |          |           |           |           |
|---|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| n | Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
|   | Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 4        | 4        | 4        | 4        | 6         | 6         | 6         |
|   | Gegenstand        | -        | -        | -        | -        | B        | B        | C        | C        | C        | C        | BC        | BC        | BC        |
| j | Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
|   | Einpackbarer Wert | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 5        | 5        | 5        | 5         | 7         | 7         |
|   | Gegenstand        | -        | -        | -        | -        | -        | -        | -        | A        | A        | A        | A         | AB        | AB        |
|   | Restkapazität     | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> | <b>11</b> | <b>12</b> |
|   | Einpackbarer Wert | 0        | 0        | 0        | 0        | 2        | 2        | 4        | 5        | 5        | 5        | 6         | 7         | 7         |
|   | Gegenstand        | -        | -        | -        | -        | B        | B        | C        | A        | A        | A        | BC        | AB        | AB        |

# Lösung mit dynam. Programmierung (13)

| Gegenst. | Wert | Gewicht |
|----------|------|---------|
| A        | 5    | 7       |
| B        | 2    | 4       |
| C        | 4    | 6       |

Kapazität: 12



Da die Wurzel immer den Wert der kompletten Restkapazität hat, kann das Endergebnis aus der letzten Tabelle direkt ablesen.

- Vorteil: Der Aufwand bleibt für jede Ebene immer gleich  
->  $O(n^2)$
- Bei Backtracking wächst der Aufwand für jede Ebene um das Doppelte ->  $O(2^n)$ .

Diese Entwurfsprinzipien werden in der Vorlesung betrachtet:

**Greedy:** Mache in jedem Teilschritt den größtmöglichen Schritt.

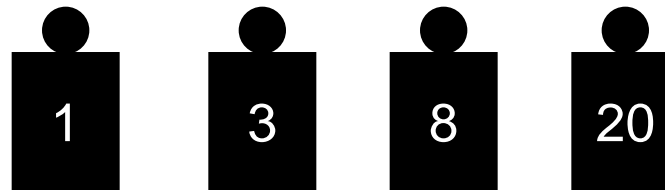
**Backtracking:** Durchlauf eines Lösungsbaums in Pre-Order-Reihenfolge.

**Branch and Bound:** Zweige des Lösungsbaums, die oberhalb einer Schranke liegen, werden abgebrochen.

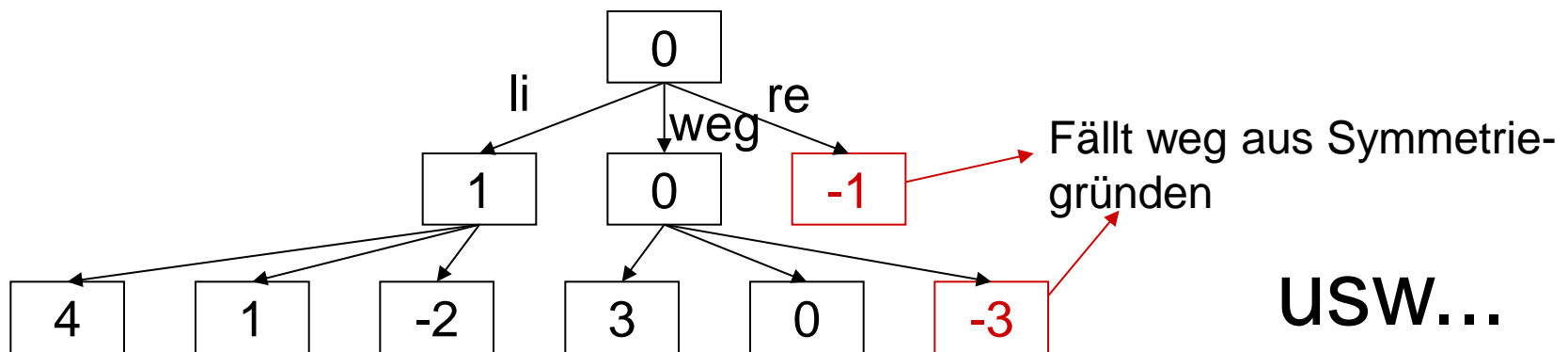
**Divide and Conquer**

**Dynamische Programmierung:** Der Lösungsbaum wird ebenelementweise durchlaufen. Die Zwischenergebnisse werden in einer Tabelle gespeichert.

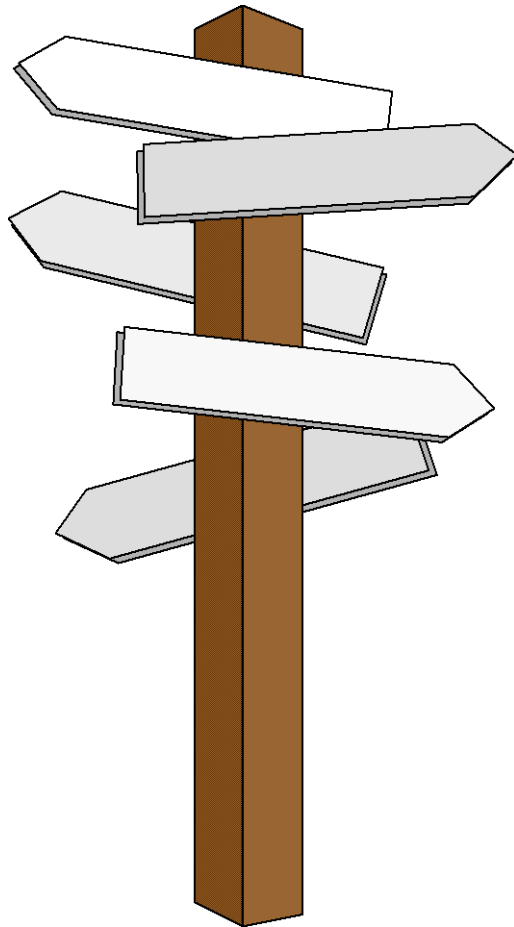
- Zur Verfügung steht eine Balkenwaage und 4 Gewichte.



- Welche Gewichte kann man ausmessen?







## Vorlesung 16

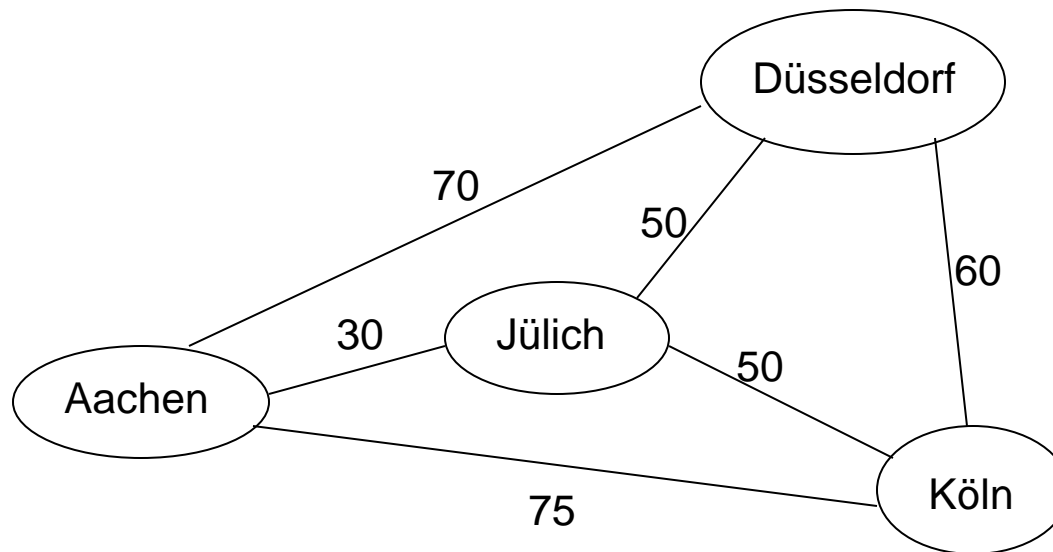
### Der Dijkstra-Algorithmus

- Günstigste Wege von einer Quelle: **Dijkstra-Algorithmus**
- Günstigste Wege zwischen allen Knotenpaaren: **Floyd-Algorithmus**
- Existiert ein Weg zwischen zwei Knoten? **Warshall-Algorithmus**

- Eigentlich: Suche nach günstigsten Wegen in *gewichteten* Graphen: Gewichte  $\cong$  Kosten;
- Bei Anwendung auf *ungewichtete* Graphen (Gewichte=1) ergibt sich: „kürzeste“ Wege.
- Praktische Bedeutung für Transport- und Kommunikationsnetzwerke
- (einige) Problemarten:
  1. Bestimme günstigsten Weg **von festem Knoten** („Quelle“) zu **allen anderen Knoten** („Single-source shortest-path problem“).
  2. Bestimme günstigste Wege **zwischen allen Knotenpaaren** („All-pairs shortest-path problem“).
  3. Bestimme günstigsten Weg **zwischen zwei gegebenen Knoten**.
  4. **Existiert ein Weg** zwischen zwei Knoten?

Spedition fährt täglich im Dreieck Aachen-Düsseldorf-Köln und möchte die **Kosten minimieren**.

Kantenkosten im Graphen sind z.B. Kilometer oder benötigte Zeit.



- 1) Kostenminimaler Weg von Aachen zu allen anderen Orten?
- 2) Kostenminimaler Weg zwischen allen Orten?
- 3) Kostenminimaler Weg zwischen Aachen und Köln?

## Algorithmus von Dijkstra (1959):

Gegeben: Graph  $G = (V, E)$  dessen Bewertungsfunktion folgende Eigenschaften hat:

- Jede **Kante** von  $v_i$  nach  $v_j$  hat nicht negative Kosten:  $C(i, j) \geq 0$
- Falls **keine Kante** zwischen  $v_i$  und  $v_j$  :  $C(i, j) = \infty$
- **Diagonalelemente**:  $C(i, i) = 0$  (spezielle Konvention, falls nötig)
- Bei **ungewichtetem Graph** (Gewichte=1):  
Ergebnis = Länge der kürzesten Wege

Idee: Iterative Erweiterung einer Menge von günstig erreichbaren Knoten

Menge  $S$ : die Knoten, deren günstigste Wegekosten von der vorgegebenen Quelle (Startknoten) bereits bekannt sind.

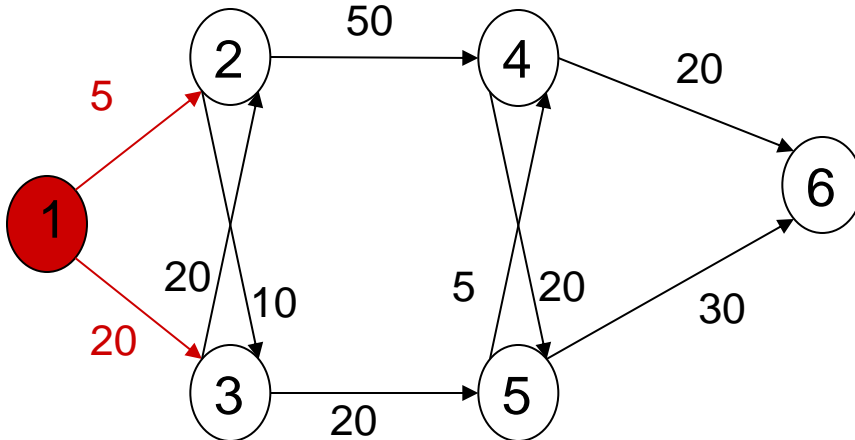
1. Initialisierung:  $S = \{ \text{Startknoten} \}$
2. Beginnend mit Quelle alle ausgehenden Kanten betrachten (analog BFS). Nachfolgerknoten  $v$  mit günstigster Kante zu  $S$  hinzunehmen.
3. Jetzt: berechnen, ob die Knoten in  $V \setminus S$  günstiger über  $v$  als „Zwischenstation“ erreichbar ist, als ohne Umweg über  $v$ .
4. Danach: denjenigen Knoten  $v'$  zu  $S$  hinzunehmen, der nun am günstigsten zu erreichen ist. Bei zwei gleich günstigen Knoten wird ein beliebiger davon ausgewählt.

Neuberechnen der Wegekosten (3.) und Erweitern von  $S$  (4.) wiederholen, bis alle Knoten in  $S$  sind.

## Beispiel (1)

Initialisierung:  $S = \{ \text{Startknoten} \}$

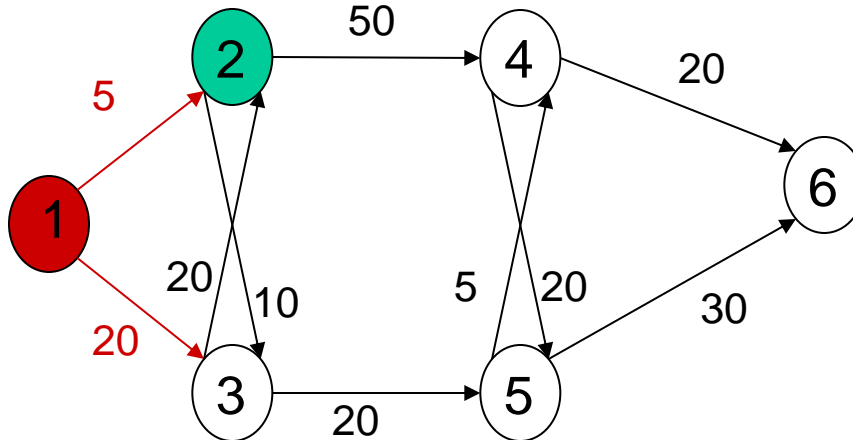
Beginnend mit Quelle alle ausgehenden Kanten betrachten (analog BFS).



| Knoten | Abstand |
|--------|---------|
| 1      | 0       |
| 2      | 5       |
| 3      | 20      |
| 4      |         |
| 5      |         |
| 6      |         |

## Beispiel (2)

Nachfolgerknoten  $v$  mit günstigster Kante zu  $S$  hinzunehmen.

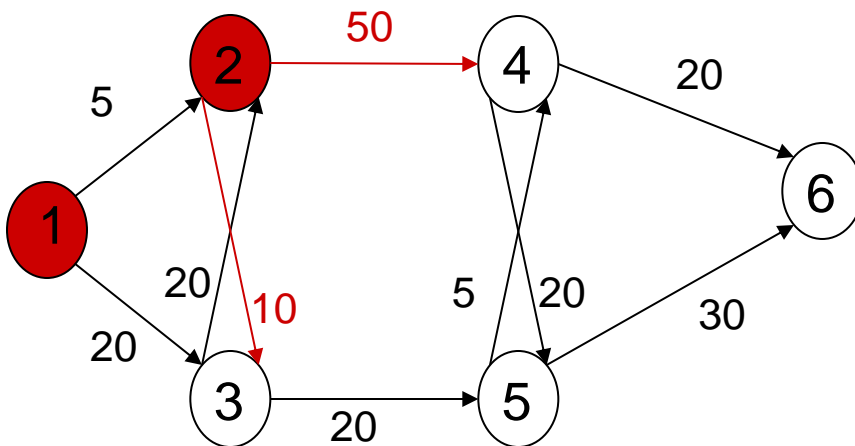


| Knoten | Abstand |
|--------|---------|
| 1      | 0       |
| 2      | 5       |
| 3      | 20      |
| 4      |         |
| 5      |         |
| 6      |         |



## Beispiel (3)

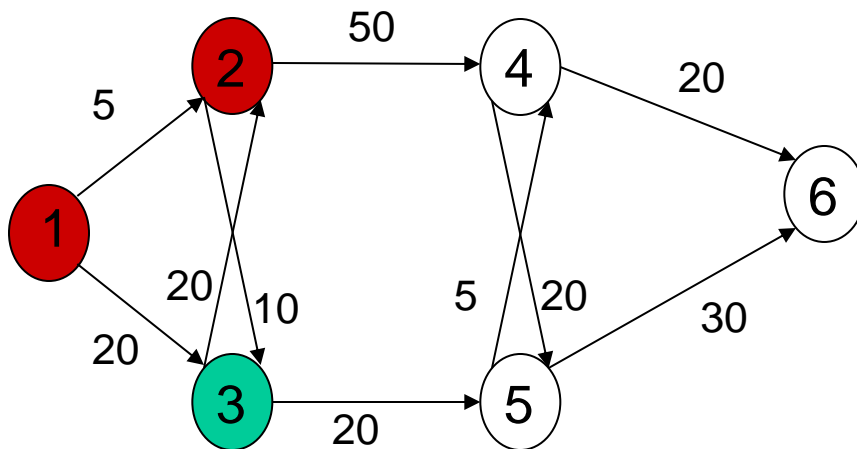
3. Jetzt: berechnen, ob die Knoten in  $V \setminus S$  günstiger über  $v$  als „Zwischenstation“ erreichbar ist, als ohne Umweg über  $v$ .



| Knoten | Abstand          |
|--------|------------------|
| 1      | 0                |
| 2      | 5                |
| 3      | <del>20</del> 15 |
| 4      | 55               |
| 5      |                  |
| 6      |                  |

## Beispiel (4)

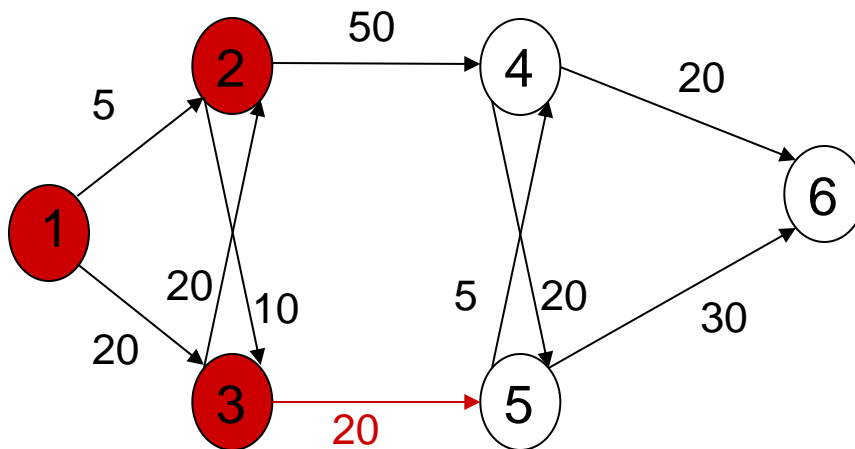
4. Danach: denjenigen Knoten  $v'$  zu  $S$  hinzunehmen, der nun am günstigsten zu erreichen ist.



| Knoten | Abstand |
|--------|---------|
| 1      | 0       |
| 2      | 5       |
| 3      | 15      |
| 4      | 55      |
| 5      |         |
| 6      |         |

## Beispiel (5)

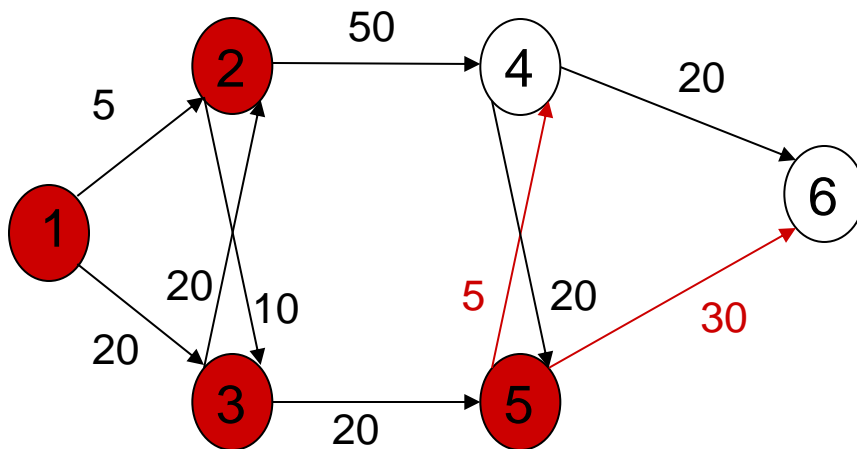
Neuberechnen der Wegekosten.



| Knoten | Abstand |
|--------|---------|
| 1      | 0       |
| 2      | 5       |
| 3      | 15      |
| 4      | 55      |
| 5      | 35      |
| 6      |         |

## Beispiel (6)

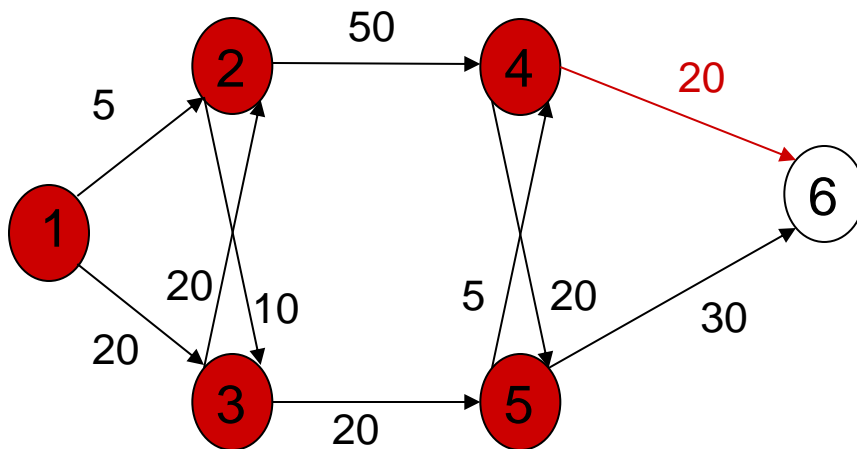
- wiederholen, bis alle Knoten in  $S$  sind.



| Knoten | Abstand          |
|--------|------------------|
| 1      | 0                |
| 2      | 5                |
| 3      | 15               |
| 4      | <del>55</del> 40 |
| 5      | 35               |
| 6      | 65               |

## Beispiel (7)

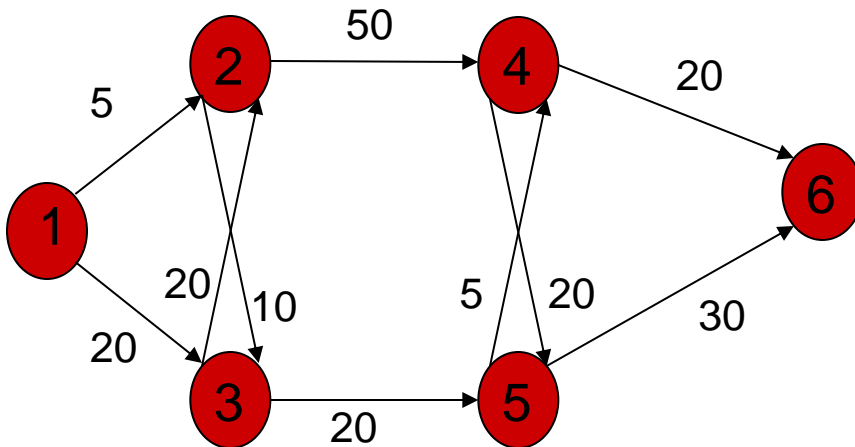
- wiederholen, bis alle Knoten in  $S$  sind.



| Knoten | Abstand          |
|--------|------------------|
| 1      | 0                |
| 2      | 5                |
| 3      | 15               |
| 4      | 40               |
| 5      | 35               |
| 6      | <del>65</del> 60 |

## Beispiel (8)

- wiederholen, bis alle Knoten in  $S$  sind.



| Knoten | Abstand |
|--------|---------|
| 1      | 0       |
| 2      | 5       |
| 3      | 15      |
| 4      | 40      |
| 5      | 35      |
| 6      | 60      |

Menge  $S$ : die Knoten, deren günstigste Wegekosten von der vorgegebenen Quelle bereits bekannt sind.

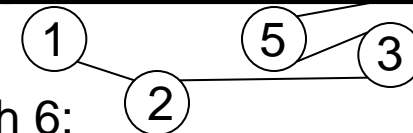
$d[i]$ : aktuell günstigste Kosten irgendeines Weges von Quelle zu  $v_i$

$p[i]$ : Wenn zusätzlich zu günstigsten Kosten auch nach günstigstem Weg zu jedem Knoten gefragt ist:

- Speichere direkten Vorgänger von  $v_i$  auf günstigstem Weg in  $p[i]$
- $p[i]$  wird mit Index der Quelle initialisiert.
- Nach Ablauf des Algorithmus kann günstigster Weg zu jedem Knoten mit Hilfe des Feldes  $p$  zurückverfolgt werden.

# Beispiel zum Ablauf bei Dijkstra

| $v_i$ | d[2] | d[3] | d[4] | d[5] | d[6] | p[2] | p[3] | p[4] | p[5] | p[6] |
|-------|------|------|------|------|------|------|------|------|------|------|
| 1     | 5    | 20   |      |      |      | 1    | 1    |      |      |      |
| 2     | 5    | 15   | 55   |      |      | 1    | 2    | 2    |      |      |
| 3     | 5    | 15   | 55   | 35   |      | 1    | 2    | 2    | 3    |      |
| 5     | 5    | 15   | 40   | 35   | 65   | 1    | 2    | 5    | 3    | 5    |
| 4     | 5    | 15   | 40   | 35   | 60   | 1    | 2    | 5    | 3    | 4    |
| 6     | 5    | 15   | 40   | 35   | 60   | 1    | 2    | 5    | 3    | 4    |



**Rückverfolgen** des günstigsten Weges z.B. von 1 nach 6:

- Vorgänger von 6 = 4
- Vorgänger von 4 = 5
- Vorgänger von 5 = 3, ....

d.h. günstigster Weg ist 1-2-3-5-4-6 (mit Kosten 60)



Quelle (Index '1') zur Menge 'S' hinzufügen

für alle anderen Knoten 'vi'

Kantenkosten zwischen Quelle und 'vi' in Feld 'd' speichern  
( $d[i]=c[1][i]$ )

Vorgänger von 'vi' ist Quelle '1':  $p[i]=1$

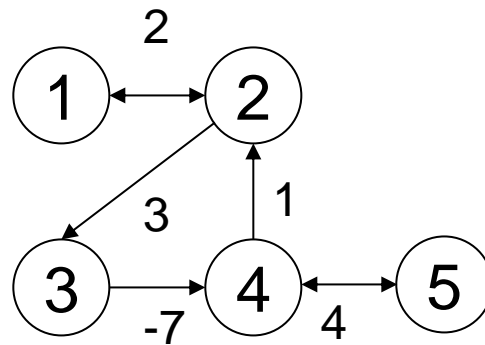
falls Kante zwischen vi und  
Quelle besteht

|                                                                          |        |
|--------------------------------------------------------------------------|--------|
| Solange nicht alle Knoten in S enthalten sind                            |        |
| wähle noch nicht in S enthaltenen Knoten 'vi' mit geringsten Kosten d[i] |        |
| füge vi zu S hinzu                                                       |        |
| für jeden noch nicht in S enthaltenen Knoten 'vj'                        |        |
| Kosten zu vj über vi ( $d[i] + c[i][j]$ )<br>geringer als d[j]?          |        |
| Wahr                                                                     | Falsch |
| neue Kosten speichern:<br>$d[j] = d[i] + c[i][j]$                        |        |
| Vorgänger von 'vj' ist 'vi': $p[j]=i$                                    |        |

- **Zeitkomplexität** des Dijkstra-Algorithmus:  
Annahme: Speicherung des Graphen mit **Adjazenzmatrix**

$$T_{\text{Dijkstra}}(|V|) \in O(|V|^2)$$

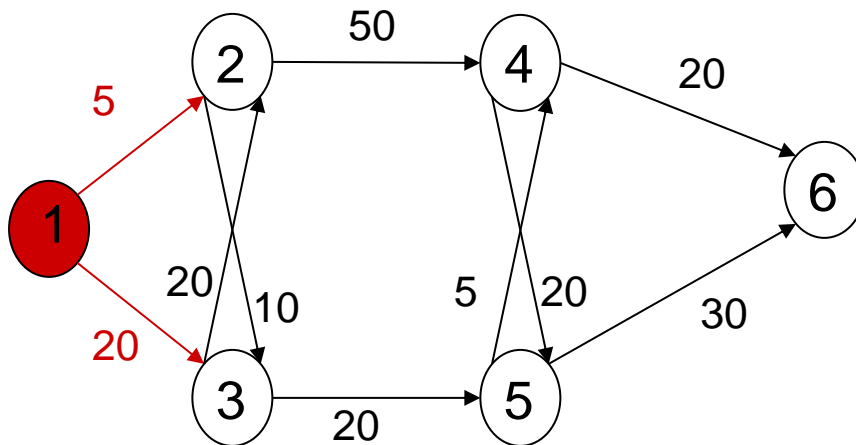
- **Nachteil: Keine negativen Gewichte möglich**
  - Ausweichen auf den Bellman-Ford-Algorithmus
  - Negative Zykel: Kein günstigster Weg möglich



Negativer Zykel 2-3-4-2  
Die Kosten 1→5 können beliebig klein werden.

- Dynamische Programmierung
- Nachteil  $O(v \cdot e)$  (statt  $O(v^2)$  bei Dijkstra)
- Vorteil: negative Kosten möglich (keine neg. Zyklen).

## Liste der Kanten aufstellen



|       |    |
|-------|----|
| 1 → 2 | 5  |
| 1 → 3 | 20 |
| 2 → 3 | 10 |
| 2 → 4 | 50 |
| 3 → 2 | 20 |
| 3 → 5 | 20 |
| 4 → 5 | 20 |
| 4 → 6 | 20 |
| 5 → 4 | 5  |
| 5 → 6 | 30 |

## Beispiel (1)

Berechnung des Abstands  $d_x$  zwischen Knoten 1 und Knoten x

- Schritt 1:  $d_1 = 0$
- Schritt 2: Für jede Kante  $a \rightarrow e$ :
  - Falls  $d_a$  bekannt
    - Falls  $d_a + g < d_e$  ( $g$ : Gewicht der Kante) oder  $d_e$  unbekannt:
      - Setze  $d_e$  auf  $d_a + g$ .

$1 \rightarrow 2$     5  
 $1 \rightarrow 3$     20  
 $2 \rightarrow 3$     10  
 $2 \rightarrow 4$     50  
 $3 \rightarrow 2$     20  
 $3 \rightarrow 5$     20  
 $4 \rightarrow 5$     20  
 $4 \rightarrow 6$     20  
 $5 \rightarrow 4$     5  
 $5 \rightarrow 6$     30

| Knoten | Abstand          |
|--------|------------------|
| 1      | 0                |
| 2      | 5                |
| 3      | <del>20</del> 15 |
| 4      | <del>55</del> 40 |
| 5      | 35               |
| 6      | <del>75</del> 65 |

- Optimaler Wert für alle Knoten, die optimal mit 1 Schritt erreicht werden können.

## Beispiel (2)

Berechnung des Abstands  $d_x$  zwischen Knoten 1 und Knoten x

- Schritt 3: Wiederhole Schritt 2

1 → 2    5  
1 → 3    20  
2 → 3    10  
2 → 4    50  
3 → 2    20  
3 → 5    20  
4 → 5    20  
4 → 6    20  
5 → 4    5  
5 → 6    30

| Knoten | Abstand          |
|--------|------------------|
| 1      | 0                |
| 2      | 5                |
| 3      | 15               |
| 4      | 55               |
| 5      | 35               |
| 6      | <del>65</del> 60 |

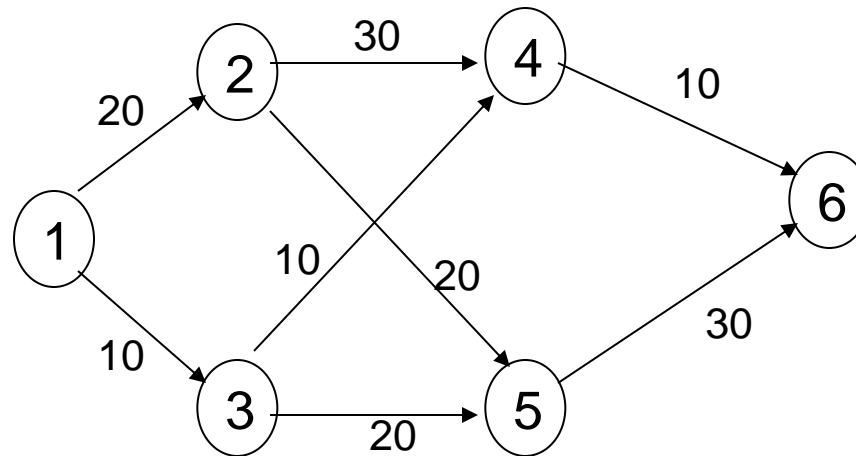
- Optimaler Wert für alle Knoten, die optimal mit 2 Schritten erreicht werden können.

## Beispiel (3)

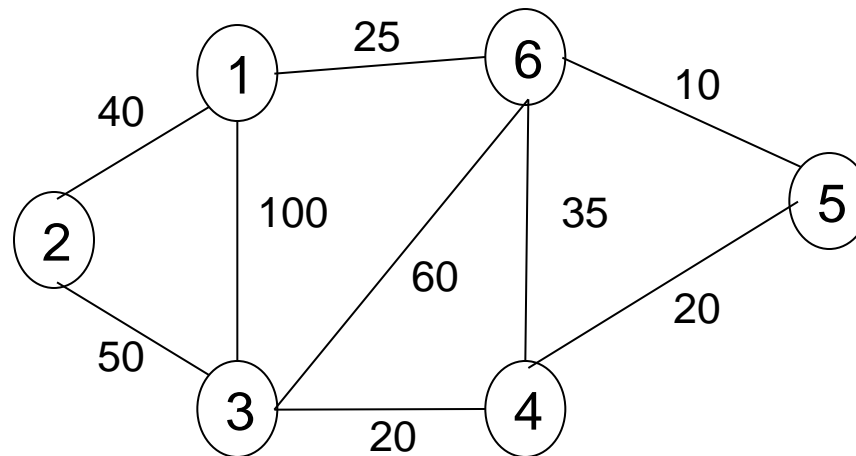
Berechnung des Abstands  $d_x$  zwischen Knoten 1 und Knoten  $x$

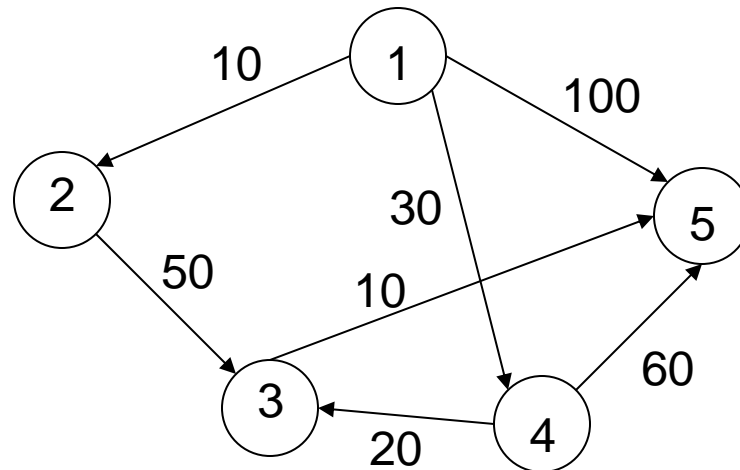
- Schritt 4: Wiederhole Schritt 2
  - Keine Änderung mehr → fertig
- Spätestens nach  $v-1$  Wiederholungen sind alle Abstände bestimmt (mehr als  $v-1$  Schritte benötigt keine optimale Wegstrecke).

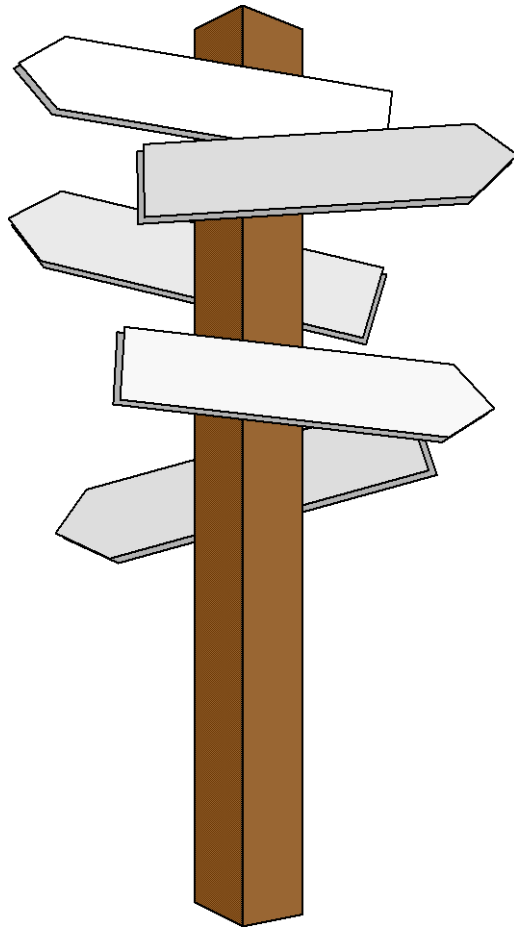
| Knoten | Abstand |
|--------|---------|
| 1      | 0       |
| 2      | 5       |
| 3      | 15      |
| 4      | 55      |
| 5      | 35      |
| 6      | 65      |











## Vorlesung 17

### Entwickeln von Lösungskonzepten

**Gegeben** ist ein algorithmisches Problem:

**Beispiel:**

Schreiben Sie eine Funktion

```
public static int[] getTwoSum(int[] arr, int sum)
```

die nach einem Zahlenpaar in `arr` sucht, das addiert `sum` ergibt und die Indizes der beiden Zahlen in einem `int[2]`-Feld zurückgibt.

Falls kein solches Zahlenpaar vorhanden ist, wird `null` zurückgegeben.

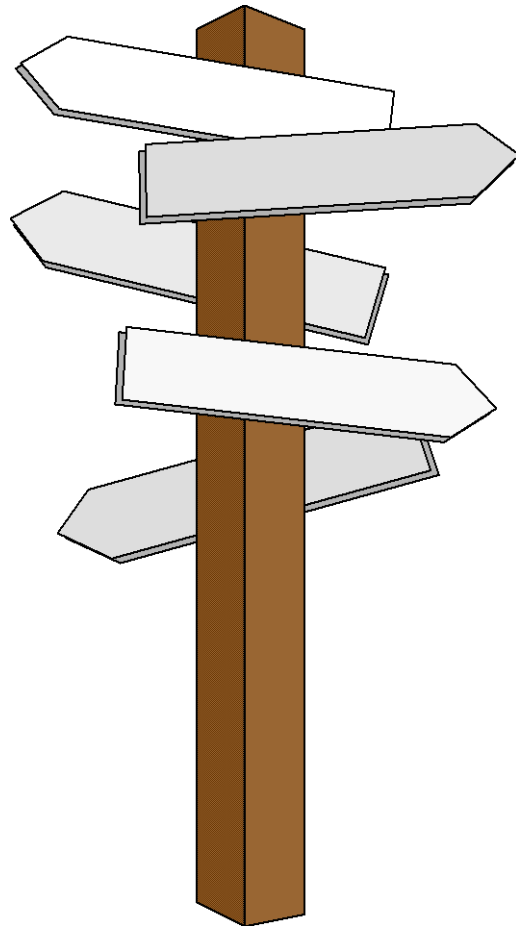
1. Schritt zur Problemlösung: Finden eines geeigneten Algorithmus.
2. Schritt zur Problemlösung: Codierung des Algorithmus.
  - Der erste Schritt ist oft ein großes Problem.
  - Leider gibt es hier kaum Patentrezepte.
  - Man braucht einfach viel Übung.

Folgende Punkte sind hilfreich:

- Es ist sinnvoll, den Lösungsansatz kurz zu notieren.
- Dazu benutzt man **Pseudocode**, eine Mischung aus Programmcode und Text.
- Für Pseudocode gibt es wenig Regeln.
- Wichtig ist, dass man hinterher seinen eigenen Pseudocode noch versteht (es schadet natürlich nicht, wenn auch andere ihn verstehen können).

Folgende Punkte sind hilfreich:

- Es ist wichtig, sehr viele Übungsaufgaben zu lösen.
- Den größten Teil dieser Vorlesung nehmen darum auch Übungsaufgaben ein.
- Für mehr Übungsaufgaben kann man das Internet nach dem Stichwort „Coding Interview“ durchsuchen.
  - **Coding Interviews verwenden große Softwareunternehmen (Microsoft, Google, Amazon, ...) in Bewerbungsgesprächen, um Programmierfähigkeiten zu testen.**
- Ein anderes Stichwort ist „code kata“.
  - **Katas heißen in verschiedenen japanischen Künsten die Grundtechniken.**



## Vorlesung 18

### Der Floyd-Warshall-Algorithmus



Gegeben (ähnlich wie für Dijkstra-Algorithmus):

Graph  $G = (V, E)$  mit folgender Bewertungsfunktion:

- Jede Kante von  $v_i$  nach  $v_j$  hat nicht negative Kosten:  $C(i, j) \geq 0$
- Falls keine Kante zwischen  $v_i$  und  $v_j$  :  $C(i, j) = \infty$

**Aufgabenstellung:** Bestimme für alle geordneten Paare  $(v, w)$  den kürzesten Weg von  $v$  nach  $w$ .

**Lösungsmöglichkeiten:**

1. Wende *Dijkstra* -Algorithmus für alle Knoten  $v$  als Quelle an  
⇒ Zeitkomplexität  $O(|V|^3)$ .
2. Verwende den *Floyd* -Algorithmus.

$|V|$  Iterationen:

- 1. Schritt: Vergleiche Kosten von
  - direkter Verbindung von Knoten  $i$  zu Knoten  $j$
  - Umweg über Knoten 1 (also: von  $i$  nach 1; von 1 nach  $j$ ).
  - Falls Umweg günstiger: alten Weg durch Umweg ersetzen.
- 2. Schritt: zusätzlich Umwege über Knoten 2 betrachten.
- $k$ -ter Schritt: Umwege über Knoten  $k$  betrachten, usw.

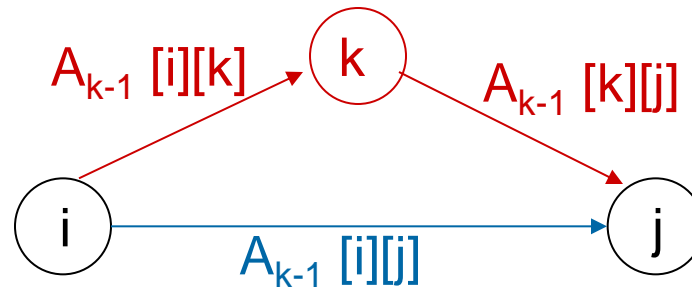
Floyd-Algorithmus nutzt eine  $|V| \times |V|$  Matrix  $A_k[i][j]$ , um Kosten der günstigsten Wege zu speichern (Subscript nur zum Bezeichnen der Iteration):

$A_k[i][j]$  := minimale Kosten, um über irgendwelche der Knoten in  $\{1, \dots, k\}$  vom Knoten  $i$  zum Knoten  $j$  zu gelangen

$A[i][j]$  = Kosten des aktuell günstigsten Wegs :

- Initialisierung:  $A_0[i][j] = C(i,j) \quad \forall i \neq j$ ; Diagonalelemente  $A_0[i][i] = 0$
- $|V|$  Iterationen mit „dynamischer Programmierung“:  
Iterationsformel zur Aktualisierung von  $A[i][j]$  :

$$A_k[i][j] = \min \{ A_{k-1}[i][j], A_{k-1}[i][k] + A_{k-1}[k][j] \}$$

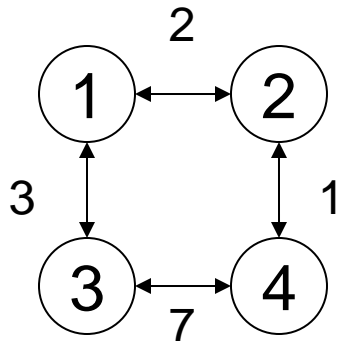


- Zum Schluss:  $A_{|V|}[i][j]$  = Kosten des günstigsten Wegs von i nach j.

Bemerkung:  $A_k[i][k] = A_{k-1}[i][k]$  und  $A_k[k][j] = A_{k-1}[k][j]$

```
for (int i=0; i<a.length; i++) {
 for (int j=0; j<a.length; j++) {
 for (int k=0; k<a.length; k++) {
 if (a[j][k]>a[j][i]+a[i][k]) {
 a[j][k]=a[j][i]+a[i][k];
 }
 }
 }
}
```

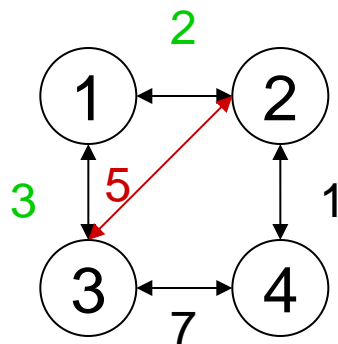
# Beispiel zum Floyd-Algorithmus (1)



$A_0$  (direkter Weg)

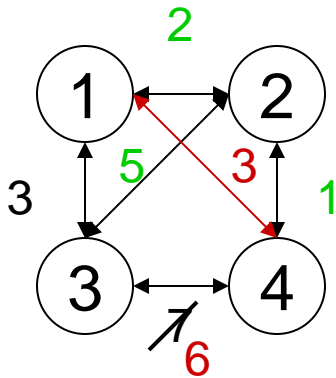
|     |   | nach     |          |          |          |   |   |   |   |
|-----|---|----------|----------|----------|----------|---|---|---|---|
|     |   | 1        | 2        | 3        | 4        |   |   |   |   |
| von | 1 | 0        | 2        | 3        | $\infty$ | - | - | - | - |
|     | 2 | 2        | 0        | $\infty$ | 1        | - | - | - | - |
|     | 3 | 3        | $\infty$ | 0        | 7        | - | - | - | - |
|     | 4 | $\infty$ | 1        | 7        | 0        | - | - | - | - |

Umweg über 1



|     |   | nach     |   |   |          |   |   |   |   |
|-----|---|----------|---|---|----------|---|---|---|---|
|     |   | 1        | 2 | 3 | 4        |   |   |   |   |
| von | 1 | 0        | 2 | 3 | $\infty$ | - | - | - | - |
|     | 2 | 2        | 0 | 5 | 1        | - | - | 1 | - |
|     | 3 | 3        | 5 | 0 | 7        | - | 1 | - | - |
|     | 4 | $\infty$ | 1 | 7 | 0        | - | - | - | - |

# Beispiel zum Floyd-Algorithmus (2)

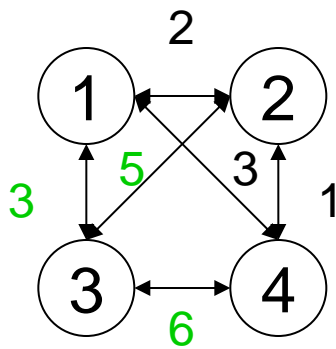


Umweg über 2

|     |   | nach |   |   |   |
|-----|---|------|---|---|---|
|     |   | 1    | 2 | 3 | 4 |
| von | 1 | 0    | 2 | 3 | 3 |
|     | 2 | 2    | 0 | 5 | 1 |
|     | 3 | 3    | 5 | 0 | 6 |
|     | 4 | 3    | 1 | 6 | 0 |

|   |   |   |   |
|---|---|---|---|
| - | - | - | 2 |
| - | - | 1 | - |
| - | 1 | - | 2 |
| 2 | - | 2 | - |



Umweg über 3

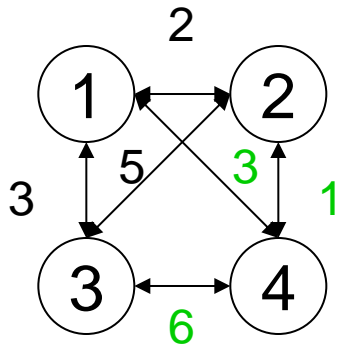
|     |   | nach |   |   |   |
|-----|---|------|---|---|---|
|     |   | 1    | 2 | 3 | 4 |
| von | 1 | 0    | 2 | 3 | 3 |
|     | 2 | 2    | 0 | 5 | 1 |
|     | 3 | 3    | 5 | 0 | 6 |
|     | 4 | 3    | 1 | 6 | 0 |

|   |   |   |   |
|---|---|---|---|
| - | - | - | 2 |
| - | - | 1 | - |
| - | 1 | - | 2 |
| 2 | - | 2 | - |

# Beispiel zum Floyd-Algorithmus (3)

Umweg über 4



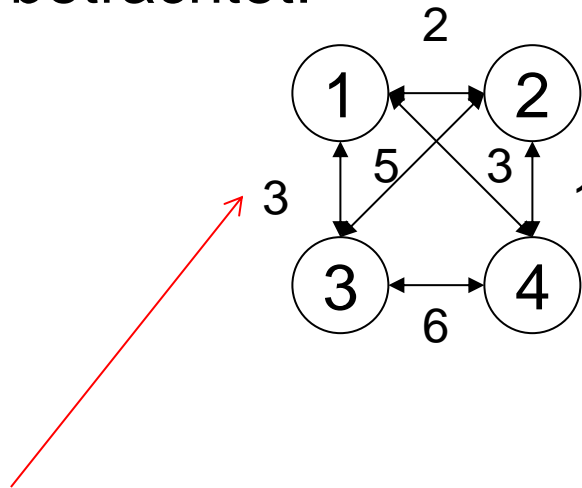
|       | nach |   |   |   |
|-------|------|---|---|---|
|       | 1    | 2 | 3 | 4 |
| von 1 | 0    | 2 | 3 | 3 |
| 2     | 2    | 0 | 5 | 1 |
| 3     | 3    | 5 | 0 | 6 |
| 4     | 3    | 1 | 6 | 0 |

|   |   |   |   |
|---|---|---|---|
| - | - | - | 2 |
| - | - | 1 | - |
| - | 1 | - | 2 |
| 2 | - | 2 | - |

- Der Floyd-Algorithmus verwendet das Entwurfsprinzip „dynamische Programmierung“.
- Warum funktioniert der Floyd-Algorithmus?
- Woran erkennt man das Entwurfsprinzip?

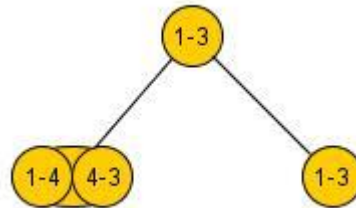


- Betrachten wir noch einmal das Ergebnis des letzten Beispiels. Im letzten Schritt hatten wir die Umwege über Knoten 4 betrachtet:



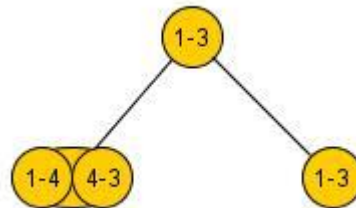
- Die Kosten 3 der Kante 1-3 könnten im letzten Schritt auf zwei Arten entstanden sein.

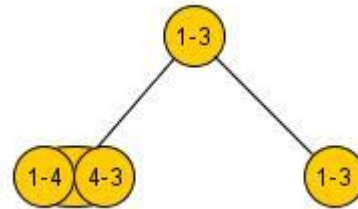
- Entweder die Kosten wurden im letzten Schritt nicht aktualisiert (rechts)
- oder sie bildeten sich aus der Summe der Kosten 1-4 und 4-3 (links)



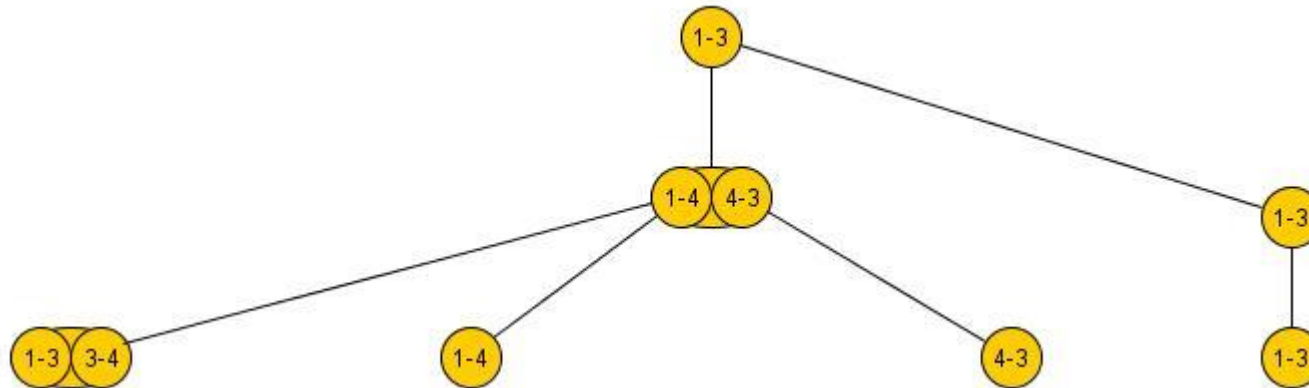
Anders ausgedrückt:

- Entweder der kürzeste Weg verläuft nicht über die Zwischenstation 4 (rechts)
- oder der kürzeste Weg verläuft über die Zwischenstation 4 und setzt sich zusammen aus den kürzesten Wegen 1-4 und 4-3 (links).

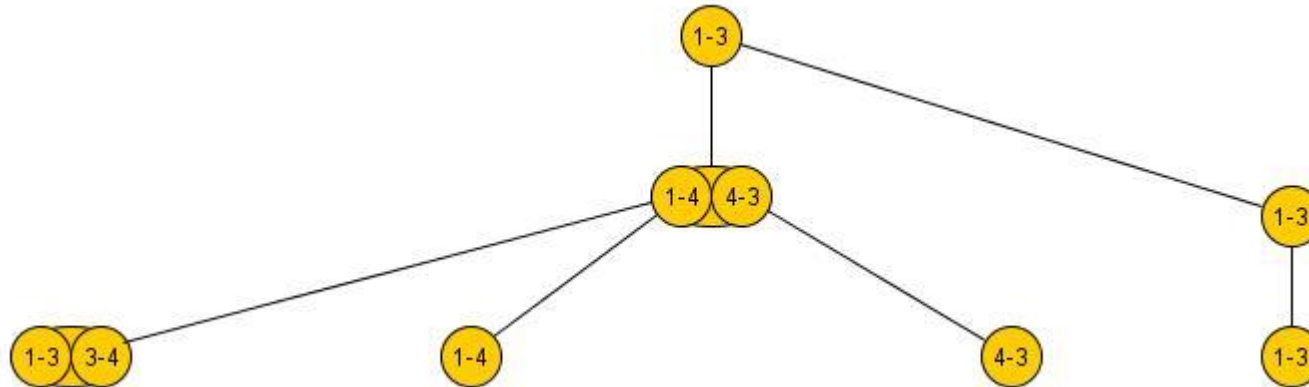




- Wichtig ist, dass in der unteren Ebene die Einzelwege nicht mehr über die Zwischenstation 4 verlaufen.
  - Links wäre das unsinnig, da 4 ja die Anfangs- oder Endstation ist.
  - Rechts ist 4 nach Definition keine Zwischenstation.

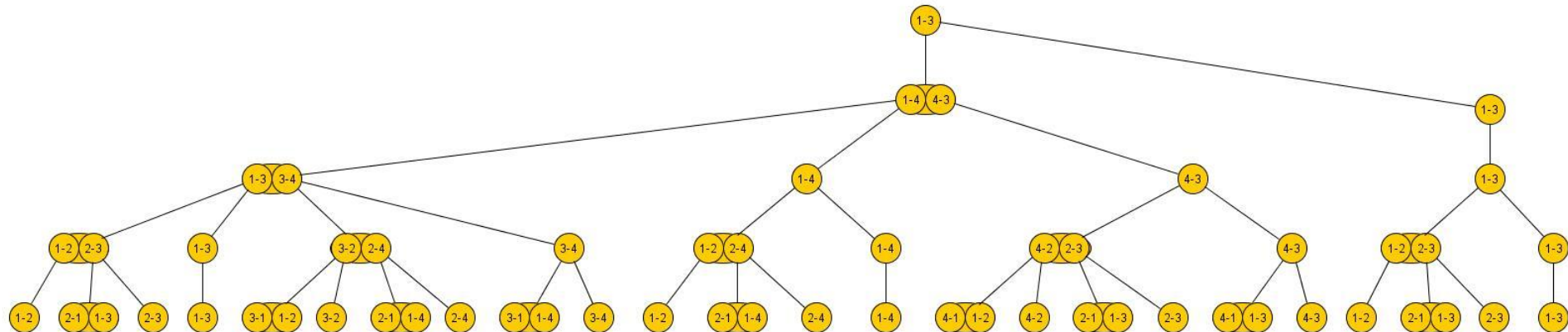


- In einer weiteren Ebene verzweigen sich die Möglichkeiten weiter:
  - **Entweder, die kürzesten Wege führen über die Zwischenstation 3 (1-3 und 3-4)**
  - **oder sie führen nicht über die Zwischenstation 3 (1-4).**
- Wenn 3 Anfangs- oder Endstation ist, gibt es nur eine Verzweigung (rechts bei 4-3 und bei 1-3).



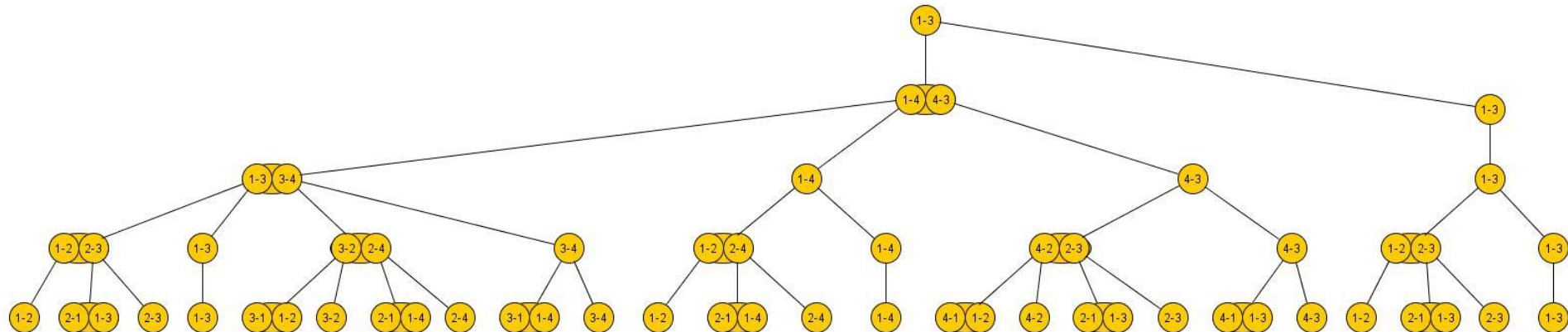
- Die Wege in der untersten Ebene führen jetzt weder über die Zwischenstation 4 noch über die Zwischenstation 3.

## Funktionsweise (8)



- Dies setzt man jetzt mit den Ebenen 2 und 1 fort.
- Die Wege in der untersten Ebene dürfen jetzt überhaupt keine Zwischenstationen mehr haben.
- Sie sind also die direkten Verbindungen und die Kosten sind zu Beginn bekannt.

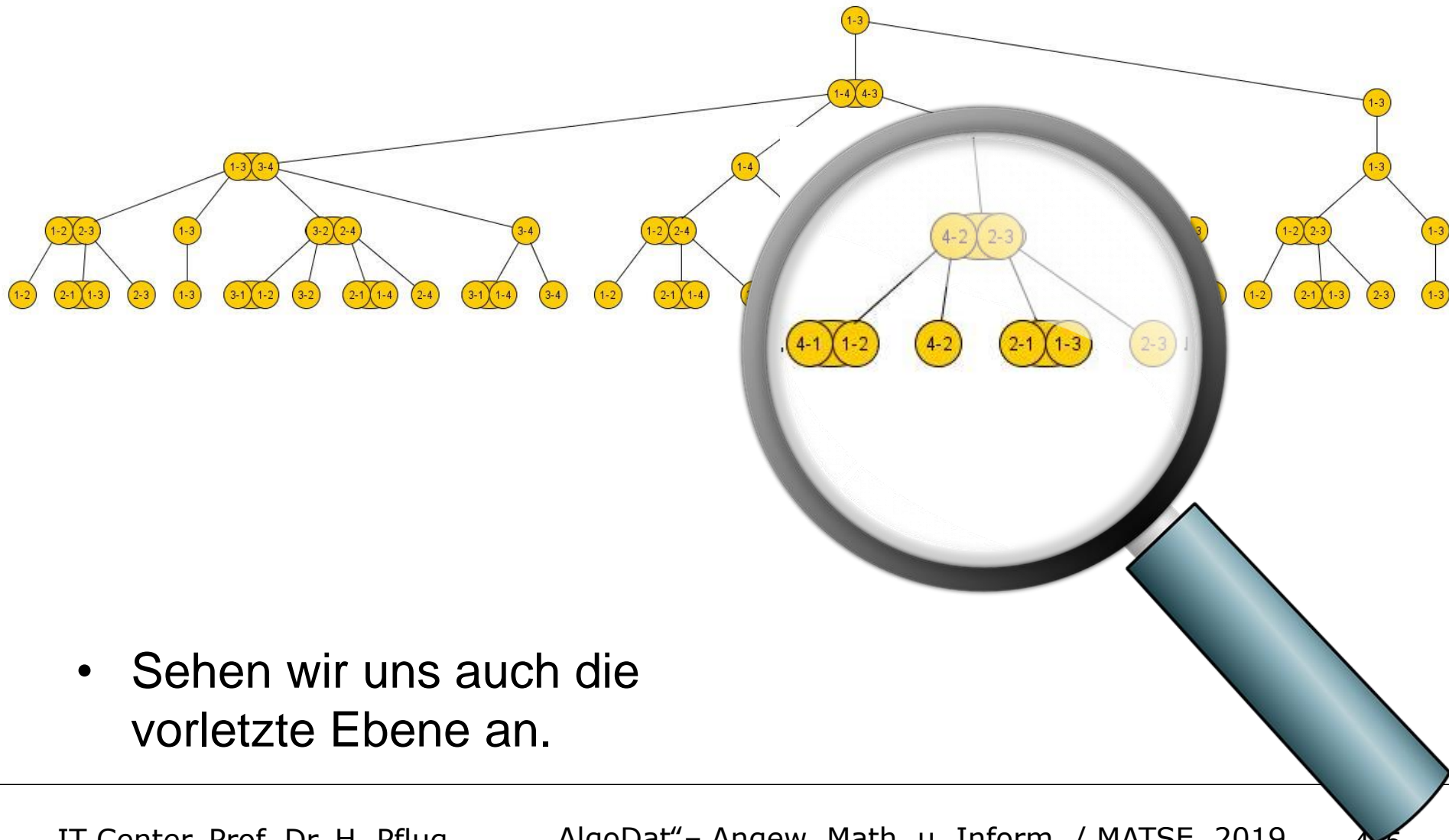
# Funktionsweise (9)



- Der Baum kann zwar sehr groß werden, aber alle Einträge in der untersten Zeile kann man in der Startmatrix ablesen.
- Es sind viele Einträge mehrfach vorhanden.
- -> Ideale Voraussetzungen für dynamische Programmierung.

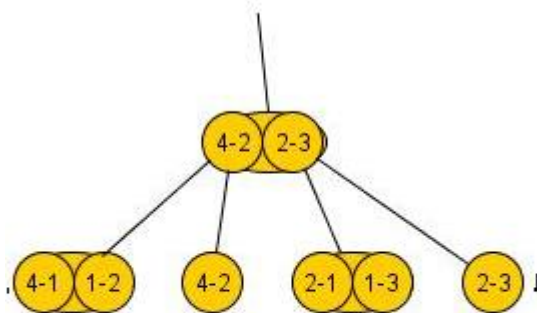
|     |   | nach     |          |          |          |
|-----|---|----------|----------|----------|----------|
|     |   | 1        | 2        | 3        | 4        |
| von | 1 | 0        | 2        | 3        | $\infty$ |
|     | 2 | 2        | 0        | $\infty$ | 1        |
|     | 3 | 3        | $\infty$ | 0        | 7        |
|     | 4 | $\infty$ | 1        | 7        | 0        |



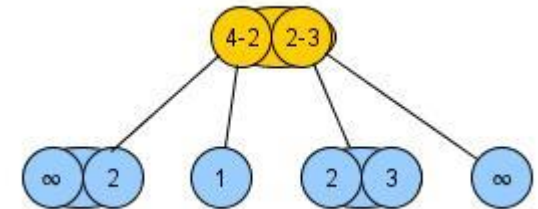


- Sehen wir uns auch die vorletzte Ebene an.

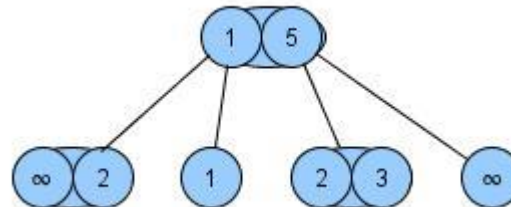
- Wenn wir die letzte Ebene eingetragen haben....



|     | nach     |          |          |          |
|-----|----------|----------|----------|----------|
|     | 1        | 2        | 3        | 4        |
| von |          |          |          |          |
| 1   | 0        | 2        | 3        | $\infty$ |
| 2   | 2        | 0        | $\infty$ | 1        |
| 3   | 3        | $\infty$ | 0        | 7        |
| 4   | $\infty$ | 1        | 7        | 0        |



- ...dann können wir einfach die Elemente der vorletzten Ebene bestimmen.



- Praktischerweise fertigen wir uns dazu eine Tabelle der vorletzten Ebene an.

|     |  | nach     |          |          |          |   |     |          | nach |   |          |  |
|-----|--|----------|----------|----------|----------|---|-----|----------|------|---|----------|--|
|     |  | 1        | 2        | 3        | 4        |   | 1   | 2        | 3    | 4 |          |  |
| von |  |          |          |          |          |   | von |          |      |   |          |  |
| 1   |  | 0        | 2        | 3        | $\infty$ | → | 1   | 0        | 2    | 3 | $\infty$ |  |
| 2   |  | 2        | 0        | $\infty$ | 1        |   | 2   | 2        | 0    | 5 | 1        |  |
| 3   |  | 3        | $\infty$ | 0        | 7        |   | 3   | 3        | 5    | 0 | 7        |  |
| 4   |  | $\infty$ | 1        | 7        | 0        |   | 4   | $\infty$ | 1    | 7 | 0        |  |

- Das ist genau der erste Schritt im Floyd-Warshall-Verfahren.

- Mit jedem Schritt des Floyd-Warshall-Verfahrens geht man eine weitere Ebene nach oben.
- In der letzten Tabelle kann man die Kosten des Wurzelknotens ablesen. Dies ist der gesuchte kürzeste Weg.
- Das entspricht genau dem Entwurfsprinzip der dynamischen Programmierung.

$T_{\text{Floyd}}(|V|) \in O(|V|^3)$  (asymptotisch nicht schneller als  $|V|$  mal Dijkstra)

## Bemerkung:

Algorithmus hat sehr einfache Struktur.

⇒ Compiler erzeugt effizienten Code.

Dadurch wird Floyd doch schneller als  $|V|$  mal Dijkstra sein.

**Gegeben:** Graph  $G = (V, E)$  (egal ob gerichtet oder nicht; gewichtet oder nicht)

**Aufgabenstellung:** Prüfe für alle geordneten Paare  $(v_i, v_j)$ , ob ein Weg (beliebiger Länge) von  $v_i$  nach  $v_j$  existiert.

Ziel: Berechne „Adjazenzmatrix  $A$  zu **transitivem Abschluss**“ von  $G$ :

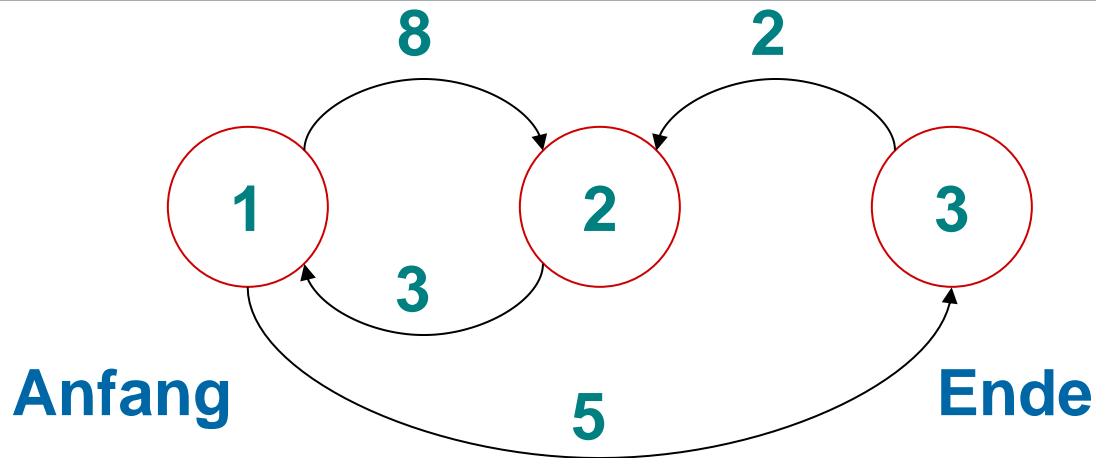
$A[i][j] = \text{true} \Leftrightarrow$  es existiert ein nicht-trivialer Weg (Länge  $> 0$ )  
von  $v_i$  nach  $v_j$

**Lösung:** Modifikation von Floyd–Algorithmus

$\Rightarrow$  **Warshall-Algorithmus**

**Iterationsformel:**  $A_k[i][j] = A_{k-1}[i][j] \vee (A_{k-1}[i][k] \wedge A_{k-1}[k][j])$

# Beispiel zum Warshall-Algorithmus



Floyd:

|          |   |          |
|----------|---|----------|
| 0        | 8 | 5        |
| 3        | 0 | $\infty$ |
| $\infty$ | 2 | 0        |

Floyd:

|   |   |   |
|---|---|---|
| 0 | 7 | 5 |
| 3 | 0 | 8 |
| 5 | 2 | 0 |

Warshall:

$A_{\text{Original}} =$

|   |   |   |
|---|---|---|
| T | T | T |
| T | T | F |
| F | T | T |

Warshall:

$A_{\text{Abschluss}} =$

|   |   |   |
|---|---|---|
| T | T | T |
| T | T | T |
| T | T | T |

```
for (int i=0; i<a.length; i++) {
 for (int j=0; j<a.length; j++) {
 for (int k=0; k<a.length; k++) {
 a[j][k]=a[j][k]|| (a[j][i]&&a[i][k]);
 }
 }
}
```



$$T_{\text{warshall}}(|V|) \in O(|V|^3)$$

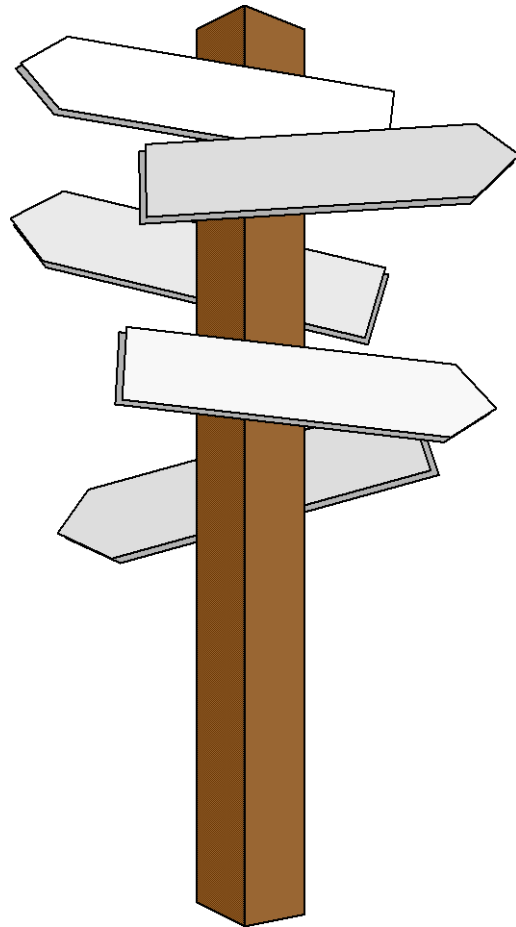
- Aufgabe 1. Finden Sie die kürzesten Wege mit dem Floyd-Algorithmus.

|     | nach     |          |          |          |
|-----|----------|----------|----------|----------|
|     | 1        | 2        | 3        | 4        |
| von |          |          |          |          |
| 1   | 0        | 5        | 1        | $\infty$ |
| 2   | $\infty$ | 0        | 8        | $\infty$ |
| 3   | $\infty$ | 3        | 0        | 4        |
| 4   | 2        | $\infty$ | $\infty$ | 0        |

- Aufgabe 2. Finden Sie die kürzesten Wege mit dem Floyd-Algorithmus.

|     | nach     |          |          |          |
|-----|----------|----------|----------|----------|
|     | 1        | 2        | 3        | 4        |
| von |          |          |          |          |
| 1   | 0        | 4        | $\infty$ | $\infty$ |
| 2   | $\infty$ | 0        | 1        | 4        |
| 3   | 5        | $\infty$ | 0        | 1        |
| 4   | $\infty$ | 2        | 1        | 0        |

- Teile dieser Algorithmen finden sich als 'Skelett' in vielen interessanten Lösungen zu Problemen aus dem wirklichen Leben:
  - Netzplantechnik
  - Compiler (Vektorisierung)
  - Problem des Handlungsreisenden
  - Produktionssteuerung, etc.
- Es gibt eine große Anzahl weiterer Probleme und Algorithmen
  - Zusammenhangskomponenten (Erreichbarkeit)
  - Flüsse auf Netzwerken (maximale, minimale)
  - Transportprobleme
    - Euler-Kreis: Kreis, in dem jede Kante einmal durchlaufen wird (einfaches Problem)
    - Hamilton-Kreis: Kreis, in dem jeder Knoten einmal durchlaufen wird (sehr schwieriges Problem)
  - Färbungsprobleme



## Vorlesung 19

### Textsuche

- Einführung
- Naives Verfahren
- Boyer/Moore
- Rabin/Karp
- Mustererkennung mit endlichen Automaten

**Problem:** Suche erstes Vorkommen von

**Muster-Zeichenfolge** `pattern[0...m-1]` in

**Text-Zeichenfolge** `text[0...n-1]`

**Anwendungen:**

- Finden von Mustern in Text-Dateien, z.B. Unix-Kommando `grep`; Suchfunktion im Editor; Web-Suchmaschine
- Algorithmen funktionieren auch bei anderen „Alphabeten“, z.B. DNS-Sequenzen oder Bitfolgen

**Verwandtes Problem:** Muster ist regulärer Ausdruck

⇒ „Pattern Matching“ („Mustererkennung“) mit **endlichen Automaten**

# Übersicht über die Textsuchverfahren

|                                                 |                                             |                                             |
|-------------------------------------------------|---------------------------------------------|---------------------------------------------|
| Naiver / grober /<br>brute force-Algorithmus    |                                             | Naives<br>Suchverfah-<br>ren $O(n \cdot m)$ |
| Knuth-Morris-Pratt                              | Rabin-Karp                                  | $\sim O(n+m)$                               |
| (Boyer-Moore)-Sunday<br>(Boyer-Moore)-Horsepool | vereinfachter<br>Boyer-Moore<br>Boyer-Moore | $> O(n + m)$                                |



- Die Textsuchverfahren „**Naiv**“, „**Knuth-Morris-Pratt**“, „**Boyer-Moore-Sunday**“ sind verwandt.
  - Für kleine Texte ist „**Naiv**“ am schnellsten.
  - Für große Texte ist „**Boyer-Moore-Sunday**“ am schnellsten.
  - **Knuth-Morris-Pratt** hat vor allem historische Bedeutung.
- „**Rabin-Karp**“ (hier nicht erklärt) benutzt ein anderes Prinzip.
- Zunächst wird die **grundsätzliche Problemstellung** bei „**Naiv**“, „**Knuth-Morris-Pratt**“ und „**Boyer-Moore-Sunday**“ erläutert.

Zu Beginn setzt man das Muster an den Beginn des Textes

text      U N G L E I C H U N G S T E I L . . .

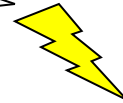
pattern   U N G L E I C H U N G E N

Dann wird der Text Zeichen für Zeichen durchgegangen, bis entweder

- eine komplette Übereinstimmung festgestellt wird
- oder ein Zeichen nicht übereinstimmt.

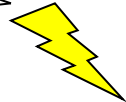
text      U N G L E I C H U N G **S** T E I L . . .

pattern   U N G L E I C H U N G **E** N



Wenn ein Zeichen nicht übereinstimmt, muss das Muster neu angesetzt werden. [Aber wohin?](#)

Das kommt auf das Verfahren an.

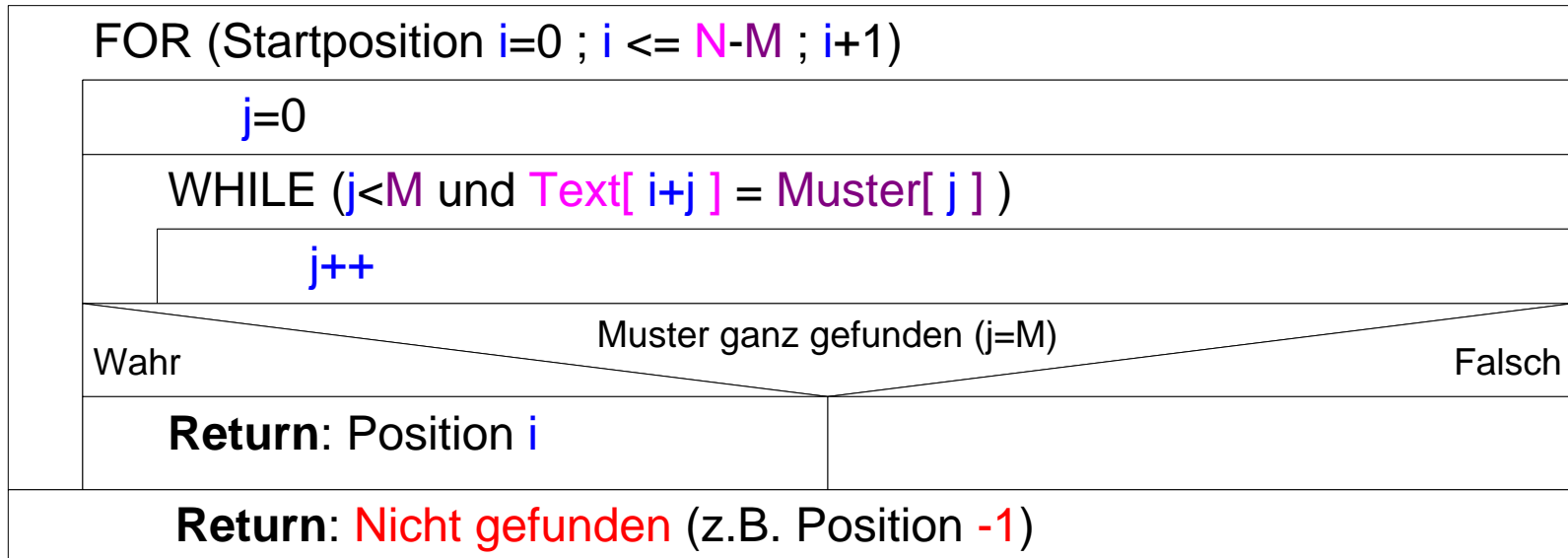
text      U N G L E I C H U N G **S** T E I L . . .  
pattern U N G L E I C H U N G **E** N 

## Naives Verfahren:

Das Muster wird einen Buchstaben weiter gesetzt.

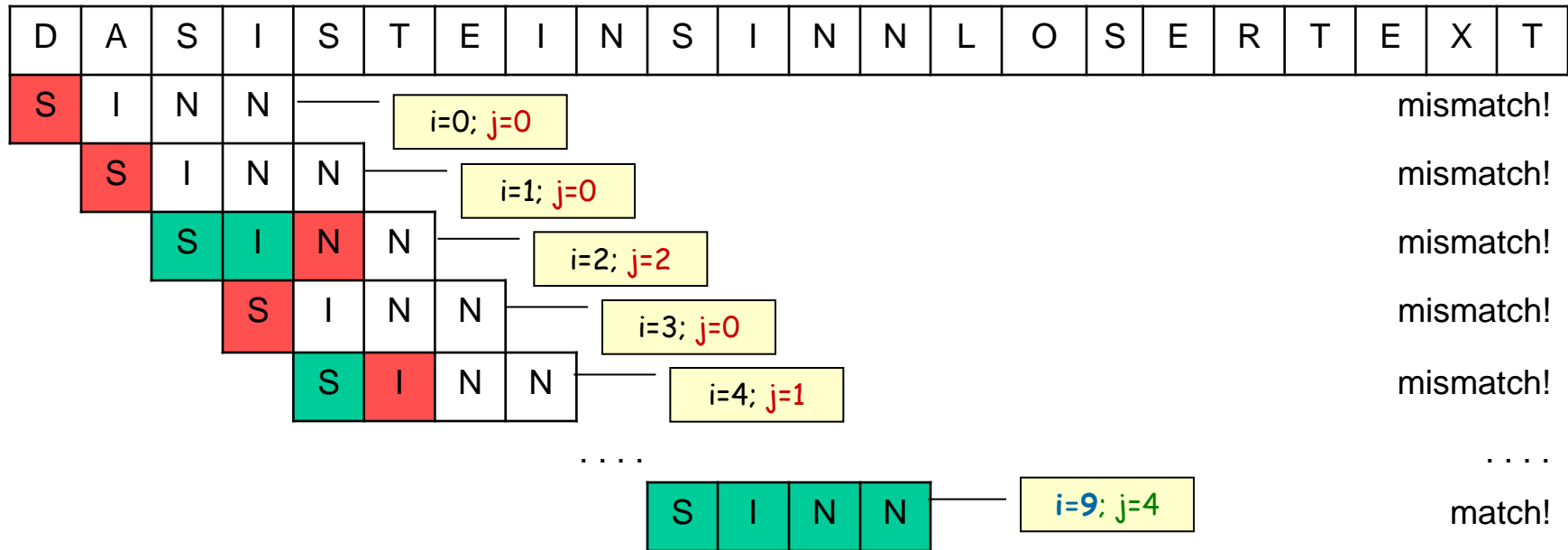
text      U N G L E I C H U N G S T E I L . . .  
pattern      U N G L E I C H U N G E N

Ab jeder Position  $i \in [0;n-m]$  prüfen, ob  $\text{text}[i \dots i + (m-1)] = \text{pattern}$



- Worst Case:  $m(n-m)$  Vergleiche; für  $n \gg m$  (Normalfall):  $O(n \cdot m)$
- Best Case (realistisch):  $O(n)$  bei Nicht-Finden;  $O(m)$  bei Finden;
- Zusätzlicher Platzbedarf:  $O(1)$

# Mustersuche mit naive Algorithmus (Bsp.)



Ein „mismatch“ liegt bei Nichtübereinstimmung vor.  
 „match“ bedeutet „komplette Übereinstimmung“.

Jetzt muss das Muster neu angesetzt werden. Aber wo?

**Regel:** Wenn die letzten überprüften Buchstaben gleich dem Anfang des Patterns sind, verschiebt man das Muster entsprechend.

|         |                                    |
|---------|------------------------------------|
| text    | U N G L E I C H U N G S T E I L... |
| pattern | U N G L E I C H U N G E N          |

und macht beim anschließenden Zeichen weiter.

|         |                                    |
|---------|------------------------------------|
| text    | U N G L E I C H U N G S T E I L... |
| pattern | U N G L E I C H U N G E N          |

Wenn die letzten überprüften Buchstaben nicht gleich dem Anfang des Musters sind, verschiebt man das Muster so, dass das erste Zeichen auf dem Mismatch zu liegen kommt.

text      U N G L E I C H **U** N G S T E I L...

pattern   U N G L E I C H **E** R

ergibt

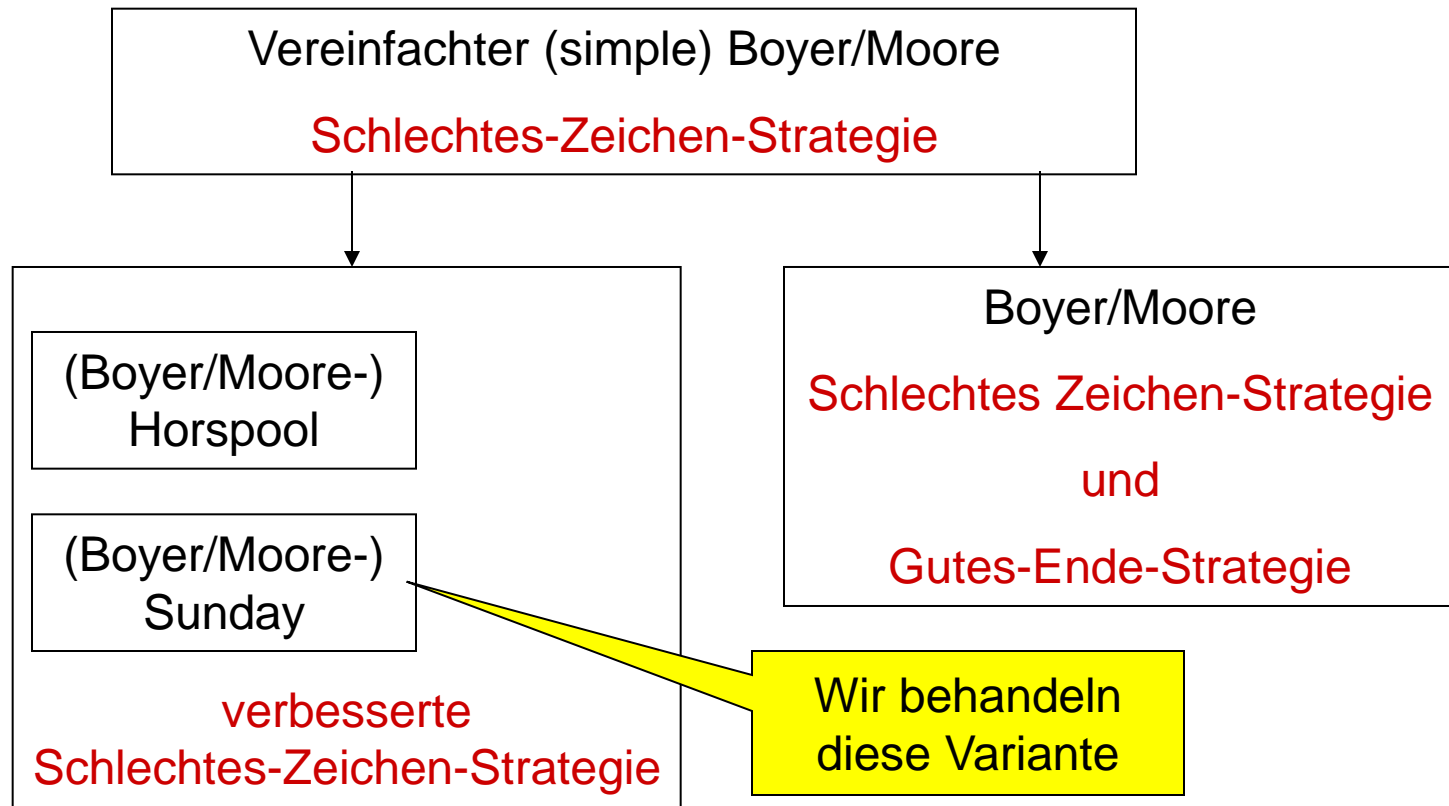
text      U N G L E I C H **U** N G S T E I L...

pattern                      U N G L E I C H E R

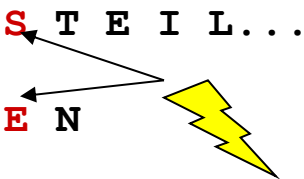
Wichtig ist zu prüfen, welcher von beiden Fällen eintritt.

- Der Algorithmus wird in der Vorlesung nicht genauer erklärt.
- Es muss eine **next-Tabelle** initialisiert werden: Aufwand  $O(m)$
- Dann beginnt die **Such-Phase**: in jedem Schritt
  - entweder im Text um 1 Zeichen weitergehen ( $++i$ ),
  - oder Muster um mindestens 1 Zeichen weiter rechts „anlegen“ ( $j = \text{next}[j]$ ). $\Rightarrow$  Aufwand  $O(n)$
- KMP-Algorithmus **insgesamt**:  $O(n + m)$
- Zusätzlicher **Platzbedarf**:  $O(m)$





text      U N G L E I C H U N G S T E I L . . .  
pattern   U N G L E I C H U N G E N

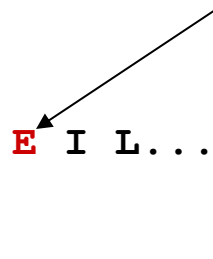


Wenn ein Zeichen nicht übereinstimmt, muss das Pattern neu angesetzt werden. [Aber wohin?](#)

## Boyer-Moore-Sunday:

Man betrachtet den Buchstaben, der hinter dem Muster liegt:

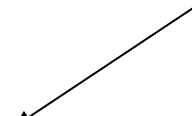
text      U N G L E I C H U N G S T E I L . . .  
pattern   U N G L E I C H U N G E N



## Boyer-Moore-Sunday:

Man betrachtet den Buchstaben, der hinter dem Muster liegt:

text      U N G L E I C H U N G S T **E** I L . . .  
pattern  U N G L E I C H U N G E N



Das Muster kann so weit nach vorne geschoben werden, bis ein Buchstabe des Musters mit diesem Buchstaben übereinstimmt.

text      U N G L E I C H U N G S T **E** I L . . .  
pattern        U N G L E I C H U N G **E** N

Anschließend wird Text und Muster wieder verglichen.

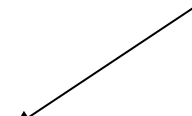
text      U N G L E I C H U N G S T E I L . . .  
pattern      U N G L I C H U N G E N

Bei Ungleichheit wird wieder der erste Buchstabe hinter dem Muster betrachtet und das Muster entsprechend vorgeschoben.

text      U N G L E I C H U N G S T E I L . . .  
pattern                      U N G L E I C H U N G E N

Der auf das Muster folgende Buchstabe kommt im Muster mehrmals vor. Wo wird das Muster angelegt?

text      U N G L E I C H U N G S T **E** I L . . .  
pattern  U N G L **E** I C H U N G **E** N



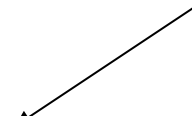
Antwort: An den Buchstaben, der im Muster am weitesten hinten liegt.

Das entspricht dem kleinsten der möglichen Sprünge.

text      U N G L E I C H U N G S T **E** I L . . .  
pattern        U N G L E I C H U N G **E** N

Der auf das Muster folgende Buchstabe kommt im Muster **nicht** vor.  
Wo wird das Muster angelegt?

```
text U N G L E I C H U N G S - T E I L . . .
pattern U N G L E I C H U N G E N
```

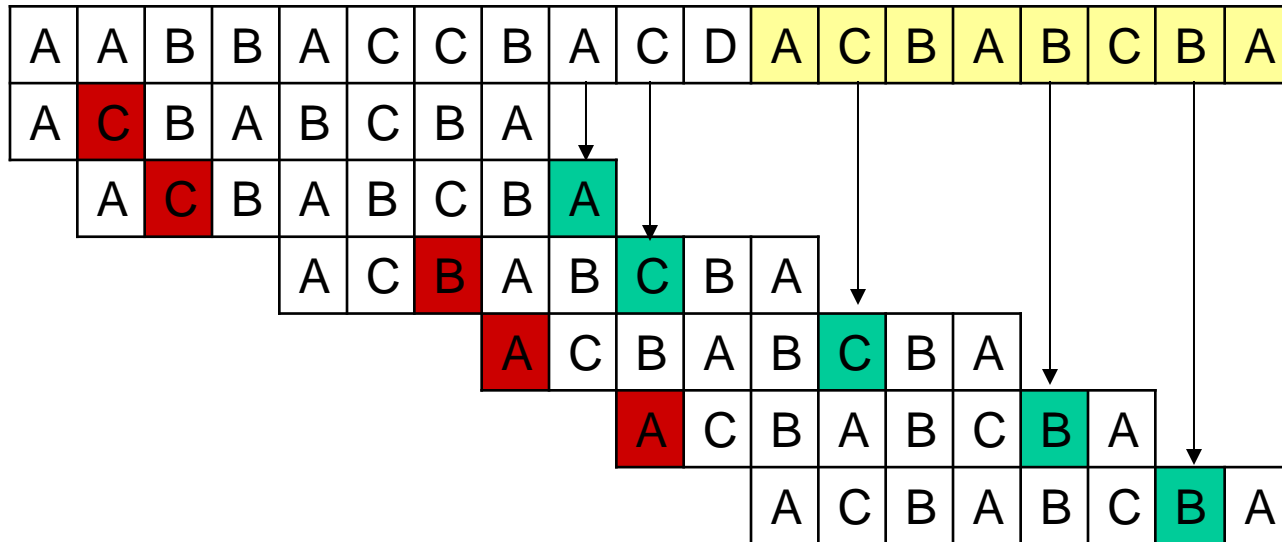


Antwort: Das Muster wird über den Buchstaben hinweggeschoben.

Alle Positionen vorher sind zwecklos, da der Buchstabe ja im Muster nicht vorkommt.

```
text U N G L E I C H U N G S - T E I L . . .
pattern U N G L E I C H U N G E N
```

# Beispiel 1

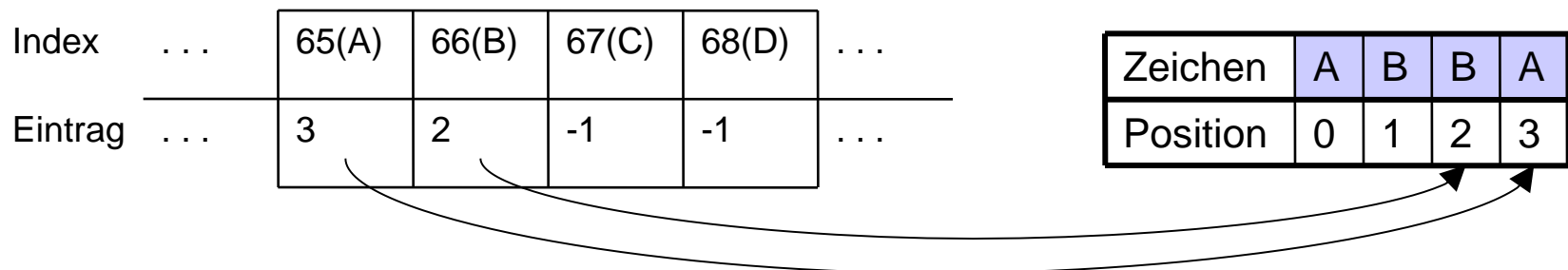


## last-Tabelle

...enthält zu jedem Zeichen des Zeichensatzes die Position des letzten Vorkommens im Muster (oder '-1', falls es nicht vorkommt).

Implementierung z.B. als Array indiziert mit (Unicode)-Zeichensatz: 'A' auf Index 65, 'B' auf 66 usw., 'a' auf Index 97, 'b' auf 98 usw..

Auszug aus last-Tabelle am Beispielpattern: "ABBA"





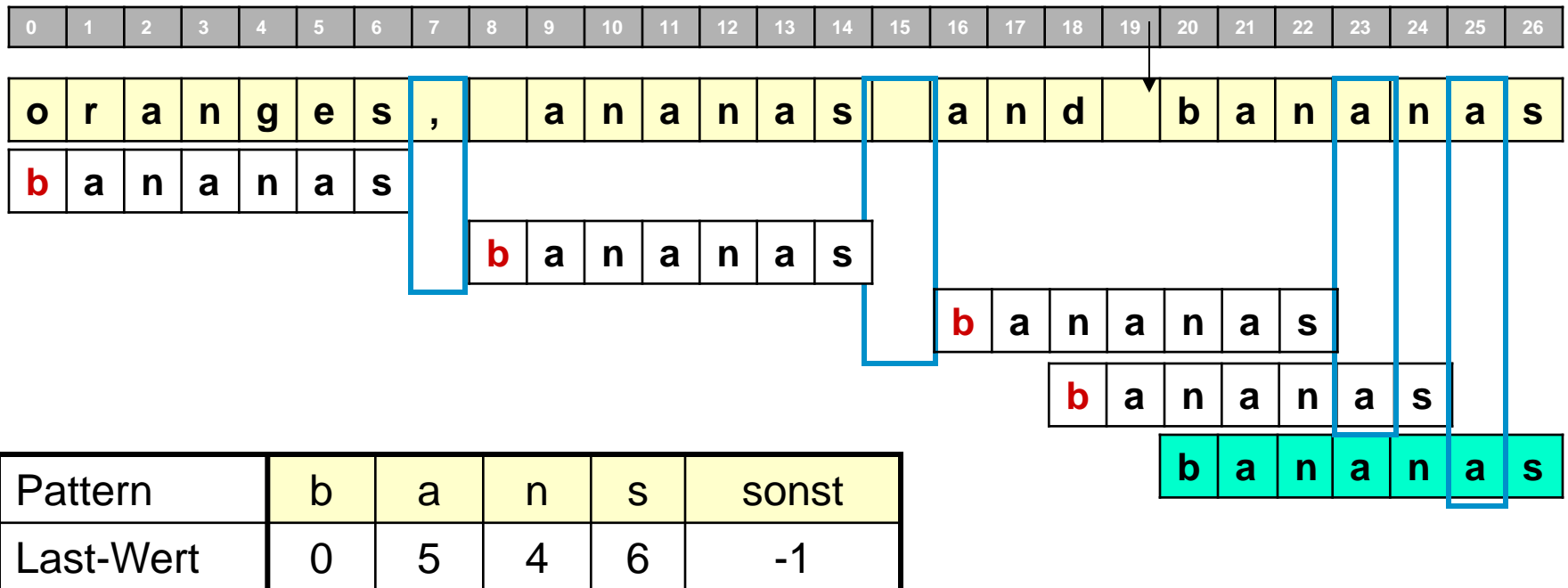
### Last-Tabelle des Musters „bananas“:

|                 |    |    |     |    |     |    |     |
|-----------------|----|----|-----|----|-----|----|-----|
| Pattern         | b  | a  | n   | a  | n   | a  | s   |
| Index im Muster | 0  | 1  | 2   | 3  | 4   | 5  | 6   |
| Unicode-Index   | 98 | 97 | 110 | 97 | 110 | 97 | 115 |
| Last-Wert       | 0  | 5  | 4   | 5  | 4   | 5  | 6   |

|      |    |    |        |     |         |     |         |
|------|----|----|--------|-----|---------|-----|---------|
| ...  | a  | b  | ...    | n   | ...     | s   | ...     |
| 0-96 | 97 | 98 | 99-109 | 110 | 111-114 | 115 | 116-127 |
| -1   | 5  | 0  | -1     | 4   | -1      | 6   | -1      |

# Boyer/Moore/Sunday (Beispiel 2)

## Suche nach Pattern ‚bananas‘



Alte Position des Patterns:  $i \dots i+m-1$   
 Verschiebedistanz  $v = m - \text{last}[\text{text}[i+m]]$   
 Neue Position des Patterns:  $i+v \dots i+v+m-1$

- Laufzeit dieser Variante (unter Annahme  $n \gg m$ ):
  - $O(n \cdot m)$  im schlimmsten Fall  
Bsp. f. ungünstige Text/Pattern-Kombination: Text:  $a^*$  Muster: aaaaaba
  - Im Normalfall (Text und Muster sind sich sehr unähnlich) geht es aber viel schneller. Wenn Alphabet des Textes groß im Vergleich zu  $m$ : nur etwa  $O(n/m)$  Vergleiche
- Nur sinnvoll, wenn das Alphabet groß ist (z.B. ASCII/Unicode). Für Bitstrings ist dieses Verfahren weniger gut geeignet.

- Prinzip:
  - Vom Pattern wird ein Hashwert gebildet.
  - Dieser wird mit dem Hashwert des jeweiligen Textausschnitts verglichen.
  - Nur wenn beide Hashwerte gleich sind, werden Textausschnitt und Pattern zeichenweise verglichen
- Beispiel:
  - Pattern: ANE
  - Text: BENANE
  - Hashverfahren: Die Positionen der Buchstaben im Alphabet werden addiert.

- Hashwert des Patterns:
  - $A+N+E = 1+14+5 = 20$
- Hashwert der ersten 3 Buchstaben des Texts:
  - $B+E+N = 2+5+14 = 21$
  - **Keine Übereinstimmung**

- Das Pattern wird einen Buchstaben weitergerückt.
  - Der erste Buchstabe fällt aus dem Pattern heraus und wird subtrahiert. Der neue Buchstabe wird addiert.
  - Je nach Hashfunktion können diese Operationen auch anders aussehen.

|     |    |   |   |    |   |   |     |
|-----|----|---|---|----|---|---|-----|
| alt | B  | E | N | A  | N | E | neu |
| 21  | -2 |   |   | +1 |   |   | 20  |

- Der Hashwert von Pattern und Textausschnitt stimmen jetzt überein.
  - Zeichenweiser Vergleich: Textausschnitt: ENA, Pattern: ANE
  - Kein Treffer

- Das Pattern wandert weiter durch den Text

|     |   |    |     |   |     |    |     |
|-----|---|----|-----|---|-----|----|-----|
| alt | B | E  | N   | A | N   | E  | neu |
| 20  |   | -5 |     |   | +14 |    | 29  |
| 29  |   |    | -14 |   |     | +5 | 20  |

- Zeichenweiser Vergleich  
– **Treffer**

- Im Worst Case sind die Hashwerte immer gleich und Rabin-Karp hat keine Vorteile gegenüber dem naiven Verfahren  $\rightarrow O(n \cdot m)$ .



- *String.indexOf(String str)* verwendet naive Textsuche.
  - Python ebenso
  - Ruby verwendet Rabin/Karp

- Bisher gesucht: konkrete Zeichenfolge
- Jetzt: Suche nach allgemeinerem Muster
- Dazu benötigt:
  - Notation, um allgemeine Muster zu beschreiben  
⇒ **reguläre Ausdrücke**
  - Mechanismus, um solche Muster zu „erkennen“  
⇒ **endliche Automaten**

- Die Notation, in der die Muster beschrieben werden, heißt „reguläre Ausdrücke“.
  - Die Bezeichnung „regulärer Ausdruck“ kann man etwa so interpretieren: „Eine Menge von Ausdrücken, die durch (einfache) Regeln beschrieben wird“.
- Jeder reguläre Ausdruck beschreibt eine Menge von Zeichenfolgen, nämlich alle, die dem Muster entsprechen.

**Zeichen** - z.B. Buchstabe, Ziffer

**Alphabet** - endliche Menge von Zeichen, z. B.  $\Sigma = \{a,b,c\}$

**Wort über Alphabet  $\Sigma$**

- endliche Folge von Zeichen aus  $\Sigma$ , z.B. abcb
- Spezialfall: "leeres Wort"  $\epsilon$

**$\Sigma^*$**

- Menge aller Wörter über  $\Sigma$ ,  
z. B.  $\Sigma = \{a,b\} \Rightarrow \Sigma^* = \{ \epsilon, a,b,aa,ab,ba,bb,aaa,\dots \}$

**Sprache  $L$  über Alphabet  $\Sigma$**

- Teilmenge:  $L \subseteq \Sigma^*$  (auch  $\{\epsilon\}$  ist Sprache über  $\Sigma$ ).

Sonderfall:  
→ reguläre Sprachen

### Regulärer Ausdruck (in der theoretischen Informatik):

- Ein regulärer Ausdruck ist eine „Formel“, die eine Sprache beschreibt, d.h. eine Teilmenge aller möglichen Worte definiert.
- Der Begriff **Regulärer Ausdruck** hat (vor allem in der Unix-Welt) noch eine andere, verwandte Bedeutung.
  - **wird am Ende des Kapitels erklärt.**

- Die Regeln für reguläre Ausdrücke setzen sich aus 3 grundlegenden Operationen zusammen.
  - **Verkettung**
  - **Oder**
  - **Hüllenbildung**

- Die erste Operation heißt „Verkettung“ (Concatenation)
- Zwei oder mehr Buchstaben werden durch diese Operation aneinandergehängt, z.B. **AB**. Der Operator wird nicht mitgeschrieben.
- Das heißt schlicht und einfach, dass die Buchstaben im Text direkt hintereinander stehen müssen.
  
- Würde man nur diese eine Operation zulassen, wäre man bei der „einfachen“ Textsuche.

- Die zweite Operation heißt „Oder“ (Or)
- Sie erlaubt die Angabe von Alternativen im Muster
- Schreibweise:  $(A|B)$  für „entweder A oder B“.
- Beispiele:
  - $(A|B)(A|B)$  bedeutet: AA, AB, BA oder BB.
  - $(A|C)((B|C)D)$  bedeutet: ABD, CBD, ACD oder CCD.
  - $C(AC|B)D$  bedeutet: CACD oder CBD.



- Die dritte Operation heißt „Hüllenbildung“ (Closure).
- Sie erlaubt es, Teile des Musters beliebig oft zu wiederholen.
- Erlaubt ist auch 0 mal Wiederholung
- Schreibweise: Hinter den zu wiederholenden Buchstaben wird ein Stern (\*) gesetzt.
- Sind mehrere Buchstaben zu wiederholen, müssen sie in Klammern gesetzt werden.
- Beispiele:
  - **A\*** bedeutet:  $\varepsilon$ , A, AA, AAA, AAAA, AAAAA, ...
    - $\varepsilon$  bedeutet „Leerstring“
  - **(ABC)\*** bedeutet:  $\varepsilon$ , ABC, ABCABC, ABCABCABC, ...
  - **DA\*B** bedeutet: DB, DAB, DAAB, DAAAB, ...

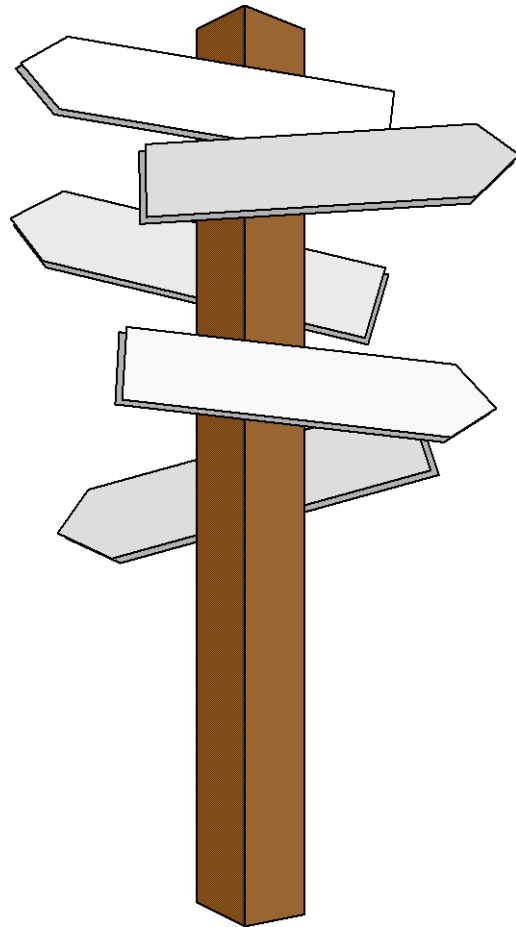
- Klammern binden am stärksten.
- Es folgt die Hüllenbildung (\*).
- Dann folgt die Verkettung.
- Am schwächsten ist der oder-Operator (|).

- Beispiele:
- $a|a(a|b)^*a$  bedeutet:
  - $a | a(a|b)^*a$  entweder ein einfaches  $a$  oder
  - $a | a(a|b)^*a$  ein führendes  $a$ , gefolgt von
  - $a | a(a|b)^*a$  einer beliebigen Kombination von  $a$  und  $b$   
(z.B.  $abbab$ )
  - $a | a(a|b)^*a$  und einem abschließenden  $a$
- Zusammengefasst: Alle Zeichenketten aus  $a$  und  $b$ , die am Anfang und am Ende ein  $a$  enthalten.

## Beispiel (2)

- $(1+01)^*(0+1)$  bedeutet:
  - $(1+01)^*(0+1)$  Eine beliebige Abfolge aus den Elementen „1“ und „01“
  - $(1+01)^*(0+1)$  gefolgt von einer abschließenden 0 oder 1.
- Zusammengefasst: Alle Kombinationen aus 1 und 0, in denen nicht mehrere 0 aufeinander folgen.

- In der theoretischen Informatik sind die Regeln für reguläre Ausdrücke:
  - $XY$ : **Verkettung**
  - $X|Y$ : **Oder**
  - $X^*$ : **Hüllenbildung**
- In Unix wurden diese Regeln in einigen Werkzeugen (ed, grep) umgesetzt.
  - **Es wurden aber der Bequemlichkeit halber viele weitere Operatoren hinzugefügt.**
- Daraus entstanden die **Perl Compatible Regular Expressions (PCRE)**, die meistens einfach „reguläre Ausdrücke“ genannt werden.



## Vorlesung 20

### Reguläre Ausdrücke

- Standardisierte Regeln zur Erzeugung regulärer Ausdrücke für die Textsuche.
- Sind in vielen Editoren und Programmiersprachen vorhanden.
  - **Z.B. Eclipse:** Im Dialogfeld „Find/Replace“ das Auswahlfeld „Regular expressions“ anklicken.
  - **Z.B. OpenOffice:** Im Dialogfeld „Find & Replace“ „More Options“ auswählen und anschließend „Regular expressions“ anklicken.
  - **Z.B. Java, C#, Python, Ruby, Perl, ...**

- Verknüpfungen

$AB$  Zeichenfolge  $AB$   
 $A|B$   $A$  oder  $B$   
 $[AB]$  Zeichenklasse „ $A$  oder  $B$ “

- Quantoren

$A\{n\}$   $A$  kommt genau  $n$  mal vor.  
 $A\{min, \}$   $A$  kommt mindestens  $min$  mal vor.  
 $A\{min, max\}$   $A$  kommt mindestens  $min$  und höchstens  $max$  mal vor.

- Abkürzungen für Quantoren

$A?$  entspricht  $A\{0, 1\}$   
 $A^*$  entspricht  $A\{0, \}$   
 $A^+$  entspricht  $A\{1, \}$   
 $A$  entspricht  $A\{1\}$



- Zeichenklassen
  - `\w` Buchstaben (Word)
  - `\d` Zahlen (Digit)
  - `.` Alles außer Zeilenvorschub
- Referenzen
  - `()` Gruppierung
  - `\x` x-te Rückwärtsreferenz (je nach Implementierung auch `$x`).
- Greedy
  - (default) Greedy (gierig)
  - ? Reluctant, non-greedy (genügsam)

- „A oder B“ lässt sich auf zwei Arten ausdrücken:
  - Mit Hilfe des oder-Operators  $A | B$
  - Mit Hilfe von Zeichenklassen  $[AB]$
- Vorteile des oder-Operators:
  - Die Alternativen können komplexe Ausdrücke sein:  $A | (BC) *$
- Vorteile von Zeichenklassen:
  - Die Auswertung geht schneller (siehe nachfolgende Folien).

| Greedy | Reluctant<br>(non-greedy) |
|--------|---------------------------|
| $A^*$  | $A^*?$                    |
| $A^+$  | $A^+?$                    |
| $A?$   | $A??$                     |

- $[AB]$  ergibt **keine** Verzweigung (kein Backtracking, schneller)

- Auch in Java kann man PCRE gebrauchen, z.B. in `String.replaceFirst(..)`, `String.split(..)`, `String.matches(..)`
  - **Siehe auch das nachfolgende Beispiel**

- Finden eines regulären Ausdrucks in Java

```
import java.util.*;
import java.io.*;
import java.util.regex.*;
..
public void testRegex() {
 //Testdatei laden
 Scanner sc = new Scanner(new File("pangalak.txt"));
 String txt = "";
 while (sc.hasNextLine()) {
 txt = txt + sc.nextLine()+"\n";
 }
}
```

```
//Vorbereiten des regulaeren Ausdrucks
Pattern p = Pattern.compile("n[a-z]*e");
Matcher m = p.matcher(txt);
//Suchen
while (m.find()) {
 System.out.println("Start: "+m.start()+" String:"
 +m.group());
}
}
```

- **Ausgabe**

Start: 17, String: ngalaktische

Start: 33, String: nnergurgle

Start: 51, String: nehme

...

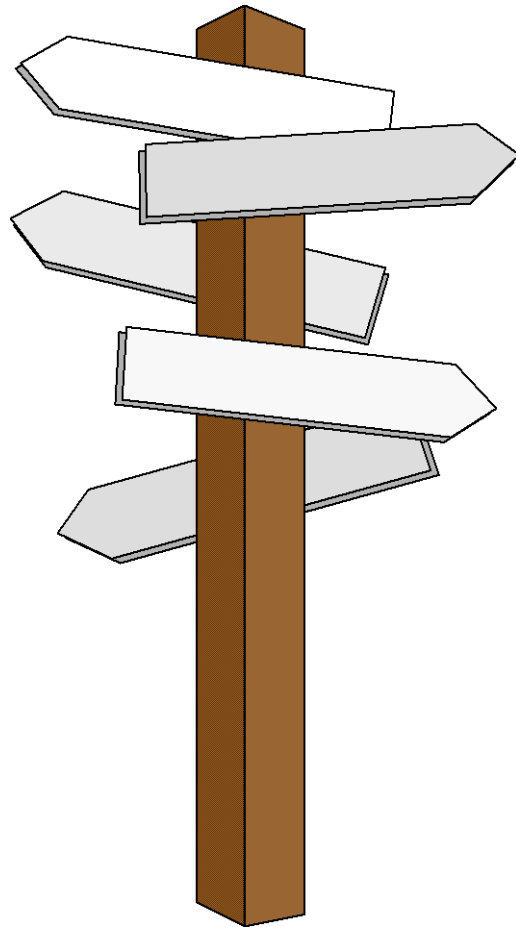
- Klammern fangen Referenzen, die intern durchnummeriert werden.
- Die Referenzen können mit Hilfe der Nummern umgesetzt werden.
- Beispiel in Java:
  - **Ersetzen aller Doppelbuchstaben durch einfache Buchstaben:**

```
s.replaceAll ("(\\w)\\1", "$1");
```

- \\w erkennt einen beliebigen Buchstaben.
- Die Klammern „fangen“ das mit \\w erkannte Zeichen und legen es in einer Liste ab. Elemente der Liste können mit \\1, \\2, \\3, ... referenziert werden.
- Für \\1 wird also das mit \\w erkannte Zeichen eingesetzt.
- Im Replace-String benötigt man statt \\1 dazu \$1 (Java-Eigenart).

- Andere Varianten von regulären Ausdrücken:
  - **Basic Regular Expressions (BRE) (POSIX-Standard)**
  - **Extended Regular Expressions (ERE) (POSIX-Standard)**
- Vereinfachte (unvollständige) Varianten:
  - **Windows (DOS-) Wildcards**
  - **MS Office-Platzhalterzeichen**





## Vorlesung 21

### RegEx-Auswertung mit endlichen Automaten

Zuerst werden wir mit regulären Ausdrücken ermitteln:

- ob ein Textstring einem regulären Ausdruck entspricht.

Später werden wir:

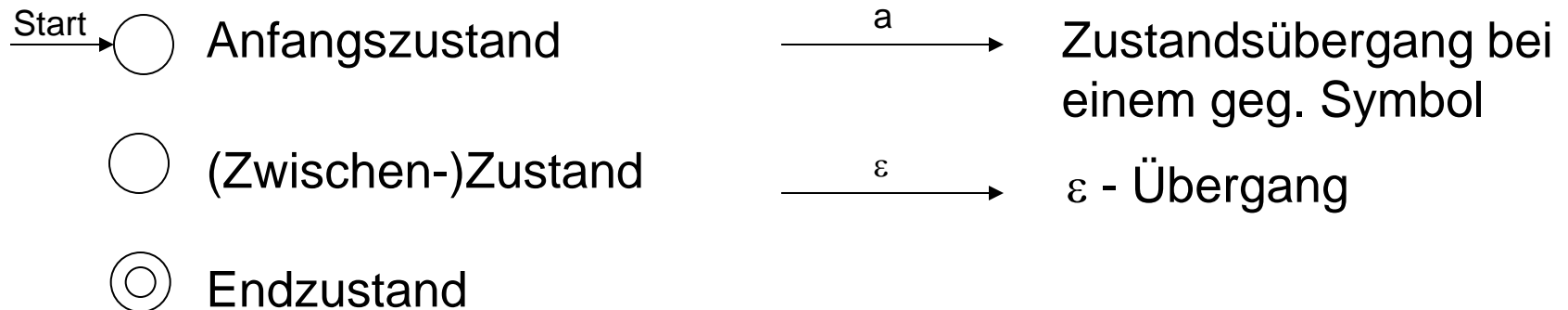
- einen regulären Ausdruck in einem längeren Text suchen.

**Endlicher Automat (finiter Automat):** abstraktes Maschinenmodell

- **Aufgabe:** Entscheiden, ob Wort zur Sprache gehört, die durch regulären Ausdruck beschrieben ist (Akzeptoren).
- Anfangs: Maschine ist in „**Anfangszustand**“
- In jedem Schritt wird ein Eingabesymbol  $\sigma$  „gelesen“. Abhängig von  $\sigma$  geht die Maschine von einem Zustand in einen bestimmten anderen über.
- Wenn nach Lesen des letzten Zeichens ein „**Endzustand**“ erreicht ist, ist das Muster ist „erkannt“ (Wort gehört zu Sprache).

- Es ist möglich, einen endlichen Automaten graphisch zu simulieren.
- Besonders einfach ist das Konstruktions-Verfahren nach Kleene, das zu einem **nichtdeterministischen finiten Automaten** (NFA) führt.
- Es gibt immer auch einen entsprechenden **deterministischen finiten Automaten** (DFA). Dort gibt es keine  $\varepsilon$ -Übergänge.

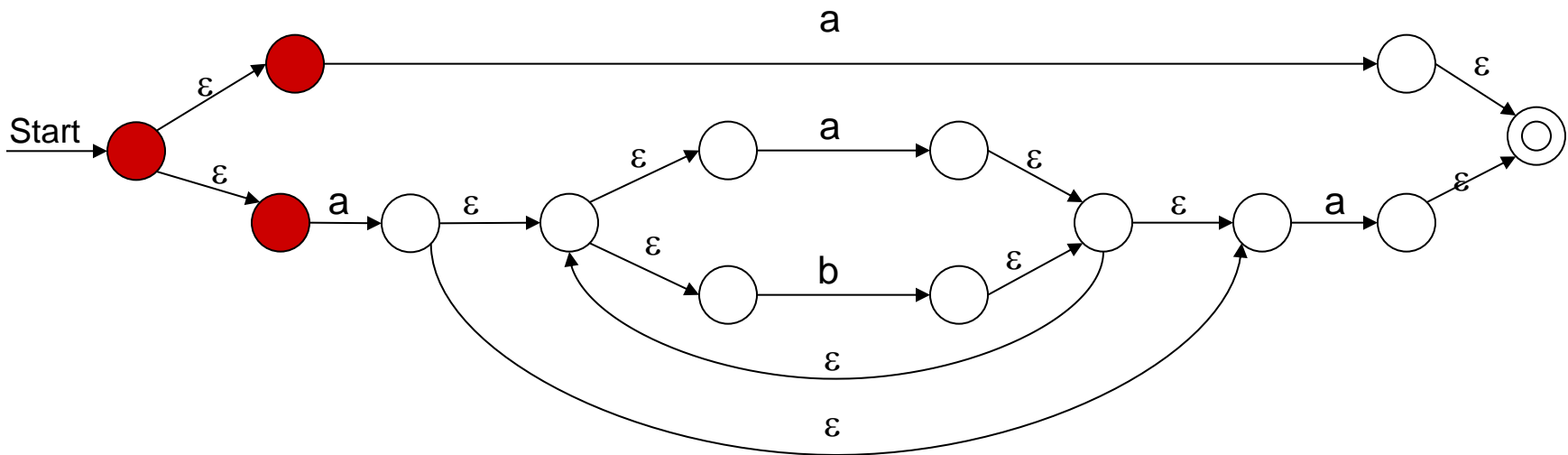
- Grundlage ist das Zustands-Übergangs-Diagramm mit folgenden Elementen.



1. Initialisierung
  1. **Markiere den Anfangszustand**
  2. **Markiere alle Zustände, die durch  $\varepsilon$ -Übergänge erreichbar sind.**
2. Für jedes gelesene Eingabesymbol
  1. **Markiere alle Zustände, die durch dieses Eingabesymbol erreichbar sind.**
  2. **Lösche alle anderen Zustände.**
  3. **Markiere alle Zustände, die jetzt durch  $\varepsilon$ -Übergänge erreichbar sind.**

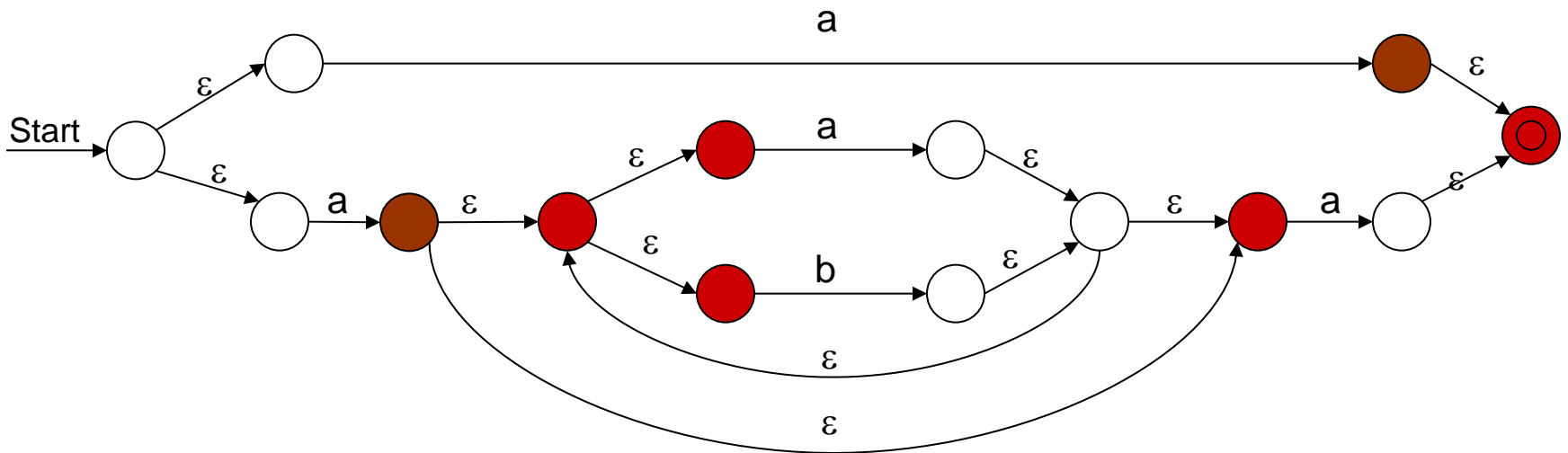
# Ablauf der Simulation (1)

- Text „ABBA“
- Automat „ $A|A(A|B)^*A$ “ (Konstruktion später)
- Anfangszustand



## Ablauf der Simulation (2)

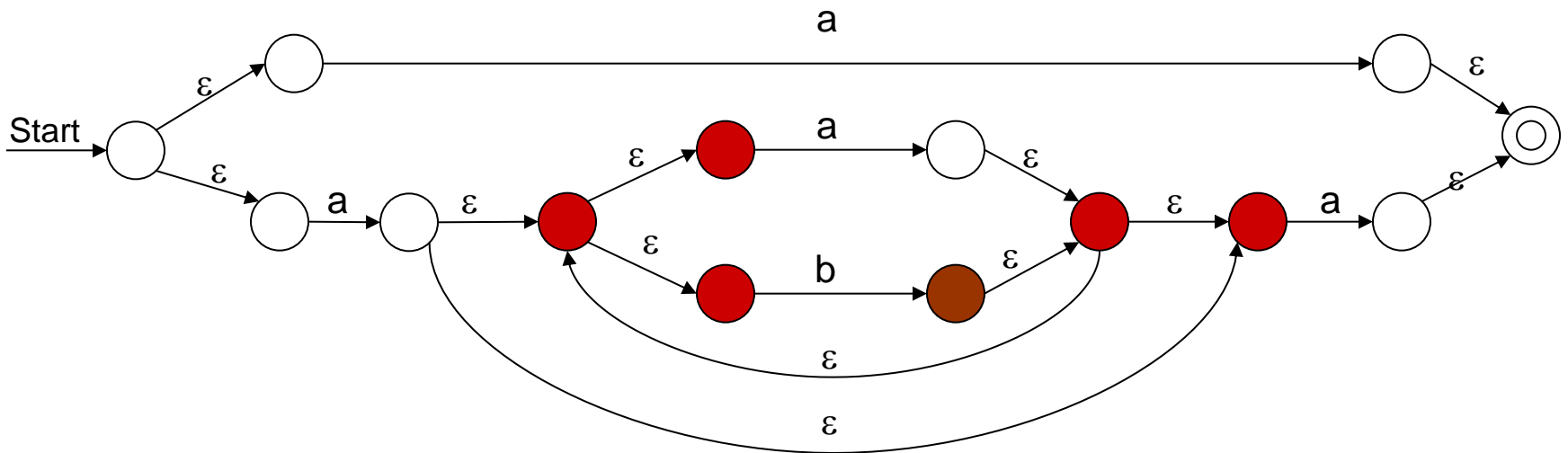
- Text „ABBA“
- 1. Buchstabe: A





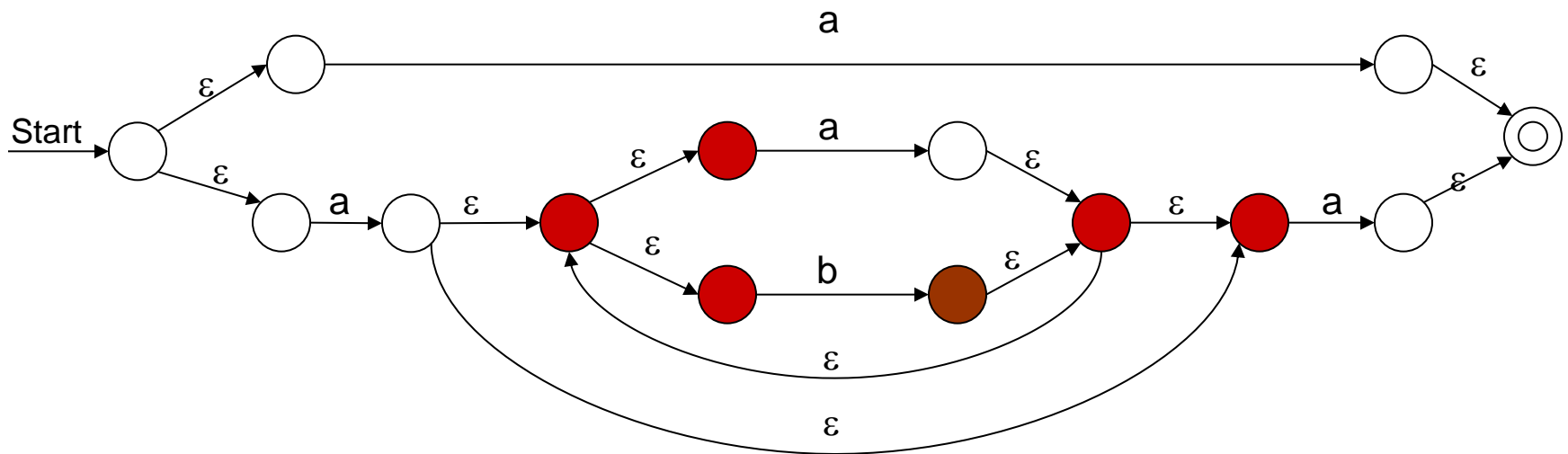
## Ablauf der Simulation (3)

- Text „ABBA“
- 2. Buchstabe: B



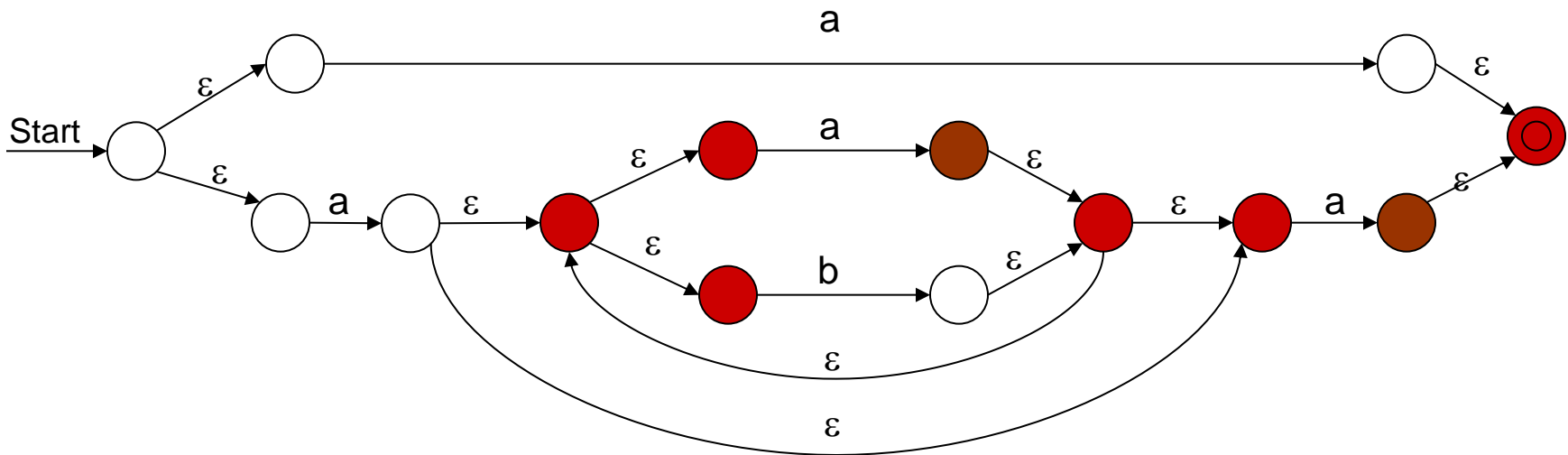
## Ablauf der Simulation (4)

- Text „ABBA“
- 3. Buchstabe: B



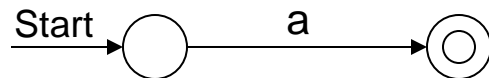
## Ablauf der Simulation (5)

- Text „ABBA“
- 4. Buchstabe: A

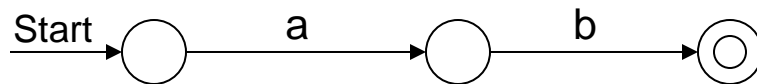


ABBA wird durch den regulären Ausdruck  $A|A(A|B)^*A$  beschrieben.

- Einzelnes Symbol: Regulärer Ausdruck „a“:

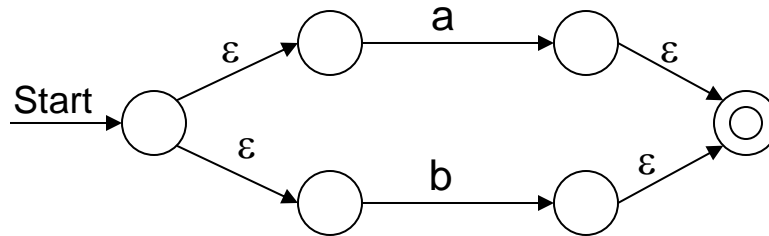


- Verkettung: Regulärer Ausdruck „ab“:

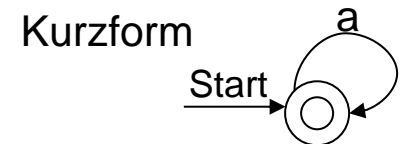
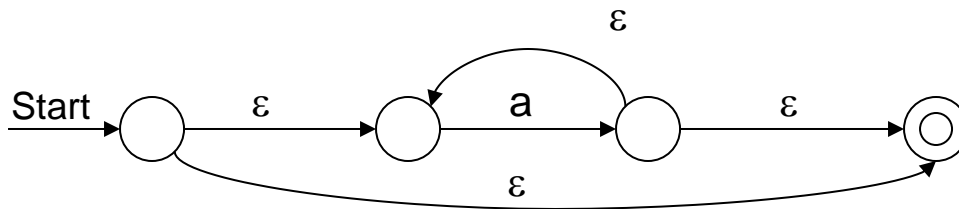


# Konstruktion des Automaten (2)

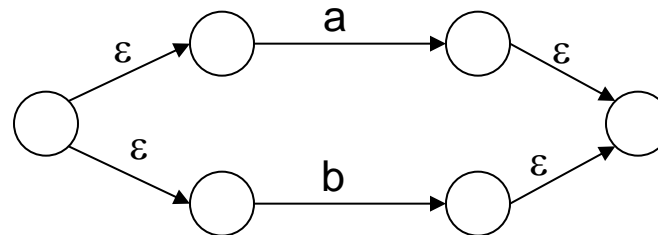
- Oder-Veknüpfung  $a|b$



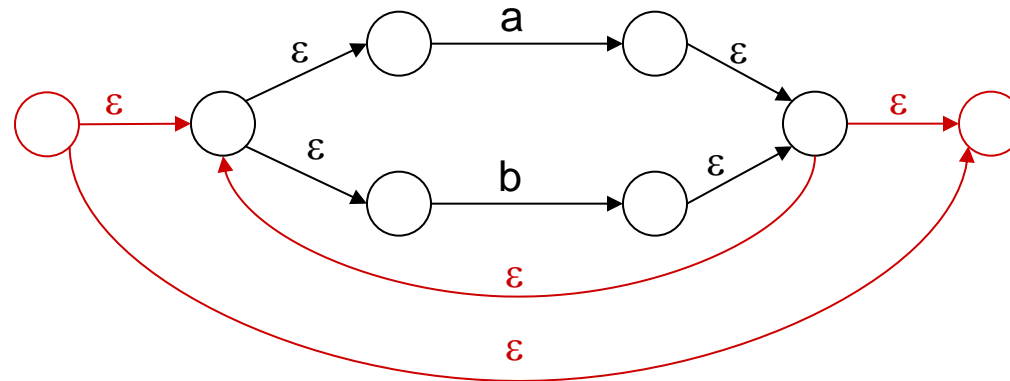
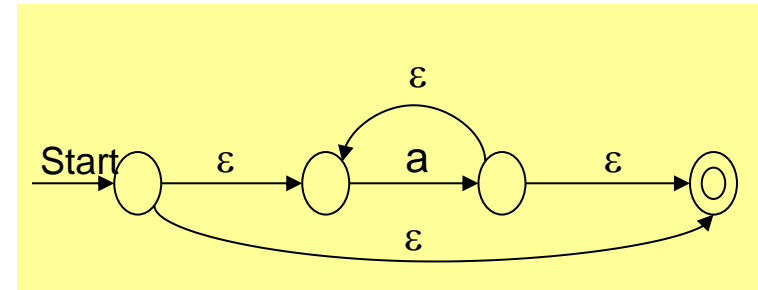
- Hüllenbildung  $a^*$



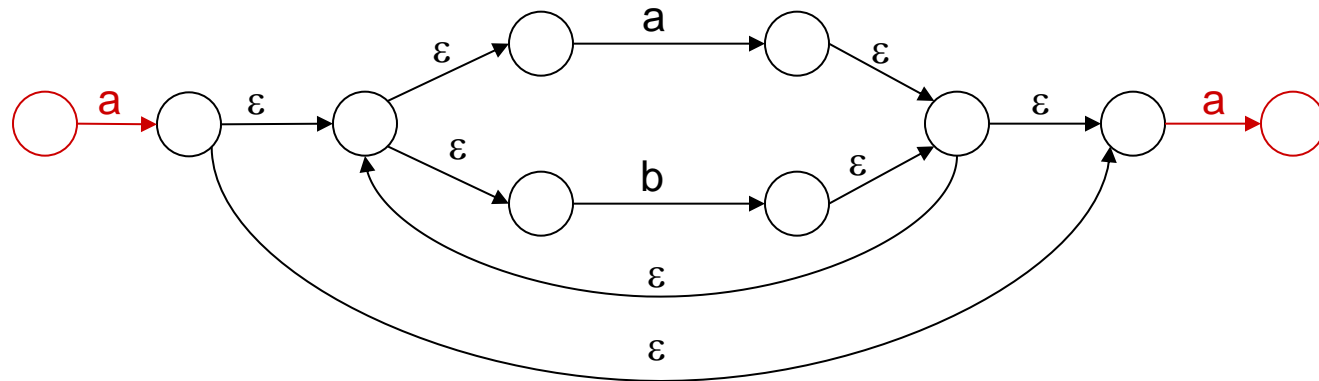
- Aufbau des Automaten für:  $a|a(a|b)^*a$
- 1. Schritt:  $a|b$



- Aufbau des Automaten für:  $a|a(a|b)^*a$
- 2. Schritt:  $(a|b)^*$



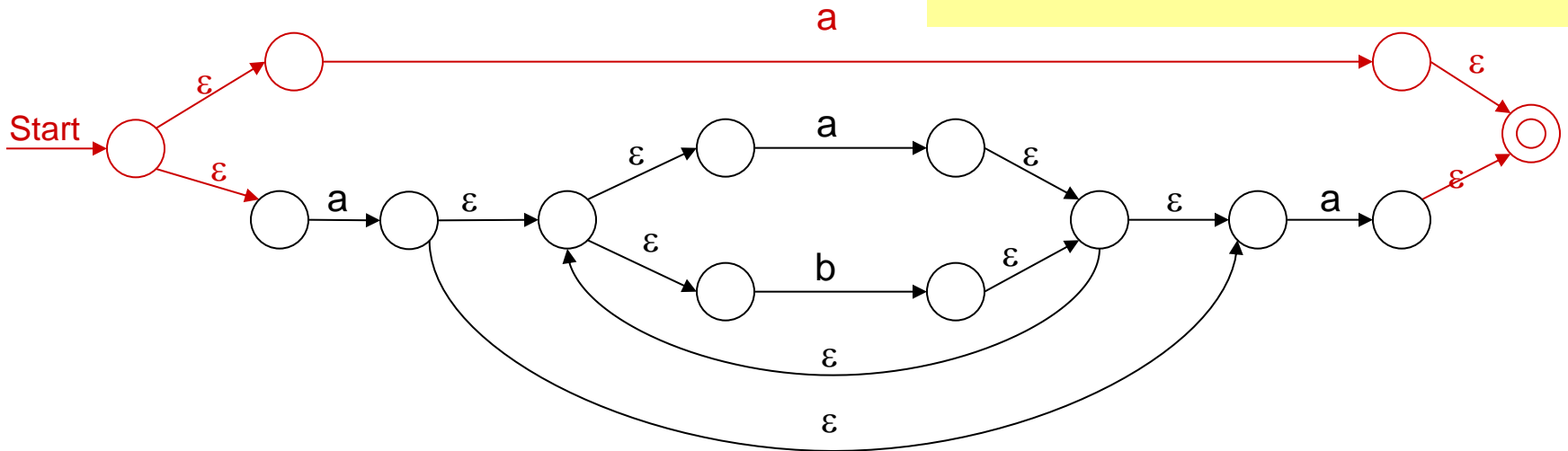
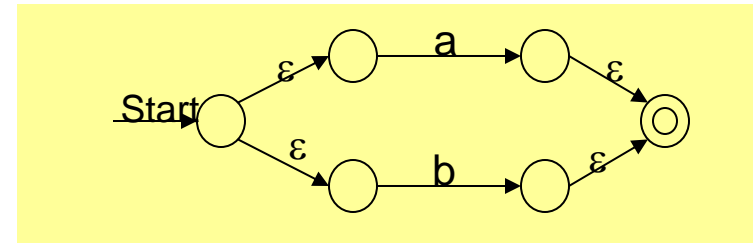
- Aufbau des Automaten für:  $a|a(a|b)^*a$
- 3. Schritt:  $a(a|b)^*a$





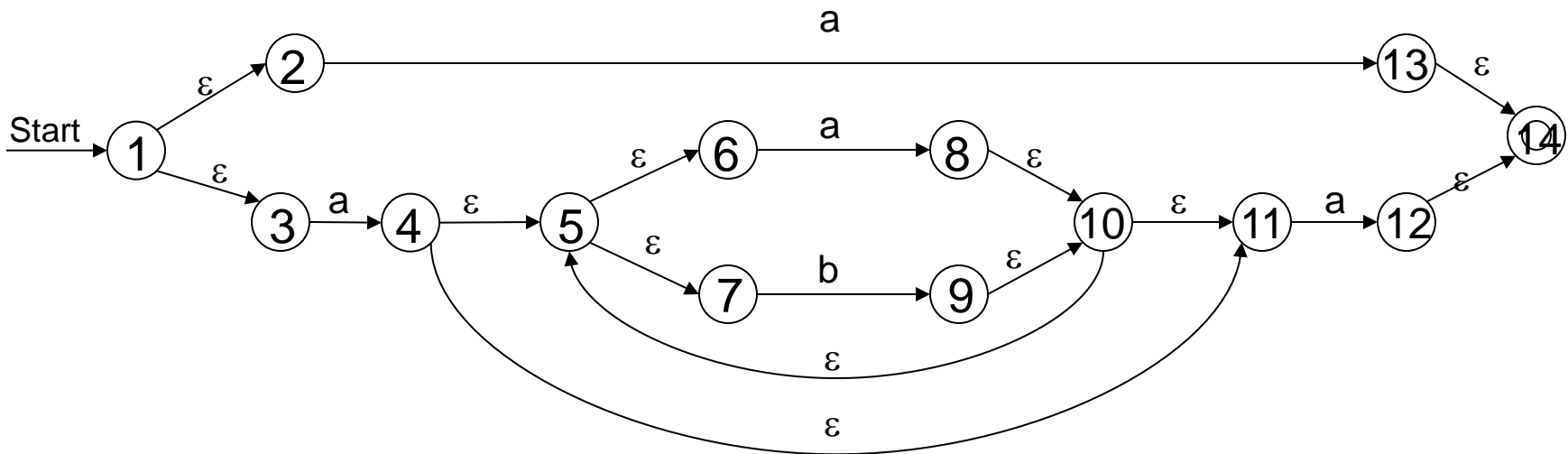
# Kombination der Elemente

- Aufbau des Automaten für:  $a|a(a|b)^*a$
- 4. Schritt: **a** $|a(a|b)^*a$



- Zustands-Übergangs-Diagramm ist ein Graph
- Speicherung in Tabellenform (z.B. Adjazenzliste)
- Neuer Zustand lässt sich aus Tabelle ermitteln.

|   |                           |  |    |                           |
|---|---------------------------|--|----|---------------------------|
| 1 | $\epsilon 2, \epsilon 3$  |  | 8  | $\epsilon 10$             |
| 2 | a13                       |  | 9  | $\epsilon 10$             |
| 3 | a4                        |  | 10 | $\epsilon 11, \epsilon 5$ |
| 4 | $\epsilon 5, \epsilon 11$ |  | 11 | a12                       |
| 5 | $\epsilon 6, \epsilon 7$  |  | 12 | $\epsilon 14$             |
| 6 | a8                        |  | 13 | $\epsilon 14$             |
| 7 | b9                        |  | 14 |                           |



# Letzter Zustandsübergang (Wiederholung)

1  
(a)

|   |                             |  |    |                             |
|---|-----------------------------|--|----|-----------------------------|
| 1 | $\epsilon_2, \epsilon_3$    |  | 8  | $\epsilon_{10}$             |
| 2 | a13                         |  | 9  | $\epsilon_{10}$             |
| 3 | a4                          |  | 10 | $\epsilon_{11}, \epsilon_5$ |
| 4 | $\epsilon_5, \epsilon_{11}$ |  | 11 | a12                         |
| 5 | $\epsilon_6, \epsilon_7$    |  | 12 | $\epsilon_{14}$             |
| 6 | a8                          |  | 13 | $\epsilon_{14}$             |
| 7 | b9                          |  | 14 |                             |

2  
( $\epsilon$ )

|   |                             |  |    |                             |
|---|-----------------------------|--|----|-----------------------------|
| 1 | $\epsilon_2, \epsilon_3$    |  | 8  | $\epsilon_{10}$             |
| 2 | a13                         |  | 9  | $\epsilon_{10}$             |
| 3 | a4                          |  | 10 | $\epsilon_{11}, \epsilon_5$ |
| 4 | $\epsilon_5, \epsilon_{11}$ |  | 11 | a12                         |
| 5 | $\epsilon_6, \epsilon_7$    |  | 12 | $\epsilon_{14}$             |
| 6 | a8                          |  | 13 | $\epsilon_{14}$             |
| 7 | b9                          |  | 14 |                             |

3  
( $\epsilon$ )

|   |                             |  |    |                             |
|---|-----------------------------|--|----|-----------------------------|
| 1 | $\epsilon_2, \epsilon_3$    |  | 8  | $\epsilon_{10}$             |
| 2 | a13                         |  | 9  | $\epsilon_{10}$             |
| 3 | a4                          |  | 10 | $\epsilon_{11}, \epsilon_5$ |
| 4 | $\epsilon_5, \epsilon_{11}$ |  | 11 | a12                         |
| 5 | $\epsilon_6, \epsilon_7$    |  | 12 | $\epsilon_{14}$             |
| 6 | a8                          |  | 13 | $\epsilon_{14}$             |
| 7 | b9                          |  | 14 |                             |

4  
( $\epsilon$ )

|   |                             |  |    |                             |
|---|-----------------------------|--|----|-----------------------------|
| 1 | $\epsilon_2, \epsilon_3$    |  | 8  | $\epsilon_{10}$             |
| 2 | a13                         |  | 9  | $\epsilon_{10}$             |
| 3 | a4                          |  | 10 | $\epsilon_{11}, \epsilon_5$ |
| 4 | $\epsilon_5, \epsilon_{11}$ |  | 11 | a12                         |
| 5 | $\epsilon_6, \epsilon_7$    |  | 12 | $\epsilon_{14}$             |
| 6 | a8                          |  | 13 | $\epsilon_{14}$             |
| 7 | b9                          |  | 14 |                             |

## Letzter Zustandsübergang (2)

5  
( $\varepsilon$ )

|   |                                   |  |    |                                   |
|---|-----------------------------------|--|----|-----------------------------------|
| 1 | $\varepsilon_2, \varepsilon_3$    |  | 8  | $\varepsilon_{10}$                |
| 2 | a13                               |  | 9  | $\varepsilon_{10}$                |
| 3 | a4                                |  | 10 | $\varepsilon_{11}, \varepsilon_5$ |
| 4 | $\varepsilon_5, \varepsilon_{11}$ |  | 11 | a12                               |
| 5 | $\varepsilon_6, \varepsilon_7$    |  | 12 | $\varepsilon_{14}$                |
| 6 | a8                                |  | 13 | $\varepsilon_{14}$                |
| 7 | b9                                |  | 14 |                                   |

- **Beispiel**
- Text: AABABBAABBCBA
- Gesucht wird das Pattern:  $AB(A|B)^*A$
- Regel: **First, longest**
- Ansetzen des Patterns mit nichtdeterministischem endlichen Automaten:
  - +: Endzustand ist markiert.
  - o: Endzustand ist nicht markiert, aber Zwischenzustände sind markiert.
  - -: Kein Zustand ist markiert. Hier kann abgebrochen werden.

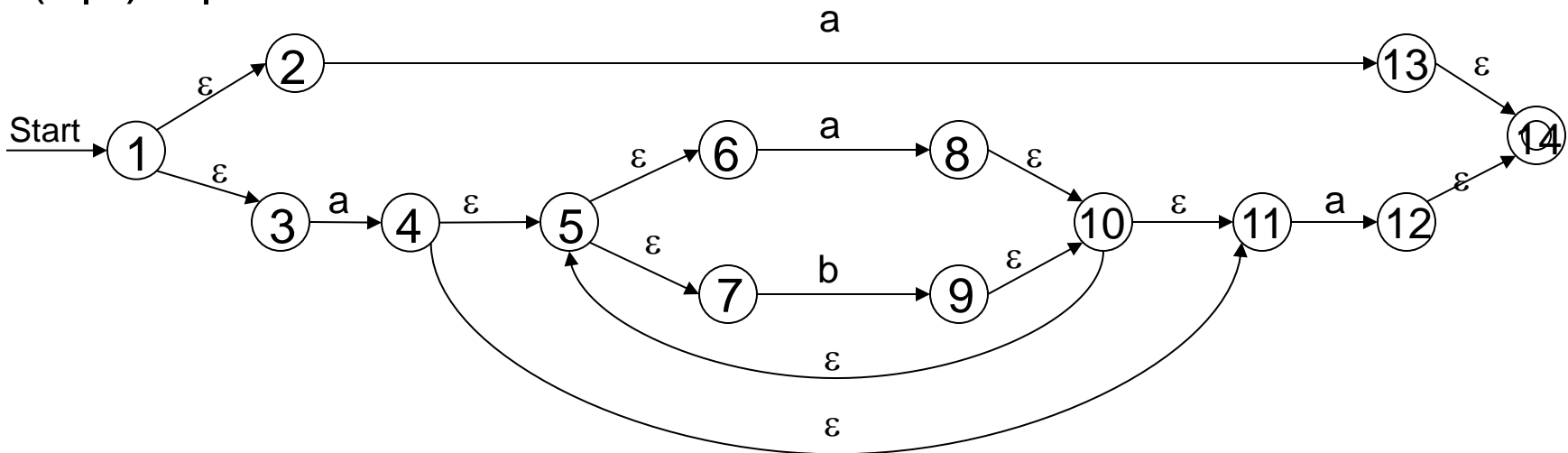
# Beispiel: Suchen nach reg. Ausdrücken

- Text: AABABBAABBCBA                      Pattern:  $AB(A|B)^*A$
- Ansatz am ersten Buchstaben → Pattern passt nicht  
AABABBAABBCBA  
○-
- Ansatz am zweiten Buchstaben  
AABABBAABBCBA  
○○+○○++○○-
  - Wenn das „-“ erreicht ist, nimmt man die Position des letzten „+“: **AABABBAA** → **First, longest**
  - First, shortest (erstes „+“; **ABA**) kann man auch ermitteln, dies wird aber selten gebraucht.
- Weitersuchen: Gewöhnlich setzt man **hinter dem gefundenen Pattern** wieder an.

## Auswertung von PCRE

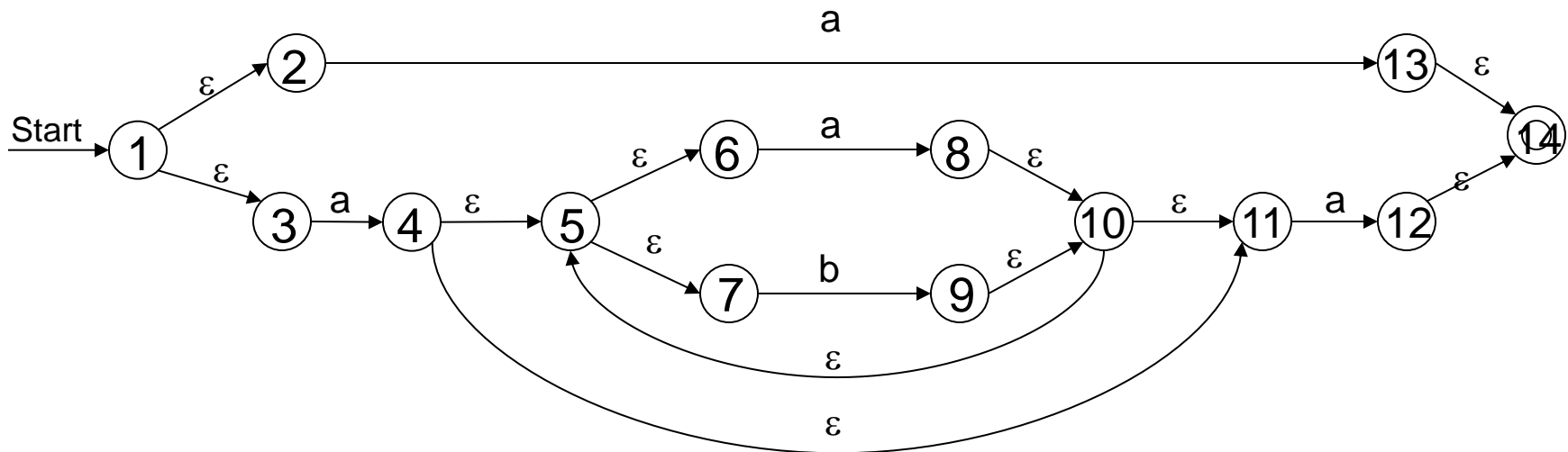
- Zur Auswertung von PCRE wird ein anderes Verfahren benutzt.
- Auch hier kann man von dem Graphen ausgehen, der bei der Simulation mit Kleene erzeugt wird.

$a(a|b)^*a|a$



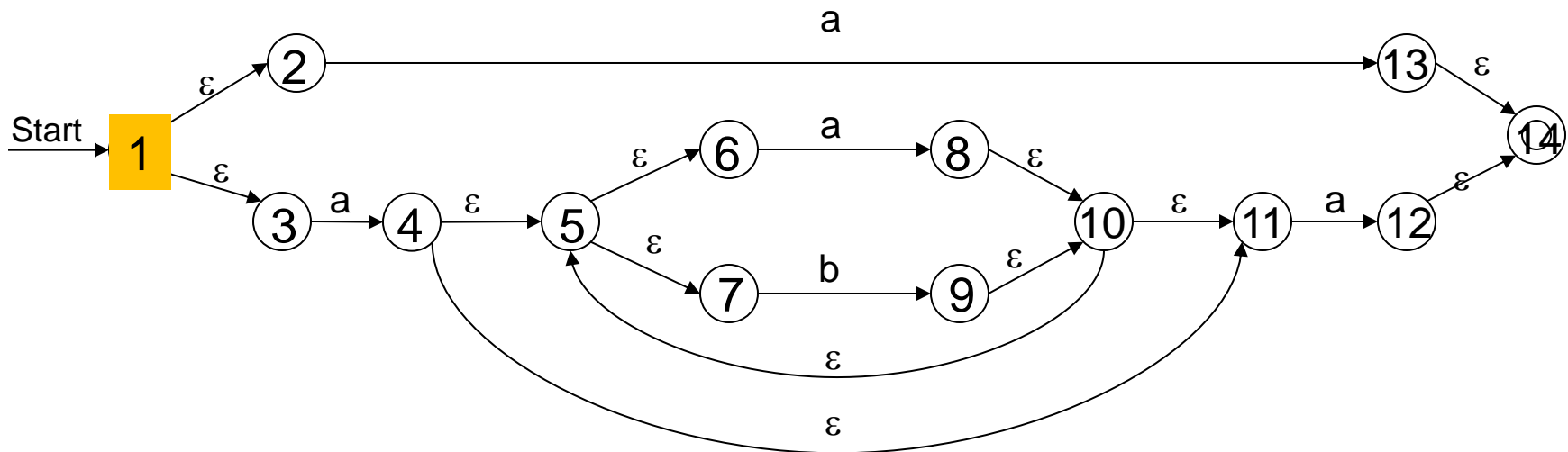


- Diesen Graphen geht man mit einer abgewandelten Tiefensuche durch.
- Man startet am Startpunkt



# Auswertung von PCRE: Beispiel 1 (1)

- Beispiel: Pattern:  $a(a|b)^*a|a$ ; Text: abba
- Man notiert sich für jeden Schritt die momentane Position im Text und die möglichen nächsten Schritte.
- Die momentane Position im Text („Cursor“) wird mit [] dargestellt.
  - **Beginn: []abba**

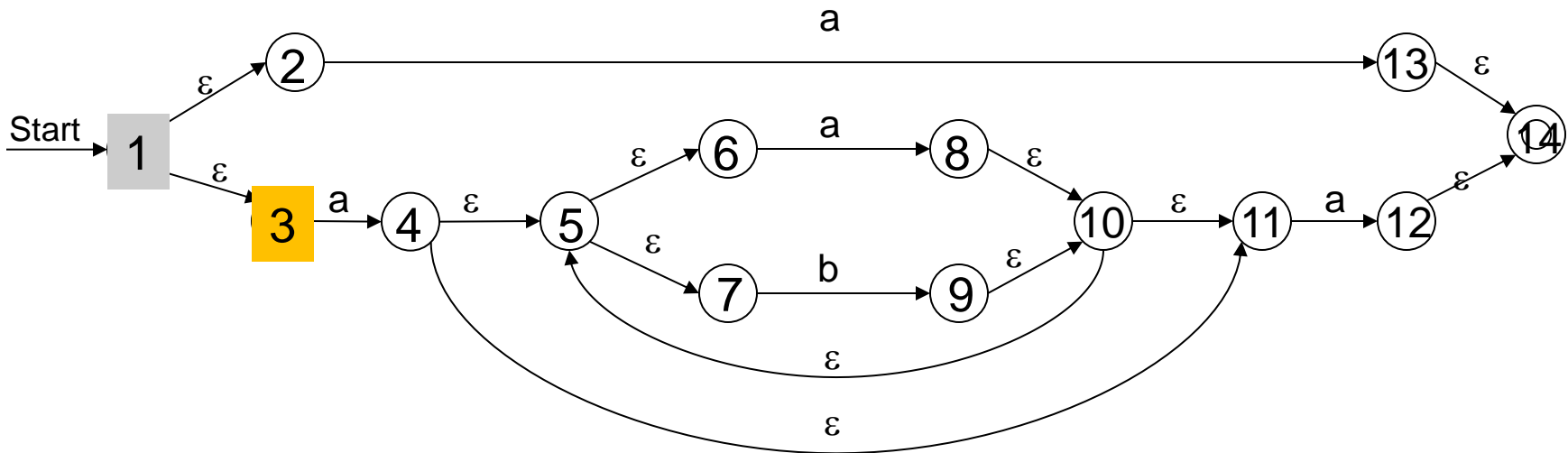


# Auswertung von PCRE: Beispiel 1 (2)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 1       | 1      | 2, 3        | []abba       |
| 2       | 3      | 4           | []abba       |

Zunächst steht die Entscheidung an, welche  $\epsilon$ -Kante verfolgt wird. Die Verzweigung ist Teil eines „oder“-Konstrukts.

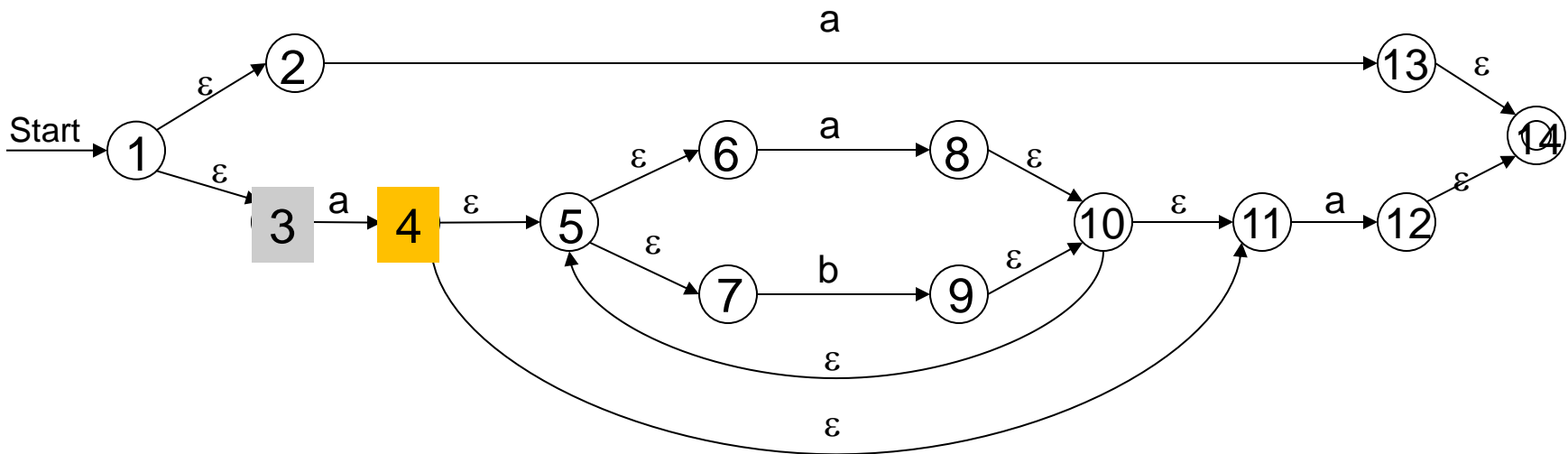
Regel: Bei oder-Entscheidungen wird zunächst die Kante verfolgt, die der linken Seite des „oder“ entspricht --> Knoten 3



# Auswertung von PCRE: Beispiel 1 (3)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 2       | 3      | 4           | []abba       |
| 3       | 4      | 5, 11       | a[]bba       |

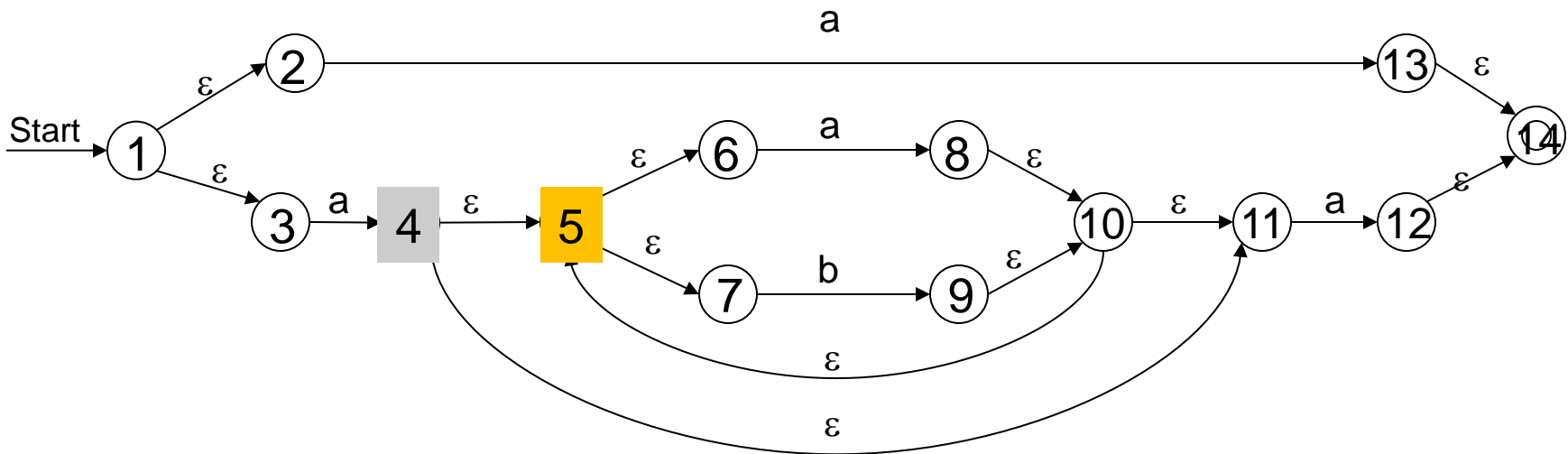
Jetzt kann mit dem Übergang 3->4 der erste Buchstabe verarbeitet werden. Der Cursor rückt eine Position weiter.



# Auswertung von PCRE: Beispiel 1 (4)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 3       | 4      | 5,11        | a[]bba       |
| 4       | 5      | 6, 7        | a[]bba       |

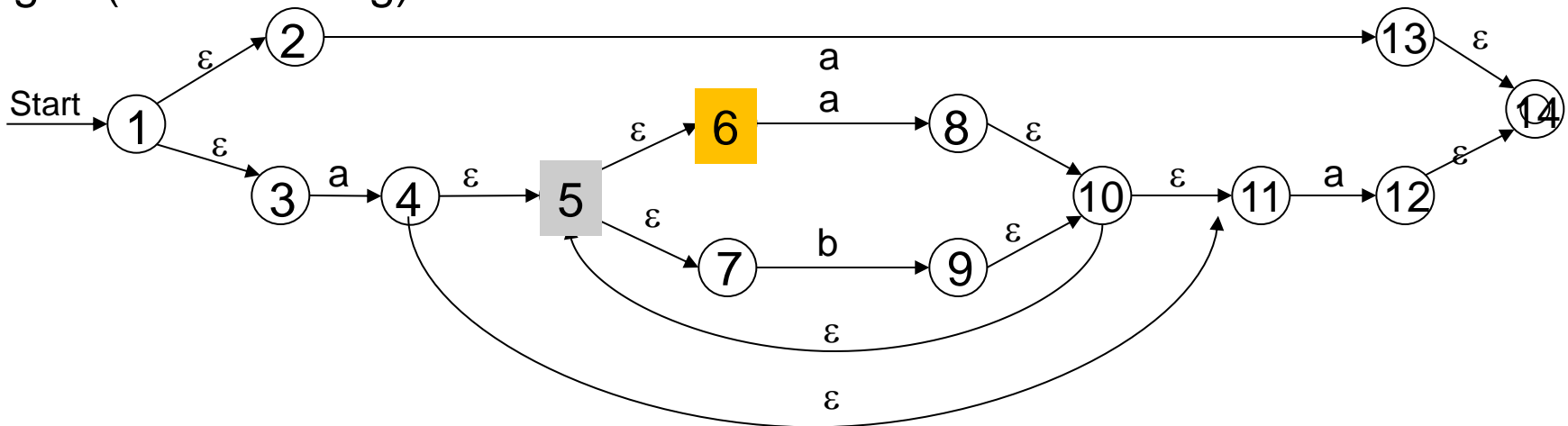
Von der 4 aus sind zwei Übergänge möglich, nämlich die Hüllenbildung (\*) zu überspringen oder nicht. Da ein Stern **greedy** ist, wird er nicht übersprungen und der Weg nach 5 gewählt.



# Auswertung von PCRE: Beispiel 1 (5)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 4       | 5      | 6, 7        | a[]bba       |
| 5       | 6      | 8           | a[]bba       |

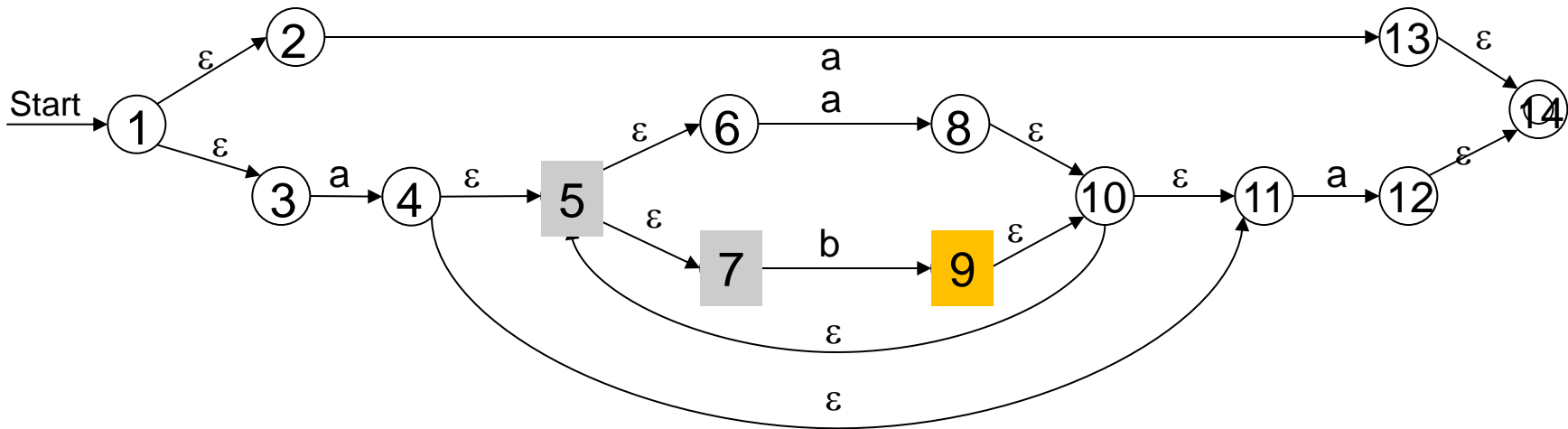
Bei der nächsten oder-Entscheidung wird wieder der linke Weg gewählt. Anschließend kann kein Übergang mehr ausgeführt werden, da nur ein a-Übergang möglich ist, das nächste Zeichen aber ein b ist. Also wird zum nächsten Punkt zurückgegangen, von dem es eine weitere Abzweigung gibt (Backtracking). Hier ist dies der Schritt 4.



# Auswertung von PCRE: Beispiel 1 (6)

| Nr. | Knoten | Kanten nach | Pos. im Text |
|-----|--------|-------------|--------------|
| 4   | 5      | 6, 7        | a[]bba       |
| 5   | 7      | 9           | a[]bba       |
| 6   | 9      | 10          | ab[]ba       |

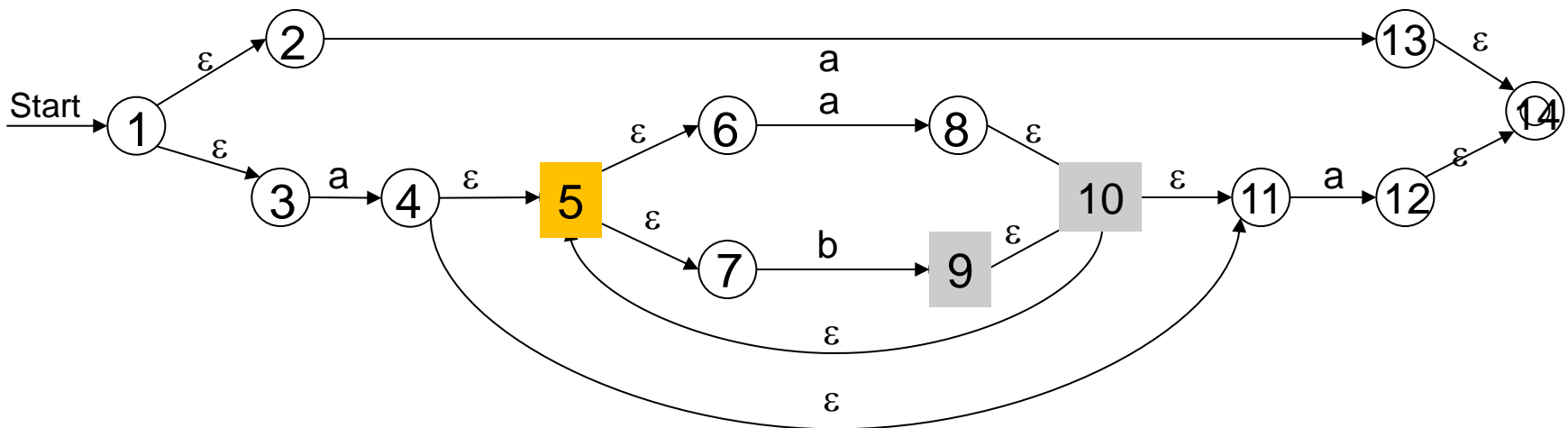
Von Knoten 5 aus wird jetzt der rechte Weg gewählt. Das b kann anschließend verarbeitet werden.



# Auswertung von PCRE: Beispiel 1 (7)

| Nr. | Knoten | Kanten nach | Pos. im Text |
|-----|--------|-------------|--------------|
| 6   | 9      | 10          | ab[]ba       |
| 7   | 10     | 5, 11       | ab[]ba       |
| 8   | 5      | 6, 7        | ab[]ba       |

Vom Knoten 10 aus gibt es wieder 2 Möglichkeiten: Zu Knoten 11 (Hüllenbildung verlassen) oder zu Knoten 5 (Hüllenbildung fortsetzen). Auch hier wird die Hüllenbildung zunächst fortgesetzt (greedy).





## Auswertung von PCRE: Beispiel 1 (8)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 8       | 5      | 6, 7        | ab[]ba       |
| 9       | 6      | 8           | ab[]ba       |

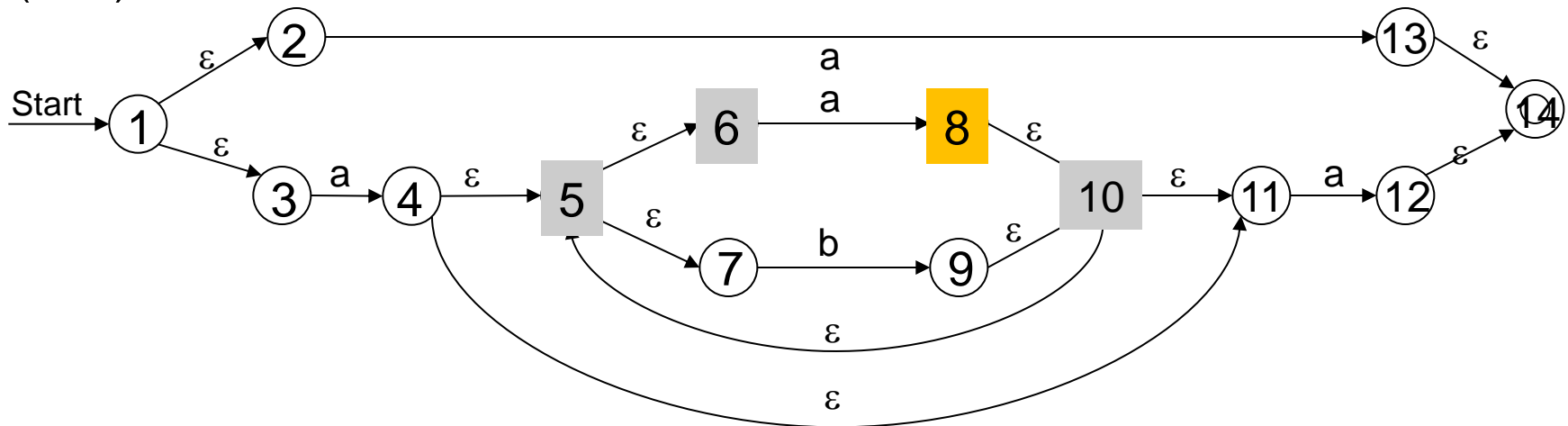
Da nächste b kann hier wieder nicht verarbeitet werden und man muss zu Schritt 8 zurück. Anschließend kann das b verarbeitet werden.

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 8       | 5      | 6, 7        | ab[]ba       |
| 9       | 7      | 9           | ab[]ba       |
| 10      | 9      | 10          | abb[]a       |
| 11      | 10     | 5, 11       | abb[]a       |

# Auswertung von PCRE: Beispiel 1 (9)

| Nr. | Knoten | Kanten nach | Pos. im Text |
|-----|--------|-------------|--------------|
| 11  | 10     | 5, 11       | abb[]a       |
| 12  | 5      | 6, 7        | abb[]a       |
| 13  | 6      | 8           | abb[]a       |
| 14  | 8      | 10          | abba[]       |

Die nächsten Entscheidungen sind wieder „greedy“ (-> 5) und „linkes oder“ (-> 6). Jetzt kann das a verarbeitet werden.



## Auswertung von PCRE: Beispiel 1 (10)

Jetzt wird versucht, über  $\varepsilon$ -Übergänge den Zielknoten zu erreichen. Das scheitert aber.

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 14      | 8      | 10          | abba[]       |
| 15      | 10     | 5, 11       | abba[]       |
| 16      | 5      | 6, 7        | abba[]       |
| 17      | 6      | 8           | abba[]       |
| Schritt | Knoten | Kanten nach | Pos. im Text |
| 16      | 5      | 6, 7        | abba[]       |
| 17      | 7      | 9           | abba[]       |
| Schritt | Knoten | Kanten nach | Pos. im Text |
| 15      | 10     | 5, 11       | abba[]       |
| 16      | 11     | 12          | abba[]       |

## Auswertung von PCRE: Beispiel 1 (11)

Man muss zunächst bis zu Schritt 12 und dann sogar bis zu Schritt 11 zurück. Der Cursor rückt im Text wieder eine Position zurück.

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 12      | 5      | 6, 7        | abb[]a       |
| 13      | 7      | 9           | abb[]a       |

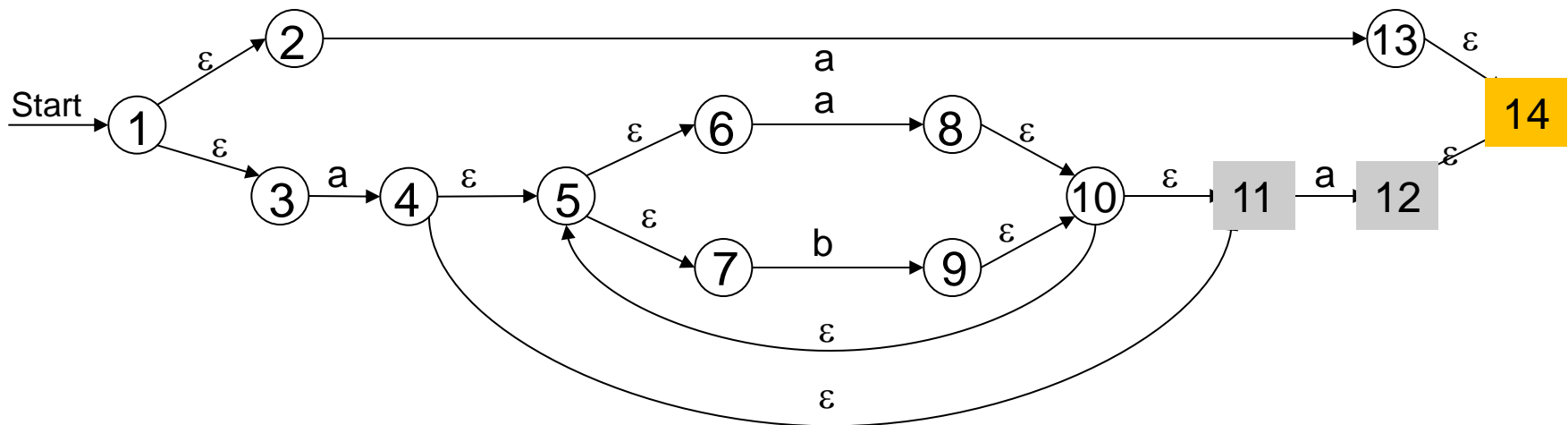
| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 11      | 10     | 5, 11       | abb[]a       |
| 12      | 11     | 12          | abb[]a       |

Die nun gewählte Möglichkeit bedeutet, dass die Hüllenbildung das letzte a nicht mehr „schluckt“.

# Auswertung von PCRE: Beispiel 1 (12)

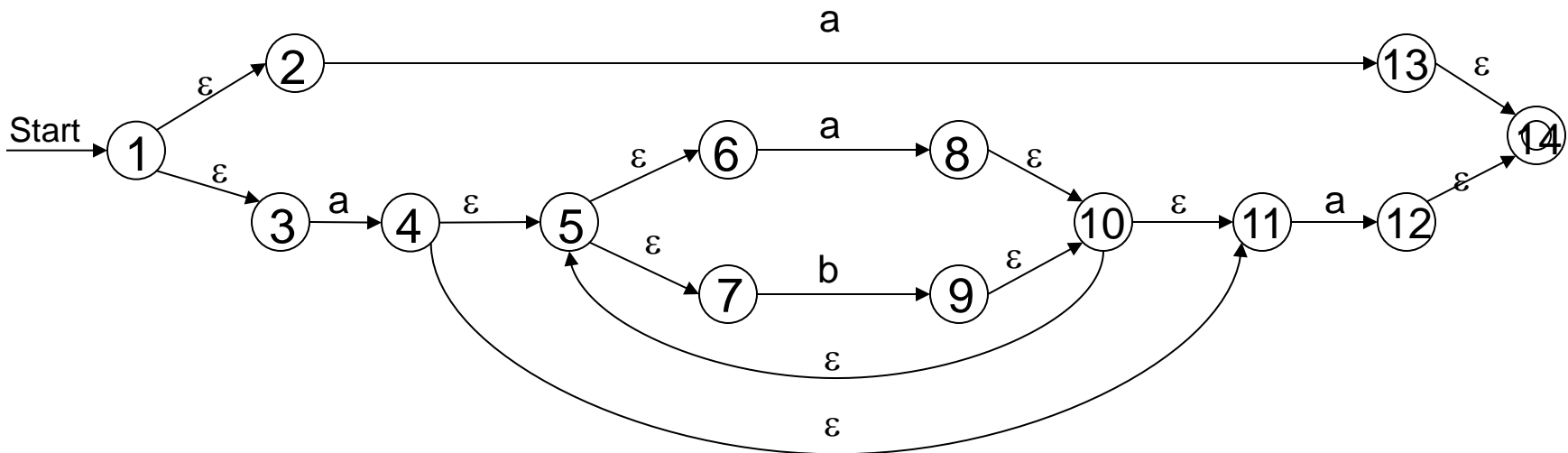
| Nr. | Knoten | Kanten nach | Pos. im Text |
|-----|--------|-------------|--------------|
| 12  | 11     | 12          | abb[]a       |
| 13  | 12     | 14          | abba[]       |
| 14  | 14     |             | abba[]       |

Der Algorithmus ist am Zielknoten angelangt und beendet sich. Das Ergebnis sind die verarbeiteten Buchstaben: abba



## Auswertung von PCRE: Beispiel 2 (1)

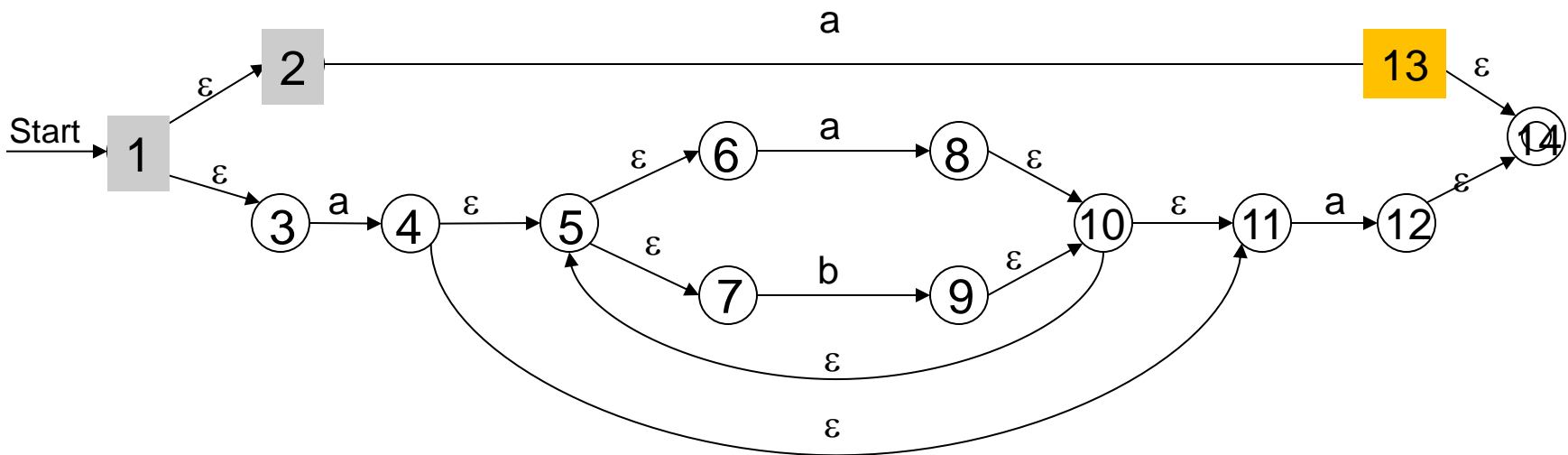
- 2. Beispiel: Pattern:  $a|a(a|b)^*a$ ; Text: abba
- Entspricht Beispiel 1, nur die Reihenfolge der äußeren oder-Funktion ist herumgedreht. Der Graph bleibt gleich.
  - Das Pattern für Beispiel 1 war  $a(a|b)^*a|a$



# Auswertung von PCRE: Beispiel 1 (2)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 1       | 1      | 2, 3        | []abba       |
| 2       | 2      | 13          | []abba       |
| 2       | 13     | 14          | a[]bba       |

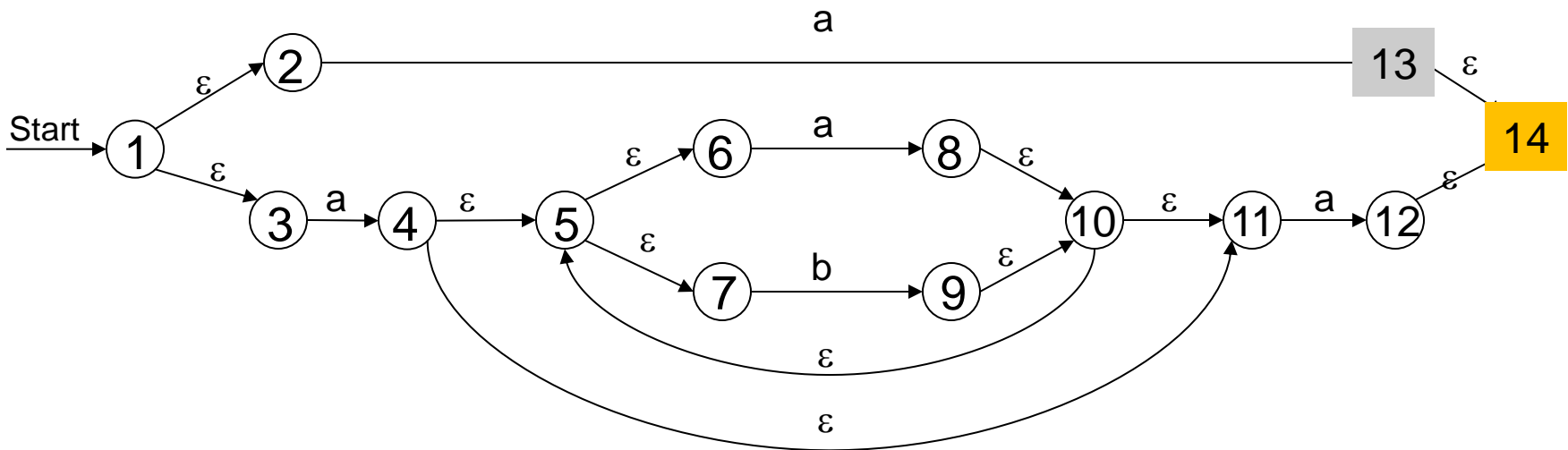
Die linke Seite des „oder“ führt nun zu Knoten 2. Das erste a kann bearbeitet werden.



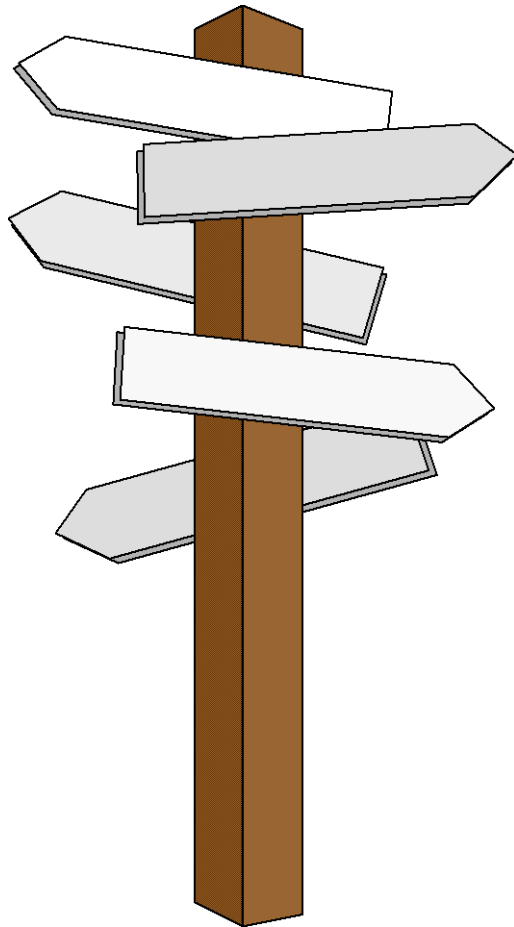
# Auswertung von PCRE: Beispiel 1 (3)

| Schritt | Knoten | Kanten nach | Pos. im Text |
|---------|--------|-------------|--------------|
| 2       | 13     | 14          | a[]bba       |
| 3       | 14     |             | a[]bba       |

Der nächste Schritt führt schon zum Zielknoten. Die verarbeiteten Zeichen (das einzelne a) werden als Treffer ausgegeben.







## Vorlesung 22

### Einfache Sortierverfahren

# Übersicht über die Sortierverfahren

- Sortierverfahren sind Bestandteil vieler Anwendungen
- Laut Statistik: ca. 25% der kommerziell verbrauchten Rechenzeit entfällt auf Sortiervorgänge.  
(Seite 63 in T. Ottmann, P. Widmayer: Algorithmen und Datenstrukturen. 1996, Spektrum, Akademischer Verlag, Heidelberg, Berlin)
- Sortierte Datensätze können
  - viel effizienter durchsucht werden (siehe Binäre Suche)
  - leichter auf Duplikate geprüft werden
  - von Menschen leichter gelesen werden
- Auch andere praxisrelevante Aufgaben können auf Sortierproblem zurückgeführt werden, z.B.
  - Median bestimmen
  - Bestimmung der  $k$  kleinsten Elemente

- **Große Anzahl von Sortierverfahren:**

Binary-Tree-Sort, Bogo-Sort, Bubble-Sort, Bucket-Sort, Comb-Sort, Counting-Sort, Gnome-Sort, Heap-Sort, Insertion-Sort, Intro-Sort, Merge-Sort, OET-Sort, Quick-Sort, Radix-Sort, Selection-Sort, Shaker-Sort, Shear-Sort, Shell-Sort, Simple-Sort, Slow-Sort, Smooth-Sort, Stooge-Sort

- Effizienz
- Speicherverbrauch
- Intern / Extern
- Stabil / Instabil
- allgemein / spezialisiert

- Das wichtigste Klassifikationskriterium ist die Effizienz.
- Einteilung in:
  - Schlechter als  $O(n^2)$ : **nicht ganz ernst gemeinte Verfahren**
  - $O(n^2)$ : **Elementare Sortierverfahren**
  - Zwischen  $O(n^2)$  und  $O(n \cdot \log n)$
  - $O(n \cdot \log n)$ : **Höhere Sortierverfahren**
  - Besser als  $O(n \cdot \log n)$ : **Spezialisierte Verfahren.**

- Von diesen Verfahren sollten Sie die Effizienz kennen:
- $O(n^2)$ : **Elementare (Einfache) Sortierverfahren**
  - **Bubble-Sort, Insertion-Sort, Selection-Sort**
- Zwischen  $O(n^2)$  und  $O(n \cdot \log n)$ 
  - **Shell-Sort**
- $O(n \cdot \log n)$ : **Höhere Sortierverfahren**
  - **Heap-Sort, Merge-Sort, Quick-Sort**
- Besser als  $O(n \cdot \log n)$ : **Spezialisierte Verfahren.**
  - **Radix-Sort**

- Spezialfälle sind:
  - Datenmenge sehr klein (<50)
  - Datenmenge sehr groß (passt nicht in Hauptspeicher)
  - Daten sind schon „vorsortiert“ (nur wenige Elemente sind nicht am richtigen Platz).
  - Daten stehen nicht in einem Feld sondern in einer verketteten Liste.



- Speicherverbrauch
- Rein vergleichsbasiertes Verfahren (sehr universell einsetzbar) oder spezialisiertes Verfahren (muss an jeweiliges Problem angepasst werden):
- Stabil / Instabil  
(behalten Datensätze mit gleichen Schlüsseln relative Reihenfolge?)  
Relevant bei mehrmaligem Sortieren nach verschiedenen Schlüsseln

- Namen nach Vor- und Nachname sortieren:  
„Meier, Martin“, „Meier, Ulla“, „Schmitz, Heinz“, „Ernst, Eva“
- 1. Sortieren zuerst nach Vornamen:  
„Ernst, **E**va“, „Schmitz, **H**einz“, „Meier, **M**artin“, „Meier, **U**lla“
- 2. Dann Sortieren nach Nachnamen:
  - Stabil: „**E**rnst, Eva“, „**M**eier, **M**artin“, „**M**eier, **U**lla“, „**S**chmitz, Heinz“
  - Instabil, z.B. „**E**rnst, Eva“, „**M**eier, **U**lla“, „**M**eier, **M**artin“, „**S**chmitz, Heinz“
- Das wichtigste stabile Verfahren ist Merge-Sort. Außerdem sind Radix-Sort, Insertion-Sort und Bubble-Sort stabil.

- Normalerweise Quicksort.
- Merge-Sort, falls
  - die Datenmenge zu groß für den Hauptspeicher ist.
  - die Daten als verkettete Liste vorliegen.
  - ein stabiles Verfahren nötig ist.
- Insertion-Sort, falls
  - wenige Elemente zu sortieren sind.
  - die Daten schon vorsortiert sind.
- Radix-Sort, falls
  - sich ein hoher Programmieraufwand für ein sehr schnelles Verfahren lohnt.

```
static void swap(int array[], int index1, int index2) {
 int temp;
 temp = array[index1];
 array[index1] = array[index2];
 array[index2] = temp;
}
```

Für Klassen, die das Interface `List` implementieren (`ArrayList`, ...) geht auch:

```
Collections.swap(List<?> list, int i, int j)
```

## Nicht ganz ernstzunehmende Verfahren

## $O(n \cdot n!)$ : Bogo-Sort

- „The archetypal perversly awful algorithm“ (Wikipedia).
- Würfelt solange alle Elemente durcheinander, bis das Feld (zufällig) sortiert ist.
- Zitat aus dem Internet:
  - **Looking at a program and seeing a dumb algorithm, one might say „Oh, I see, this program uses bogo-sort.“**

```
public void sort(List a) {
 while (isSorted(a)==false) {
 Collections.shuffle(a);
 }
}
```

```
public boolean isSorted(List a) {
 for (int i=1; i<a.size(); i++) {
 if isLastSmaller(a.get(i-1),
 a.get(i)) {
 return false;
 }
 }
 return true;
}
```

- Vorgehensweise wird hier nicht erklärt
  - -> **Wikipedia**
- Versucht, die Arbeit soweit wie möglich zu vervielfachen (multiply and surrender).
- Das Ergebnis wird erst dann berechnet, wenn die Lösung nicht weiter hinausgezögert werden kann.
- Ziel: Auch im besten Fall ineffizienter als alle anderen Sortierverfahren.
- Ist aber trotzdem ein echtes Sortierverfahren.

## $O(n^3)$ : Stupid-Sort

- Sehr kurzer Code.
- Variante von Bubble-Sort.
- Kann auch rekursiv implementiert werden („*with recursion, stupid-sort can be made even more stupid*“ (Wikipedia)).

```
public void sort(int[] a) {
 for (int i=0; i<a.length-1; i++) {
 if (a[i]>a[i+1]) {
 swap(a, i, i+1);
 i=-1;
 }
 }
}
```



## **$O(n^2)$ : Einfache Sortierverfahren**

- Die bekannten einfachen Sortierverfahren sind:

**Bubble-Sort**

**Selection-Sort**

**Insertion-Sort**

- Das beste Verfahren der 3 ist **Insertion-Sort**. Es wird ausführlich vorgestellt.
- Selection-Sort wird kurz angesprochen.
- Bubble-Sort wird nicht weiter erklärt.
- Dafür wird ein besonders einfacher Algorithmus namens **Simple-Sort** behandelt.

- Simple-Sort
  - $O(n^2)$ , mit hohem Vorfaktor.
  - Leicht zu merken.
- Wenden Sie es an:
  - wenn Sie keine Zeit oder Lust zum Nachdenken haben.
  - wenn die Felder so klein sind, dass der Algorithmus nicht effektiv sein muss.
  - wenn niemand sonst Ihren Code zu sehen bekommt.
- Selection-Sort
  - Eines der wichtigeren elementaren Verfahren.
  - Wird ein elementares Verfahren benötigt, wird aber meistens das etwas schnellere InsertionSort verwendet.

- Das Grundprinzip ist für beide Sortierverfahren gleich
- SimpleSort ergibt einen besonders einfachen Code, ist aber langsamer.
- Grundprinzip:

```
for (int i=0; i<array.length-1; i++)
```

Suche das kleinste Element zwischen i und dem rechten Feldende.

Vertausche dieses Element mit dem Element i.

- Das Suchen und Vertauschen wird bei SimpleSort auf ganz spezielle Weise gemacht:
  - Gehe vom **i. Element** aus nach rechts.
  - Jedes Mal, wenn ein kleineres Element als das auf Position **i** auftaucht, dann vertausche es mit dem **i. Element**.

```
public void simpleSort (int[] a) {

 for (int i=0; i<a.length; i++) {
 for (int j=i+1; j<a.length; j++) {
 if (a[i]>a[j]) {
 swap (a,i,j);
 }
 }
 }
}
```

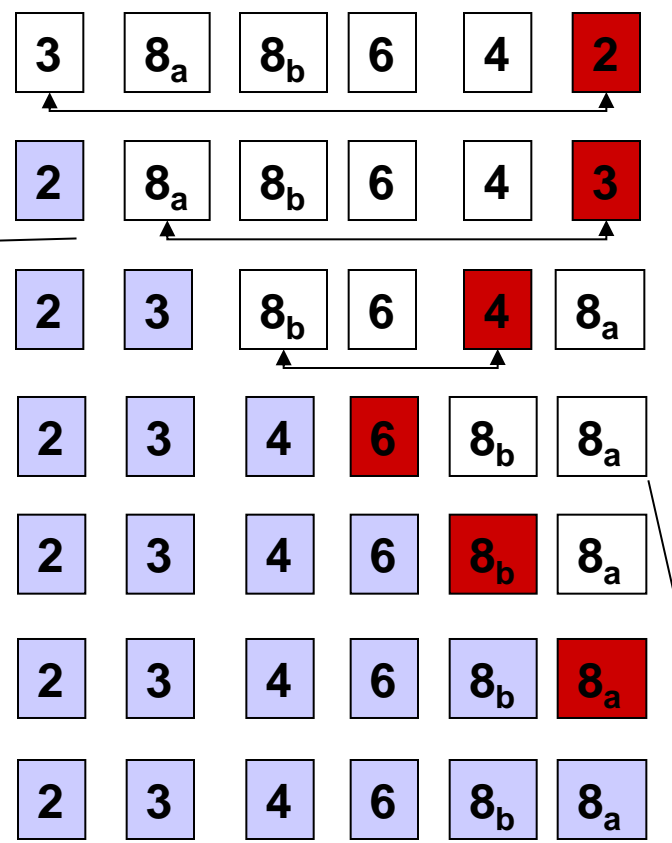
- „Normale“ Vorgehensweise: Erst kleinstes Element suchen, dann mit Element  $i$  vertauschen.

```
public void selectionSort (int[] a) {

 for (int i=0; i<a.length; i++) {
 int small = i;
 for (int j=i+1; j<a.length; j++) {
 if (a[small]>a[j]) {
 small = j;
 }
 }
 swap(a, i, small);
 }
}
```

# Beispiel zu Selection-Sort

SelectionSort ist instabil:



Sortierte Teilfolge  
 Kleinstes Element

Keine Vorteile, wenn Feld schon sortiert.

- In den meisten Fällen der schnellste elementare Suchalgorithmus.

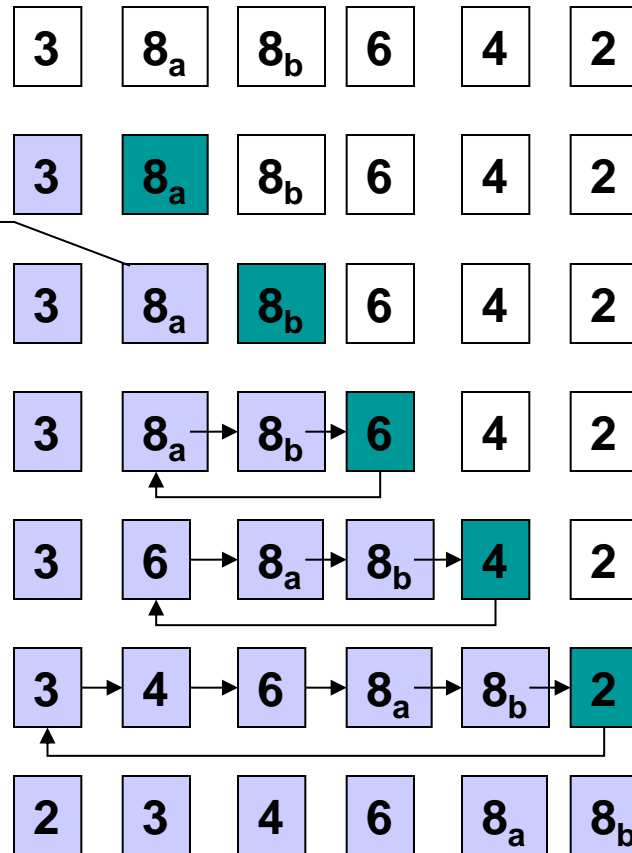
**Grundidee** (am Beispiel der Sortierung eines Kartenstapels)

- 1. Starte mit der ersten Karte einen neuen Stapel.**
- 2. Nimm jeweils die nächste Karte des Originalstapels und füge diesen an der richtigen Stelle in den neuen Stapel ein.**



# Beispiel zu InsertionSort

InsertionSort  
ist stabil:  
Vertauschung  
nur bei  
Ungleichheit



 **Sortierte Teilfolge**  
 **Einzusortierendes Element**

# InsertionSort in Java

```
public static void insertionSort(int[] array) {

 for (int i=1; i < array.length; i++) {
 int m = array[i];

 // fuer alle Elemente links von aktuellem Element
 int j;
 for (j=i; j>0; j--) {
 if (array[j-1]<=m) {
 break;
 }
 // größere Elemente nach hinten schieben
 array[j] = array[j-1];
 }

 // m an freiem Platz einfügen
 array[j] = m;
 }
}
```

- Wir zählen die Anzahl der **Vergleiche** (Anzahl der „**Bewegungen**“ ist ungefähr gleich).
- **$n-1$  Durchläufe**: Im Durchlauf mit Nummer  $k$  ( $k \in [2;n]$ ):
  - höchstens  $k-1$  Vergleiche.
  - mindestens 1 Vergleich (und keine Bewegung).
- **Best Case** - vollständig **sortierte** Folge:  $n-1$  Vergleiche, keine Bewegungen  $\rightarrow O(n)$ .
- **Average Case** - jedes Element wandert **etwa in Mitte des sortierten Teils**:

$$\sum_{k=2}^n k/2 = \frac{1}{2} \left( \frac{n(n+1)}{2} - 1 \right) \approx \frac{n^2}{4} \text{ Vergleiche } \rightarrow O(n^2).$$

## SimpleSort:

- Einfach zu implementieren.
- Langsam.

## SelectionSort:

- Aufwand ist unabhängig von der Eingangsverteilung (Vorsortierung).
- Es werden nie mehr als  $O(n)$  Vertauschungen benötigt.

## BubbleSort:

- Stabil
- Vorsortierung wird ausgenutzt.
- Langsam

## InsertionSort:

- Stabil
- Vorsortierung wird ausgenutzt
- Für ein elementares Suchverfahren ( $O(n^2)$ ) schnell.

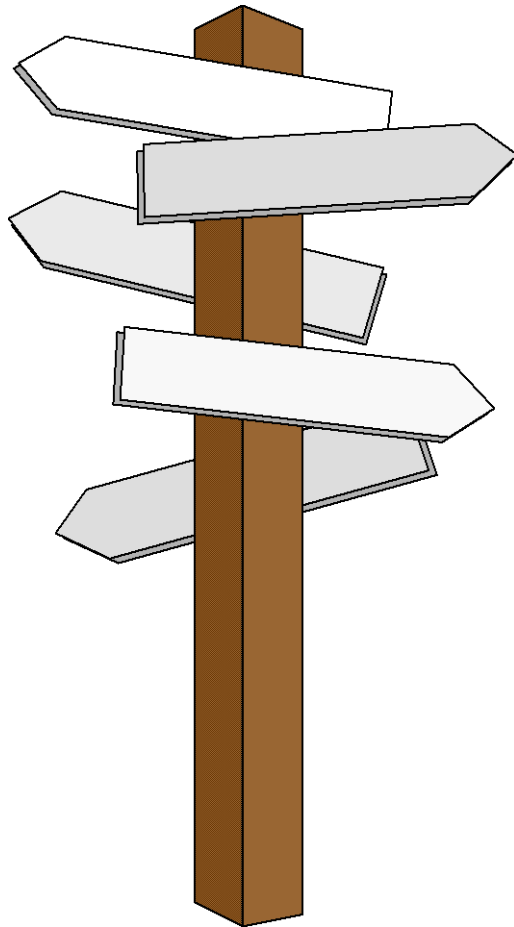
## Analyse einfacher Sortierverfahren

| Verfahren     | Laufzeitmessungen<br>(nach Wirth, Sedgewick) | Vorsortierung<br>ausnutzen | Stabil |
|---------------|----------------------------------------------|----------------------------|--------|
| SimpleSort    | 330*                                         |                            |        |
| SelectionSort | 120-200                                      |                            |        |
| InsertionSort | 100                                          | X                          | X      |
| BubbleSort    | 250-400                                      | X                          | X      |

\* Eigene Messung

### InsertionSort:

- Stabil
- Vorsortierung wird ausgenutzt
- Für ein elementares Suchverfahren ( $O(n^2)$ ) schnell.



## Vorlesung 23

### Quick-Sort

- Die bekannten höheren Sortierverfahren sind:

**Quick-Sort**

**Merge-Sort**

**Heap-Sort**

- Das meist beste Verfahren der 3 ist **Quick-Sort**. Es wird ausführlich vorgestellt.
- Ebenso **Merge-Sort**, der in einigen Fällen Vorteile hat.
- Heap-Sort wird kurz angesprochen.
- Gute Beschreibung aller 3 Verfahren in:
  - [http://www.linux-related.de/coding/sort/sort\\_main.htm](http://www.linux-related.de/coding/sort/sort_main.htm)

- Die überwiegende Mehrheit der Programmbibliotheken benutzt Quick-Sort.
  - z.B. **Sun-Java, Gnu-Java, C#/.NET (nach Wikipedia), C#/Mono, Ruby, NAG, ...**
- In fast allen Fällen sind zwei Optimierungen eingebaut: „Median of three“ und „Behandlung kleiner Teilfelder“.
- Gnu-C++ benutzt Intro-Sort (Quick-Sort-Variante).
- Bei Objekten kann Stabilität wichtig sein. Hier verwendet Java (wie Python) Merge-Sort.



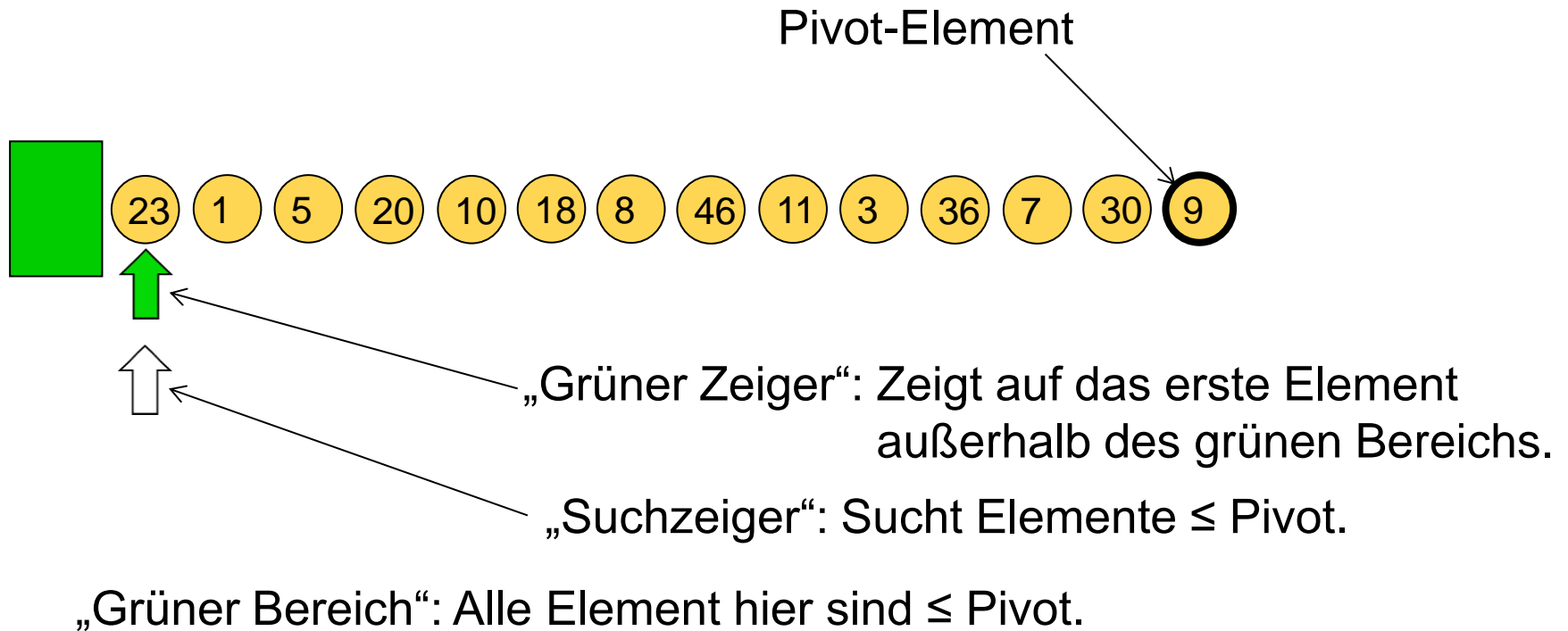
**Prinzip:** *divide-and-conquer* (,teile‘ und ,herrsche‘)

## Rekursiver Algorithmus

- Müssen 0 oder 1 Elemente sortiert werden  $\Rightarrow$  Rekursionsabbruch
- Wähle ein Element als „**Pivot-Element**“ aus.
- Teile das Feld in 2 Teile:
  - **Ein Teil mit den Elementen größer als das Pivot.**
  - **Ein Teil mit den Elementen kleiner als das Pivot.**
- Wende den Algorithmus rekursiv für beide Teilfelder an.

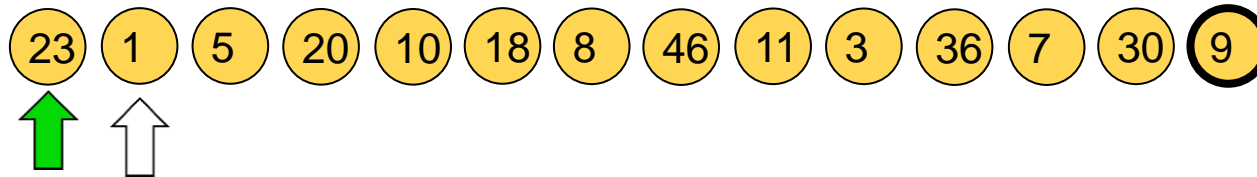
## Quicksort: Beispiel (1)

- Start

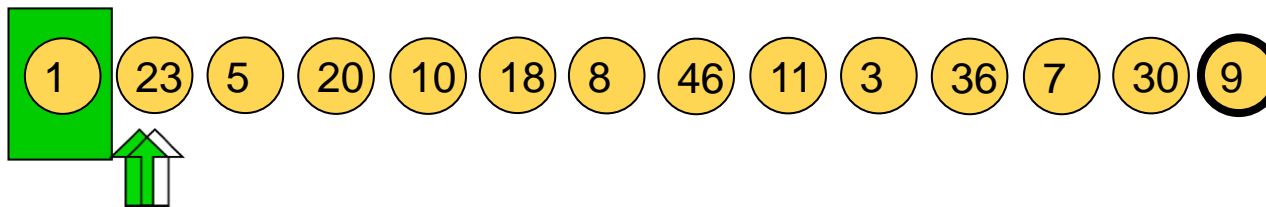


- Erster Schritt

Der Suchzeiger rückt nach rechts und sucht ein Element  $\leq$  Pivot.

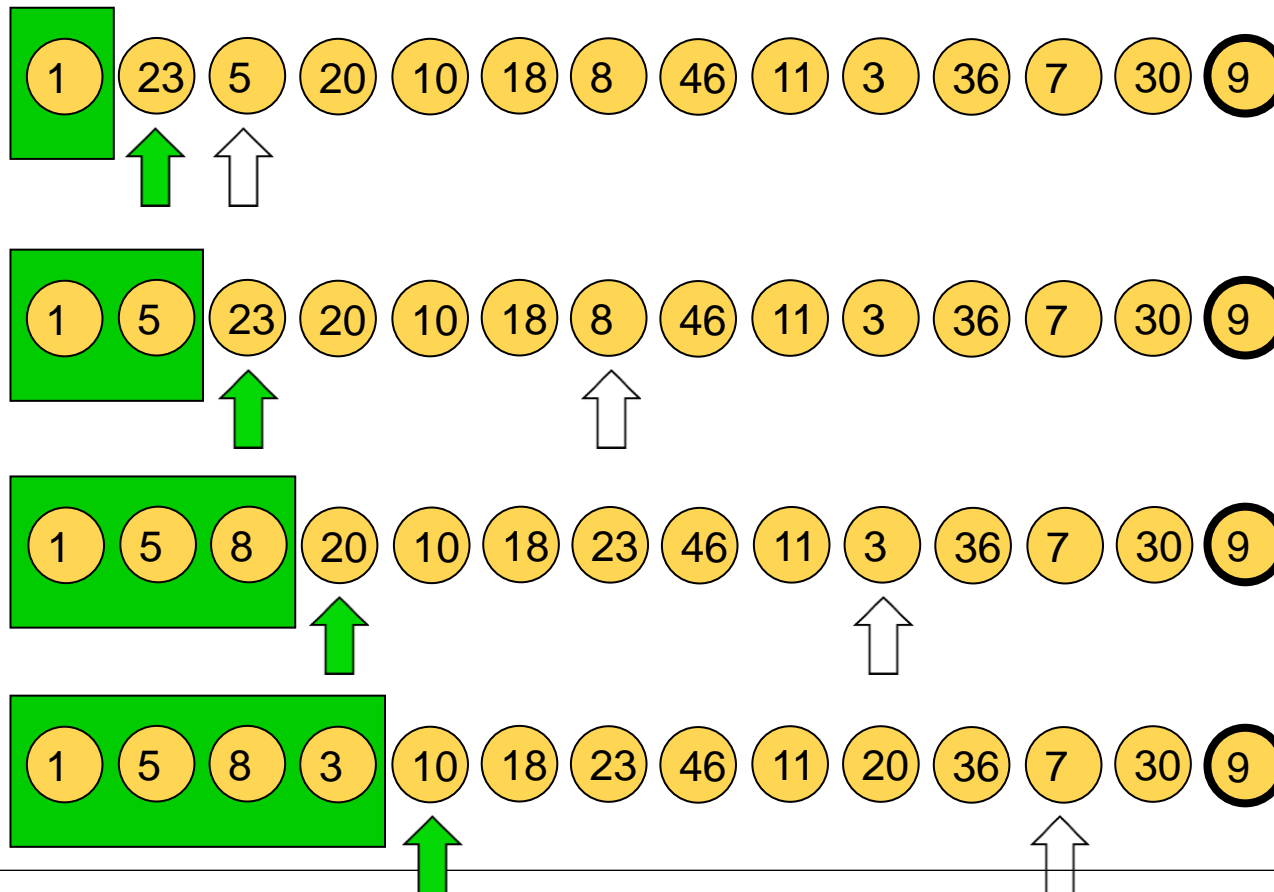


Die Elemente unter dem grünen und dem Suchzeiger werden getauscht. Der grüne Bereich wird um 1 Element nach rechts erweitert.

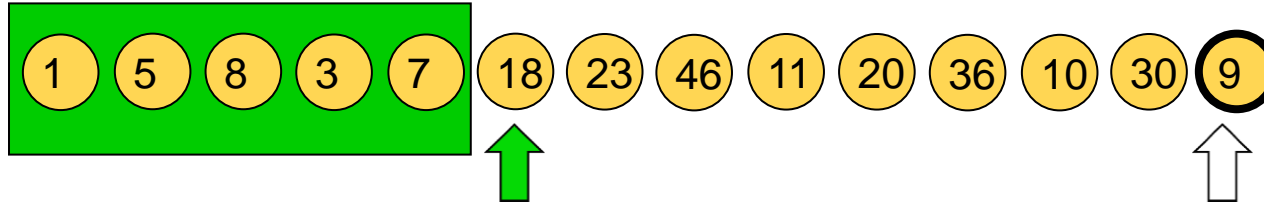


## Quicksort: Beispiel (3)

Dieser Schritt wird solange wiederholt, bis der Suchzeiger am rechten Rand des Feldes angekommen ist.

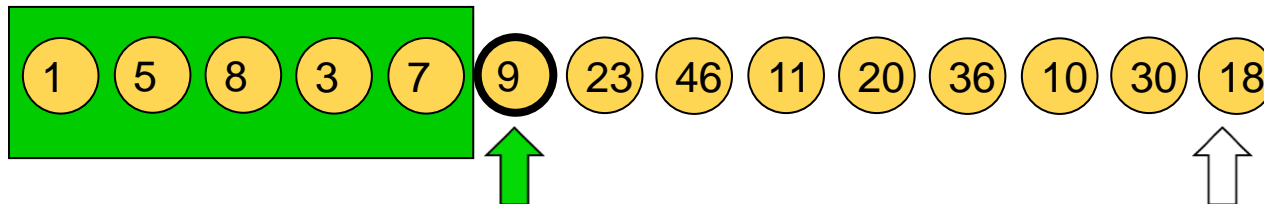


## Quicksort: Beispiel (4)

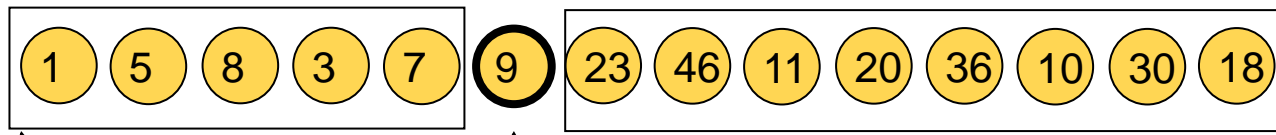


Jetzt sind alle Elemente  $\leq$  Pivot (außer dem Pivot-Element selbst) im grünen Bereich.

Zum Abschluss wird das Pivot-Element mit dem ersten Element hinter dem grünen Bereich vertauscht.



Es gibt jetzt 3 Bereiche:



Bereich 2: Das Pivot-Element  
befindet sich bereits an der  
richtigen Stelle.

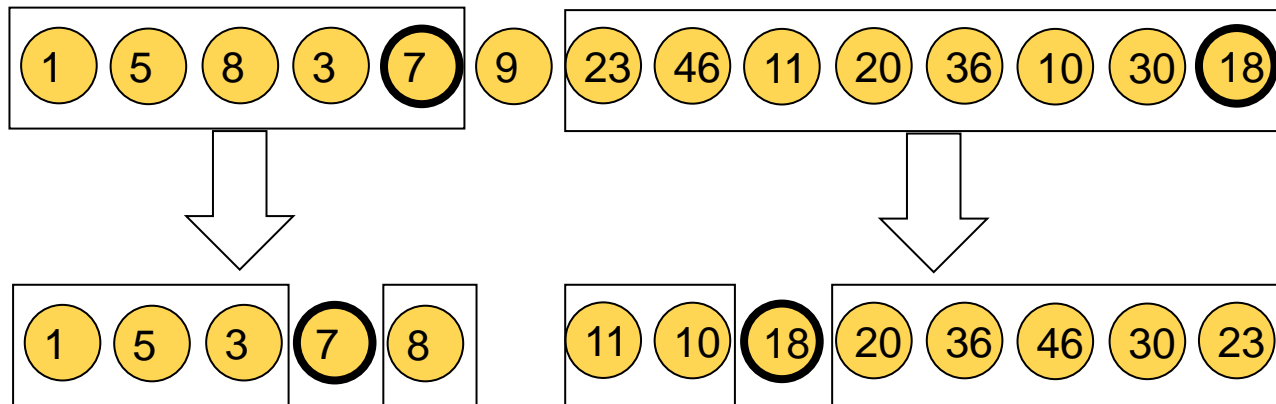
Bereich 1 (vor dem Pivot-Element):  
Alle Wert sind  $\leq$  Pivot.

Bereich 3 (hinter dem Pivot-Element):  
Alle Wert sind  $>$  Pivot.

Die Bereiche 1 und 3 werden rekursiv mit Quicksort sortiert.

## Quicksort: Beispiel (5)

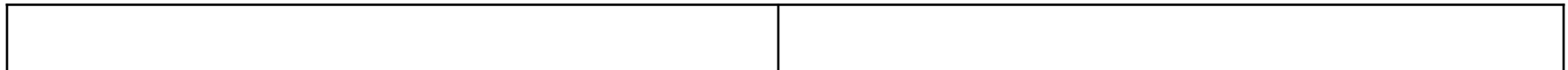
Die Bereiche 1 und 3 werden rekursiv mit Quicksort sortiert.



Die Rekursionen setzen sich solange fort, bis die Teilbereiche nur noch 1 Element enthalten (siehe auch Folie „Optimierung: Rekursionsabbruch“).

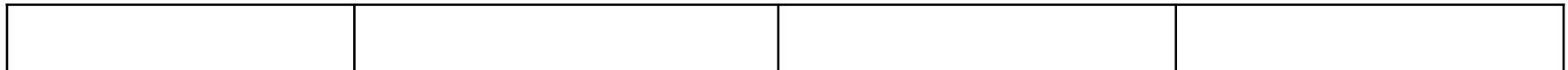
*Pivot-Wert ist immer Median der Teilliste  $\Rightarrow$  Teillisten werden stets halbiert.*

Stufe 1



1 n  
n Elemente werden mit dem Pivotwert verglichen und maximal  $n/2$  Paare vertauscht.

Stufe 2



1  $n/2$  n

In jeder der beiden Teillisten werden  $n/2$  Elemente (also insgesamt  $n$ ) mit dem jeweiligen Pivotwert verglichen und maximal  $n/4$  Paare vertauscht.

Allgemein: In jeder Stufe werden  $n$  Elemente betrachtet. Abbruch bei Teillisten der Länge  $n \leq 1 \Rightarrow \lfloor \lg n \rfloor$  Stufen (mit Halbierung der Teillisten).

Also:  $T_{\text{quicksort}}^b(n) \in O(n \log n)$



*Als Pivot-Wert wird stets das größte oder kleinste Element der Teilliste ausgewählt.*

Dann gilt:

- Die Länge der längsten Teilliste ist  $(n-1)$  bei Stufe 1,  
 $(n-2)$  bei Stufe 2, etc.  
Allgemein :  $(n-i)$  bei Stufe  $i$ .
- Es sind  $(n-1)$  Stufen nötig.  
In jeder Stufe  $i$  werden  $n-i$  Elemente betrachtet.
- Also:  $T^w_{\text{quicksort}}(n) \in \mathbf{O(n^2)}$

*Aufwand im Mittel:*

Genauere Analyse ist aufwändig. Resultat:  $T_{\text{quicksort}}^{\text{av}}(n) \in O(n \log n)$

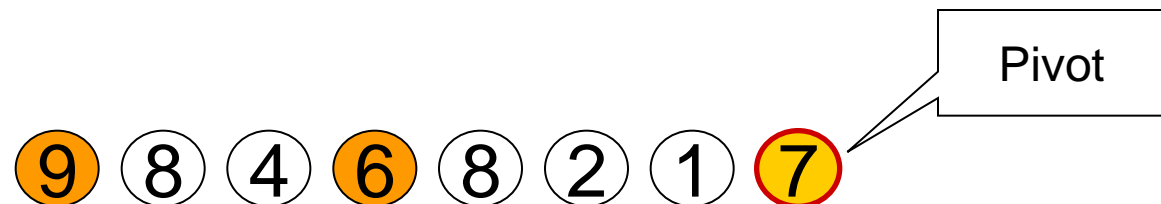
Genauer:

#Vergleiche im mittleren Fall ist nur um etwa 39% größer als im Besten Fall. QuickSort ist also auch im Mittleren Fall eine sehr gute Wahl.

Hauptziel ist Vermeidung des schlimmsten Falls.

- Wenn die Pivot-Elemente unglücklich gewählt sind, erhält man das wesentlich schlechtere Zeitverhalten  $O(n^2)$ .
- Dazu Anmerkung:  
Wenn man beim Zerlegen das erste oder letzte Feldelement als Pivot wählt, dann liegt der Worst Case bei sortierten Feldern vor.
  - **Auch nahezu sortierte Felder haben schon  $O(n^2)$ .**
- Pivot geschickter wählen.
  - **Einfachste Möglichkeit: Mittleres Element als Pivot wählen. Dann liegt der Best case bei sortierten Felder vor.**
  - **Man will den Quick-Sort aber noch sicherer machen.**

- *Median-of-three-Methode* zum Auswählen des Pivots:
  - Es werden drei Elemente als Referenz-Elemente, z.B. vom Listenanfang, vom Listende und aus der Mitte gewählt. Das Element mit dem mittleren Schlüsselwert wird als Pivot-Element gewählt.
  - Kann auf mehr als drei Elemente ausgebaut werden.
  - Das Pivot-Element wird vor dem Quicksort-Durchgang mit dem letzten Element getauscht.



- Einfache Lösung: Rekursionsabbruch, wenn die Teilliste 0 oder 1 Element enthält.
- Aber: Die letzten Rekursionsdurchgänge sind nicht mehr effektiv.
- Daher wird die Rekursion schon früher abgebrochen.

- **Kleine Teilliste mit InsertionSort sortieren:**

- Die Grenze für eine kleine Teilliste ist nicht klar festgelegt.
  - Die Standardwerke (Knuth, Sedgewick) empfehlen 9.
  - Im Internet findet man aber auch andere Werte zwischen 3 und 32.

| Sprache    | Bibl.-Aufruf                                   | Rekursionsabbruch bei max. n Elementen |
|------------|------------------------------------------------|----------------------------------------|
| Java (Sun) | <code>java.util.Arrays.sort</code>             | 6                                      |
| Java (Gnu) | <code>java.util.Arrays.sort</code>             | 7                                      |
| C# / Mono  | <code>System.Collections.ArrayList.sort</code> | 3                                      |
| Ruby       | <code>Array.sort</code>                        | 2                                      |

## Quicksort:

- Schnellster der oft verwendeten allgemeinen Algorithmen.
- Programmbibliotheken implementieren ganz überwiegend QuickSort.
- Nutzt Vorsortierung nicht aus.
- Im Worst Case nur  $O(n^2)$ .
- Schlecht für kleine  $n$  ( $<20$ ).
- Zwei Standard-Optimierungen

## Standard-Optimierung: Rekursionsabbruch für kleine $n$ :

- Verwendung von InsertionSort
- Grenze liegt zwischen 3 und 32.

## Standard-Optimierung: Pivot ist Median mehrerer Elemente:

- 3 oder mehr Pivot-Elemente.
- Dadurch Worst Case sehr unwahrscheinlich.
- Langsamer aufgrund zusätzlicher Rechenschritte.

- Optimierte Quicksort-Variante
  - Z.B. in **Gnu C++**
- Ab einer gewissen Rekursionstiefe wird zu **Heap-Sort** gewechselt.
  - **Heap-Sort hat  $O(n \log n)$  auch im schlechtesten Fall. Damit kann  $O(n^2)$  nicht auftreten.**
  - **Allgemein ist Heap-Sort aber langsamer als Quick-Sort.**
  - **Große Rekursionstiefe deutet bei Quick-Sort auf Worst-Case-Probleme. Daher wechselt man hier zu Heap-Sort.**

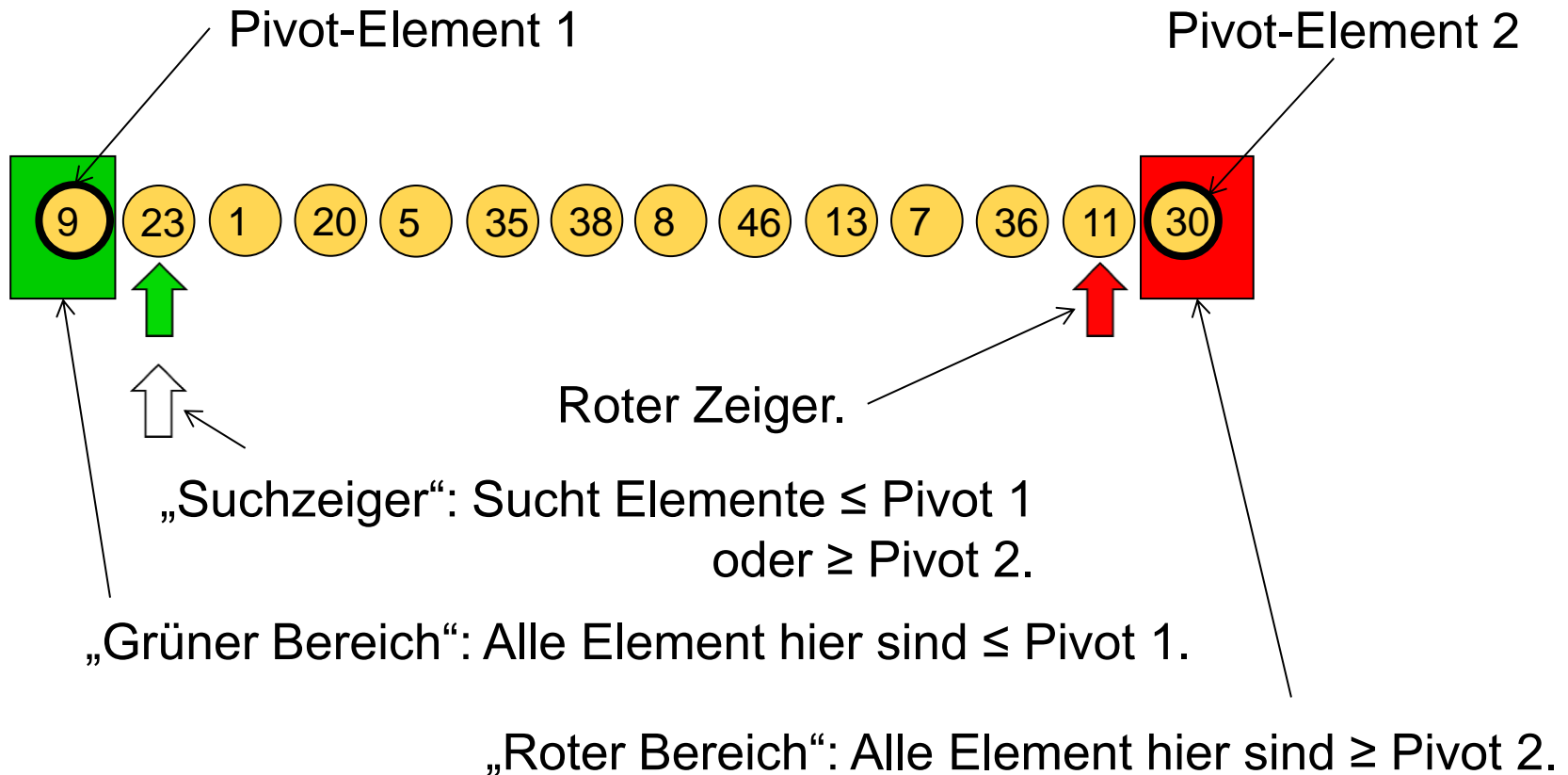


- Yaroslavskiy, Bentley, Bloch (2009)
- Java benutzt Double-Pivot Quicksort seit **Java 7**
  
- 2 Pivot-Elemente
- Werte werden in 3 Bereiche geteilt

- Es gibt Ähnlichkeiten zur vorgestellten Quick-Sort-Variante.
- Zunächst werden zwei Pivot-Elemente ausgewählt.
- Das größere Element wird ganz nach rechts kopiert.
- Das kleinere Element wird ganz nach links kopiert.

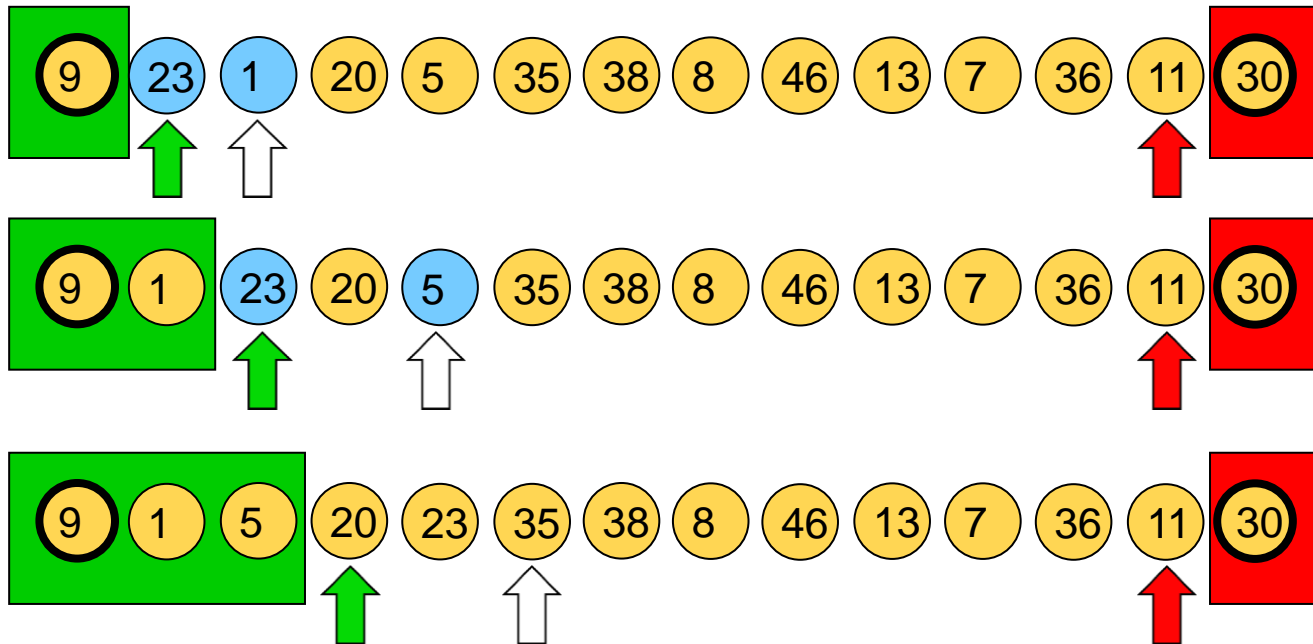
# Dual-Pivot-Quicksort: Beispiel (1)

- Start



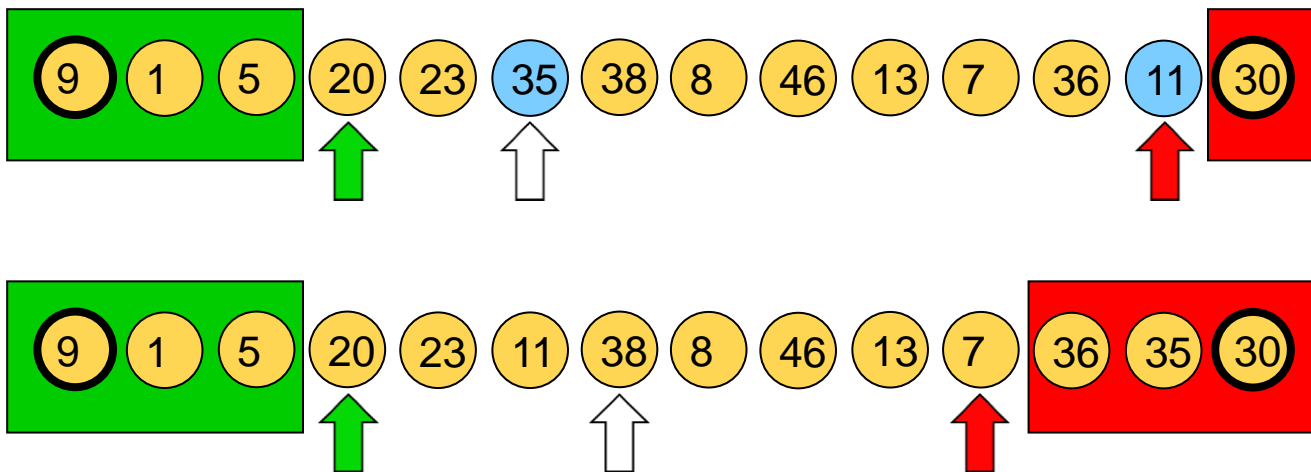
## Dual-Pivot-Quicksort: Beispiel (2)

- Trifft der Suchzeiger auf Elemente  $\leq$  Pivot 1 ist das Verfahren genau so wie bei der vorgestellten „normalen“ Quick-Sort-Variante.



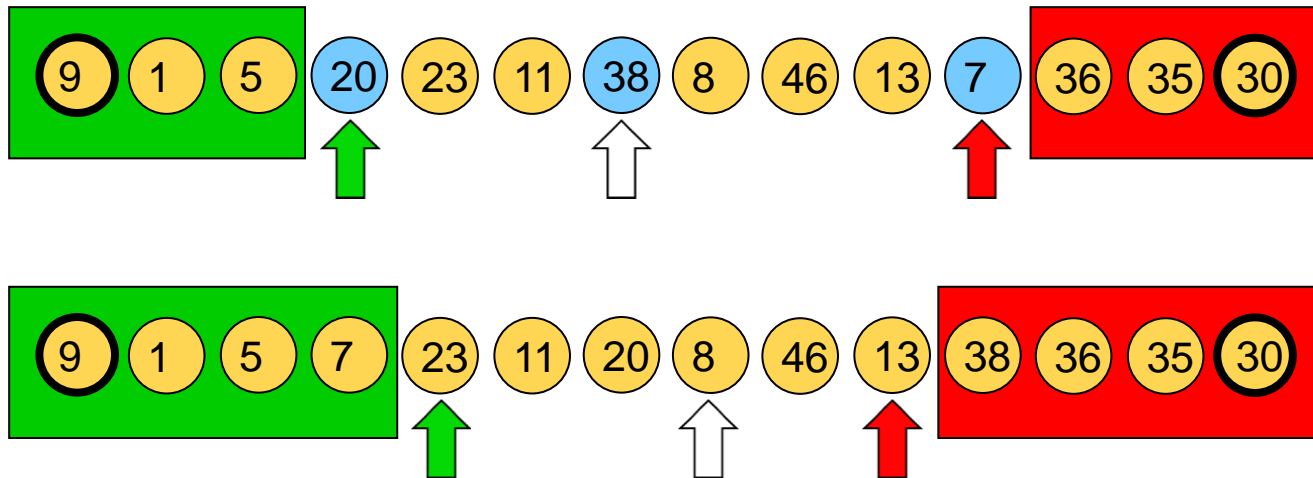
## Dual-Pivot-Quicksort: Beispiel (3)

- Trifft der Suchzeiger auf Elemente  $\geq \text{Pivot2}$ , dann vertauschen die Elemente unter dem Suchzeiger und dem roten Zeiger. Der rote Zeiger rückt zunächst um 1 Element vor.
- Besonderheit: Solange der rote Zeiger jetzt auf Elementen  $\geq \text{Pivot2}$  steht, rückt er weiter nach links vor.
- Das zurückgetauschte Element (11) gehört nicht in den grünen Bereich und kann stehenbleiben.



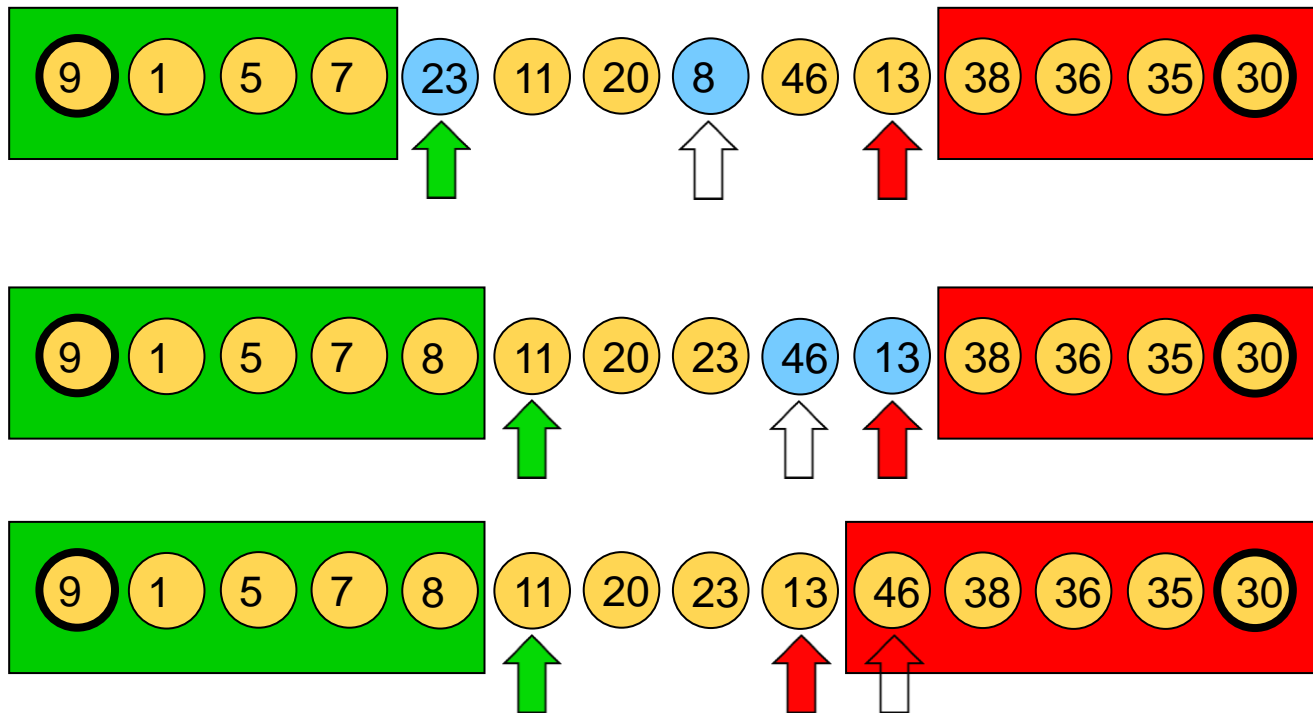
## Dual-Pivot-Quicksort: Beispiel (4)

- Die 38 vertauscht mit der 7. Da die 7 in den grünen Bereich gehört, vertauscht sie anschließend noch mit der 20.
- Sowohl der rote als auch der grüne Zeiger rücken danach eine Position nach vorne.

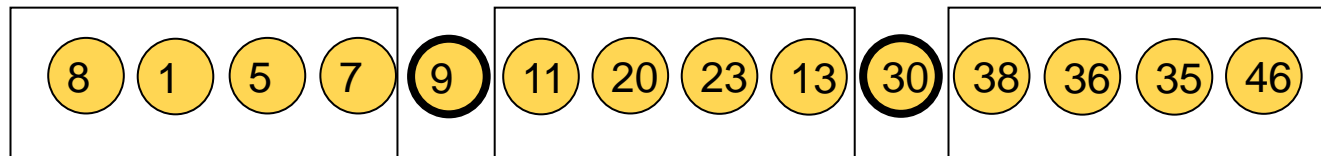


## Dual-Pivot-Quicksort: Beispiel (5)

- Sobald der Suchzeiger in den roten Bereich wandert, kann man die Vertauschungen beenden.



- Man kopiert die beiden Pivots an den inneren Rand ihres Bereichs. Sie haben dort bereits ihren endgültigen Platz. Die restlichen drei Bereiche sortiert man rekursiv mit Quicksort weiter.





## Kurze Wiederholung: Heap-Sort

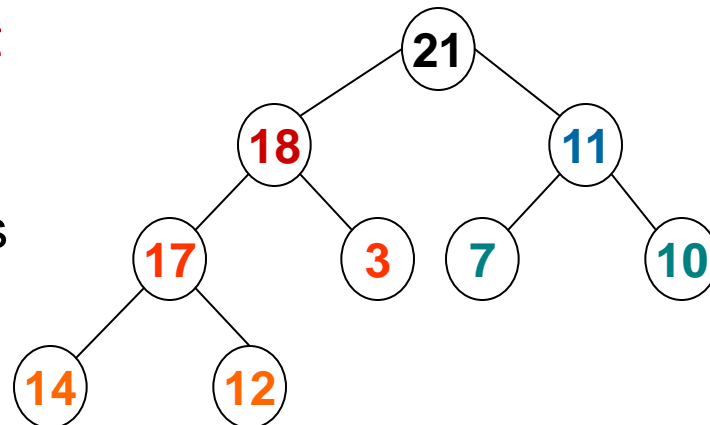
## Definition:

Ein *Heap* ist ein Binärbaum mit folgenden Eigenschaften:

- Er ist links-vollständig
- Die Kinder eines Knotens sind höchstens so groß wie der Knoten selbst.

⇒ das größte Element befindet sich an der Wurzel des Heaps

Achtung: In der Literatur gibt es auch die umgekehrte Definition



- Zunächst wird das Feld komplett in einen Heap umgewandelt:



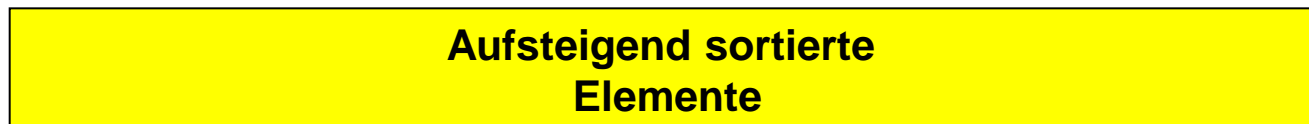
- Wenn die Wurzel entfernt wird, schrumpft der Heap um 1 Element und das letzte Feldelement gehört nicht mehr zum Heap.
- Hier kann das 1. sortierte Element untergebracht werden



- Wenn weitere Male die Wurzel entfernt wird, schrumpft der Heap immer weiter. Die freiwerdenden Stellen werden mit den entnommenen Werten besetzt, die jetzt aufsteigend sortiert sind.



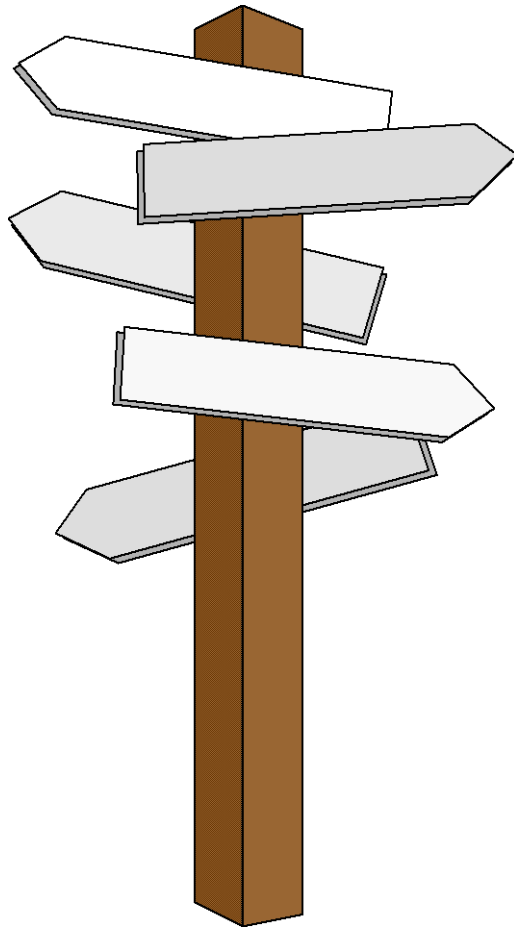
- Zuletzt ist der Heap auf 0 geschrumpft und das Feld ist sortiert.



1. **Aufbau des Heap (heapCreate):**  
Für  $n/2$  Elemente wird sink ausgeführt.  
Maximal werden dabei  $\log(n+1)$  Vergleiche durchgeführt.  
 $\Rightarrow T_{\text{heapCreate}}^w(n) \leq n/2 \log(n+1) \in O(n \log n)$
  2. Beim **sortierten Entnehmen** der Elemente mit Wiederherstellen des Heap wird sink für  $n-1$  Elemente ausgeführt.  
 $\Rightarrow T_{\text{scndPhase}}(n) \leq (n-1) \log(n+1)$
- $\Rightarrow$  Für HeapSort insgesamt gilt dann:  
 $T_{\text{heapSort}}^w(n) \leq 3/2 \cdot n \cdot \log(n+1) \in O(n \log n)$

## HeapSort:

- $O(n \log n)$ , auch im worst case
  - **Hauptvorteil gegenüber Quick-Sort**
  - **Wird daher in Quick-Sort-Optimierung „Intro-Sort“ verwendet.**
- Kein zusätzlicher Speicher nötig.
- Nicht stabil.
- Vorsortierung wird nicht ausgenutzt.
- Im Normalfall langsamer als QuickSort.



## Vorlesung 24

### Merge-Sort und Radix-Sort

## Internes / Externes Sortieren

- Voraussetzung für bisher behandelte Verfahren:  
Schneller Zugriff auf einen beliebigen Datensatz (**wahlfreiem Zugriff**).
  - Bezeichnung „**Internes Sortieren**“
  
- Dies ist in manchen Fällen nicht möglich:
  - bei sehr großen Datenbeständen z.B. auf Hintergrundspeichern (externen Speichern) mit sequentiellm Zugriff.
  - bei verketteten Listen.
  
- Hier werden Verfahren verwendet, die lediglich **sequentiellen Zugriff** benötigen.
  - Bezeichnung „**Externes Sortieren**“.
  - Wichtigstes Verfahren: **Merge-Sort**.



- **Prinzip:**

- Sortiere die Daten paarweise



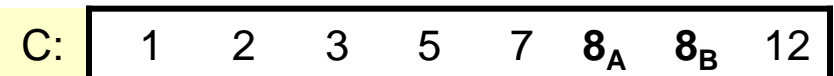
- Füge jeweils zwei Paare zu sortierten Viererfolgen zusammen (Mischen)



- Füge zwei Vierfolgen zu einer sortierten Achterfolge zusammen (Mischen)



- Usw. bei größeren Datensätzen.
- Der eigentliche Aufwand liegt in den Mischvorgängen.



8 aus A hat Vorrang vor 8 aus B ⇒ MergeSort ist stabil.

- Prinzip (iterative Variante):

C: 

|   |   |   |   |    |   |
|---|---|---|---|----|---|
| 5 | 1 | 7 | 3 | 12 | 2 |
|---|---|---|---|----|---|

C: 

|   |   |   |   |   |    |
|---|---|---|---|---|----|
| 1 | 5 | 3 | 7 | 2 | 12 |
|---|---|---|---|---|----|

C: 

|   |   |   |   |   |    |
|---|---|---|---|---|----|
| 1 | 3 | 5 | 7 | 2 | 12 |
|---|---|---|---|---|----|

C: 

|   |   |   |   |   |    |
|---|---|---|---|---|----|
| 1 | 2 | 3 | 5 | 7 | 12 |
|---|---|---|---|---|----|

Die letzten Daten „passen“ nicht mehr in Vierergruppe.

- Außer der bisher behandelten iterativen Variante gibt es auch eine rekursive Variante.
- Krumme Zahlen werden dabei gleichmäßiger aufgeteilt, wodurch die rekursive Variante etwas schneller ist.



## Merge

Eingabe:  $n_A$  sortierte Gruppen in Sequenz  $A$ ,  
 $n_B$  sortierte Gruppen in Sequenz  $B$ ,  
Gruppenlänge  $len$  mit  $(n_A+n_B)*len = N$   
Ausgabe:  $N$  Elemente in Sequenz  $C$   
in sortierten Gruppen der Länge  $2*len$

Über alle Paare  $g_A, g_B$  von Gruppen der Länge  $len$  aus Sequenz  $A$   
bzw. aus Sequenz  $B$

Solange noch Elemente sowohl in  $g_A$  als auch in  $g_B$

$a =$  kleinstes Element aus  $g_A$ ,  $b =$  kleinstes Element aus  $g_B$

Wahr

$a \leq b$

Falsch

Entnehme  $a$  aus  $g_A$  und  
hänge es an  $C$  an

Entnehme  $b$  aus  $g_B$  und  
hänge es an  $C$  an

Verschiebe restliche Elemente von  $g_A$  nach  $C$

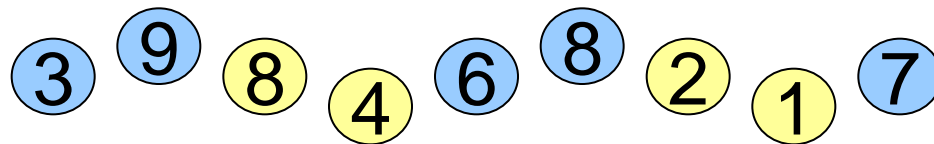
Verschiebe restliche Elemente von  $g_B$  nach  $C$

- Sequenz  $C$  wird  $k = (\lg n)$ -mal zerlegt und dann wieder zusammengemischt.
  - Verteilen und Mischen erfordern jeweils  $O(n)$  Operationen.
- $\Rightarrow T_{\text{mergesort}}(n) \in O(n \log n)$  (auch im worst case)

- Viele sinnvolle Optimierungsmöglichkeiten.
  1. Feld wird nicht in Einzelelemente geteilt, sondern in Gruppen zu  $n$  Elementen, die im 1. Schritt mit InsertionSort sortiert werden.
    - **Ähnlich wie bei QuickSort.**
    - **gcj-Java nimmt 6 Elemente als Grenze.**
    - **Python nimmt 64 Elemente als Grenze (variiert aber manchmal).**
    - **Sun-Java hat diese Optimierung nicht.**

- Java überprüft beim Zusammenfügen:
  - **ist das kleinste Element der einen Teilfolge größer ist als das größte Element der anderen Teilfolge?**
  - **wenn ja, beschränkt sich das Zusammenfügen auf das Hintereinandersetzen der beiden Teilfolgen.**
  - **Damit wird eine Vorsortierung ausgenutzt.**

- Weitergehende Ausnutzung der Vorsortierung.
- Jede Zahlenfolge besteht aus Teilstücken, die abwechselnd monoton steigend und monoton fallend sind.



Die Idee ist, diese bereits sortierten Teilstücke als Ausgangsbasis des Merge-Sorts zu nehmen.

- Bei nahezu sortierten Feldern werden die Teilstücke sehr groß und das Verfahren sehr schnell ( $O(n)$  im Best case).



| Optimierung           | Java (Sun) | Java (gcj) | Python |
|-----------------------|------------|------------|--------|
| Felduntergrenze       | -          | 6          | 64     |
| Trivialfall           | x          | x          | x      |
| Natürlicher Mergesort | -          | -          | x      |

- Merge-Sort auf externen Laufwerken:
  - Teile die Daten in zwei gleich große Dateien A und B.
  - Lese jeweils ein Datum von A und B. Schreibe das sortierte Paar abwechselnd in zwei neue Dateien C und D.
  - Lese jeweils ein Paar aus C und D. Verschmelze die Paare und schreibe die Vierergruppen abwechselnd in A und B.
  - usw.
  - Wiederhole, bis in einer Datei die komplette sortierte Folgesthet.

## MergeSort:

- $O(n \log n)$ , auch im Worst case
- MergeSort kann auch externe Daten sortieren.
- Stabil
- Benötigt zusätzlichen Speicherplatz.
- Durchschnittlich langsamer als Quicksort.
- Nutzt Vorsortierung nicht aus

## Optimierung: Natural MergeSort

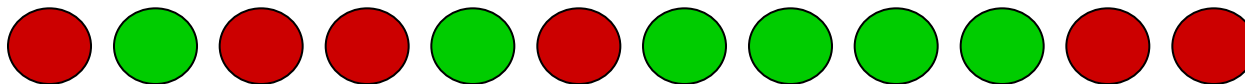
- Nutzt Vorsortierung aus.
- Bei vorsortierten Daten häufig schneller als  $O(n \log n)$ .

# Besser als $O(n \log n)$ : Spezialisierte Sortierverfahren

- **Radix-Sort** heißt eine Gruppe von Sortierverfahren mit folgender Eigenschaft:
  - **Der Sortiervorgang erfolgt mehrstufig.**
  - **Zunächst wird eine Grobsortierung vorgenommen, zu der nur ein Teil des Schlüssels (z.B. der erste Buchstabe) herangezogen wird.**
  - **Die grob sortierten Bereiche werden dann feinsortiert, wobei schrittweise der restliche Teil des Schlüssels verwendet wird.**
- **Spezialisiert** heißt hier, dass die Aufteilung in Grob- und Feinstufen auf das Problem angepasst sein muss.

- Radix-Sort baut auf einer Sortierung in einzelnen Stufen auf.
  - **Diese Stufen können unterschiedlich sein.**
- Beispiel: Zweistellige Integer-Werte sollen sortiert werden.

- Möglichkeit 1: In der 1. Stufe werden die Werte  $\geq 50$  von den Werten  $< 50$  getrennt.
  - Innerhalb jeder Stufe Aufteilung in nur 2 Gruppen.
  - Große Ähnlichkeit mit Quick-Sort.
  - Namen:  
**Radix-Quick-Sort, Binärer Quick-Sort, Radix-Exchange-Sort**
  - Leider auch nur so schnell wie Quicksort ( $O(n \log n)$ ).
  - Wird nicht weiter behandelt.



Die anderen Varianten benutzen Bucket-Sort (nächste Folien).

- Möglichkeit 2: In der 1. Stufe wird die erste Ziffer betrachtet.
  - **Benutzt Bucket-Sort mit Teillisten.**
  - **Namen:**  
**MSD-Radix-Sort** (MSD = most significant digits)
- Möglichkeit 3: In der 1. Stufe wird die letzte Ziffer betrachtet.
  - **Benutzt Bucket-Sort mit schlüsselindiziertem Zählen.**
  - **Namen:**  
**LSD-Radix-Sort, Straight Radix Sort**  
(LSD = least significant digits)



- Eignet sich für den Fall, dass der Wertebereich der Schlüssel eng begrenzt ist.
- Beispiel: Sortieren von Fußballspielern nach Toren:

|    |            |
|----|------------|
| 20 | Altintop   |
| 12 | Amanatidis |
| 14 | Ballack    |
| 21 | Berbatov   |
| 15 | Klasnic    |
| 12 | Klimowicz  |
| 25 | Klose      |
| 12 | Marcelinho |
| 17 | Makaay     |
| 12 | Podolski   |
| 13 | Smolarek   |
| 12 | Thurk      |
| 16 | Vittek     |



- Datensätze in Eimer werfen.
- Inhalt der Eimer der Reihe nach in Hauptliste einfügen.
- Datensätze sind nach Toren sortiert.

- Die Eimer sind Array-Lists
- 1. Schritt (Partitionierungsphase): Verteilung der Daten auf die Teillisten (Eimer).
- 2. Schritt (Sammelphase): Die Daten werden nach der Reihe aus den Eimern wieder in die Liste kopiert.

- Dieses Verfahren wird mehrstufig durchgeführt. Der Inhalt eines Eimers wird noch einmal mit Bucket-Sort sortiert.



- Wie bei Quick-Sort bricht man auch bei MSB-Radix-Sort ab, wenn die Teilfelder (Eimer-Inhalte) zu klein werden.
- Zum letzten Schritt benutzt man auch hier Insertion-Sort

- In der Praxis nimmt man nicht 10 Eimer (Dezimalsystem) sondern richtet sich nach dem Binärsystem, d.h.  $2^n$  Eimer.
- Wie groß sollte man  $n$  wählen?
  - **Zu wenige Eimer sind schlecht, weil man dann zu viele Schritte braucht.**
  - **Zu viele Eimer sind auch schlecht, weil man dann zu viele leere Eimer durchsuchen muss.**
- Es wird vorgeschlagen:
  - **Sedgewick: Für 64-bit-Schlüssel (long)  $2^{16}$  (65536) Eimer.**
  - **Linux-Related: Für 32-bit-Schlüssel (int)  $2^{11}$  (2048) Eimer.**
- Einsatz lohnt sich nur, wenn die Anzahl der zu sortierenden Werte deutlich größer ist, als die Anzahl der Eimer.

# Vergleich der Sortierverfahren

# Vergleich der Sortierverfahren

- Eigene Laufzeitmessungen

| <b>Verfahren</b> | <b>Laufzeit<br/>1000 Elem.</b> | <b>Laufzeit<br/>10.000.000<br/>Elem.</b> |
|------------------|--------------------------------|------------------------------------------|
| Simple Sort      | <i>1,2 ms</i>                  | <i>1d 15 h</i>                           |
| Insertion Sort   | <i>0,2 ms</i>                  | <i>4h 20 min</i>                         |
| Quick Sort       | <i>0,07 ms</i>                 | <i>1,2 s</i>                             |

## Vergleich der Sortierverfahren

| Verfahren          | Laufzeitmessungen<br>(nach Wirth, Sedgewick) | Vorsortierung<br>ausnutzen | Worst<br>Case ok | Zus.<br>Speicher | Stabil |
|--------------------|----------------------------------------------|----------------------------|------------------|------------------|--------|
| QuickSort          | 100                                          |                            |                  | $\Phi(\log n)$   |        |
| HeapSort           | 150-200                                      |                            | X                | $\Phi(1)$        |        |
| MergeSort          | 150-200                                      | X                          | X                | $\Phi(n)$        | X      |
| MSB-<br>Radix-Sort | 85<br>(bei 100.000 El.)                      |                            | X                | $\Phi(n)$        | X      |



## Vorsortierte Felder

- Welches Verfahren ist für **nahezu** sortierte Felder am besten?
- Experiment
  - **2.000.000 Elemente**
  - **Feldinhalt = Feldindex**
  - **n Elemente werden paarweise vertauscht**
  - **n variiert zwischen 0 (sortiert) und 3.000.000 (unsortiert).**

| Sortierverfahren   | Sortiertes Feld | Unsortiertes Feld |
|--------------------|-----------------|-------------------|
| Bubble             | $O(n)$          | $O(n^2)$          |
| Insertion          | $O(n)$          | $O(n^2)$          |
| Quick („Pur“)      | $O(n^2)$        | $O(n \log n)$     |
| Heap, Merge, Quick | $O(n \log n)$   | $O(n \log n)$     |
| Nat. Merge         | $O(n)$          | $O(n \log n)$     |
| Radix              | $O(n)$          | $O(n)$            |

# Vorsortierte Felder (2)

