

Datenbank-Implementierungstechniken | Vortrag

RANDOMISIERTE VERFAHREN BEI DER ANFRAGEOPTIMIERUNG

Patrick Naumann
7. Juli 2015
HTWK Leipzig



Randomisierte Verfahren

- **Vermeidung des vollständigen Durchsuchens des Suchraumes**
 - NP-Schweres Problem
 - Hohe Laufzeit bei großen Verbundsanfragen
 - Ca. 10–15 Relationen für linksorientierte Bäume
 - Ca. 6–7 Relationen für buschige Bäume
- **Keine Garantie für optimalen Startplan**
 - Hoffnung auf gute Annäherung an das Optimum

Randomisierte Verfahren

▪ Linksortorientierte Bäume

- Beispiel:
- Verbesserung durch:
 - Swap (Tauschen)
 - 3Cycle (Rotieren)

$$(((r_1 \bowtie r_2) \bowtie r_3) \bowtie r_4)$$

$$r_1 \bowtie r_2 \Rightarrow r_2 \bowtie r_1$$

$$r_1 \bowtie (r_2 \bowtie r_3) \Rightarrow r_3 \bowtie (r_1 \bowtie r_2)$$

▪ Buschige Bäume

- Beispiel:
- Verbesserung durch:
 - Kommutativität:
 - Assoziativität:
 - Linkstausch:
 - Rechtstausch:

$$((r_1 \bowtie r_2) \bowtie (r_3 \bowtie r_4))$$

$$r_1 \bowtie r_2 \Rightarrow r_2 \bowtie r_1$$

$$(r_1 \bowtie r_2) \bowtie r_3 \Rightarrow r_1 \bowtie (r_2 \bowtie r_3)$$

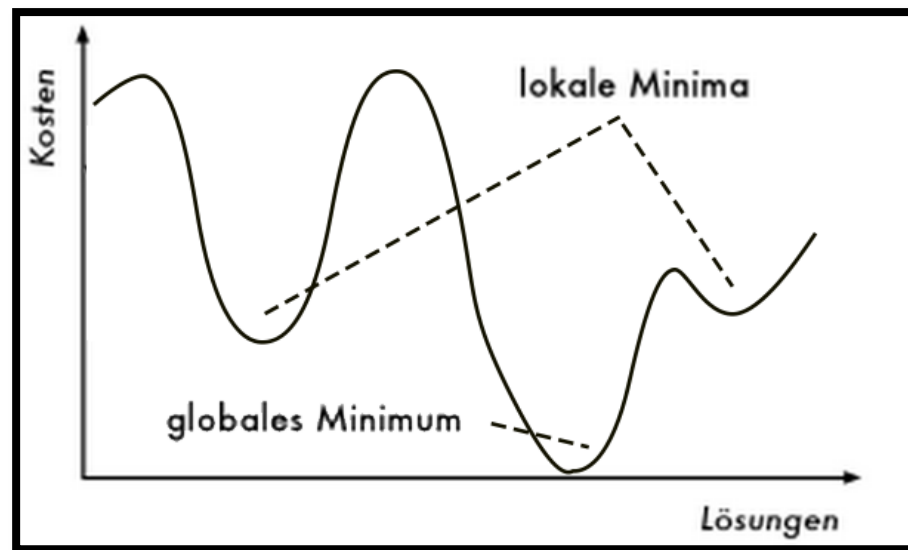
$$(r_1 \bowtie r_2) \bowtie r_3 \Rightarrow (r_1 \bowtie r_3) \bowtie r_2$$

$$r_1 \bowtie (r_2 \bowtie r_3) \Rightarrow r_2 \bowtie (r_1 \bowtie r_3)$$

Hill-Climbing

- **Betrachtung von Nachbarn eines zufälligen Startpunkts (Plan)**
 - Verfolgung solange bessere Ergebnisse erzielt werden
 - Es wird jedes bessere Ergebnis Angenommen
- **Optimal wenn nur ein Minimum existiert**
 - Unwahrscheinlich...
 - Bleibt im ersten (lokalen) Minimum hängen
 - Hohe Wahrscheinlichkeit nicht optimal zu sein

Hill-Climbing



Iterative Improvement

- **Abwandlung des Hill-Climbing-Algorithmus**
 - Betrachtung zufälliger Nachbarlösungen
 - Reduktion der Anzahl der umliegenden Punkte
 - Mögl. Vermeidung lokaler Minima
 - Betrachtung mehrerer zufälliger Startpunkte
 - Annahme des niedrigsten lokalen Minimums
- **Mögl. Abbruchbedingungen**
 - Zeitlimit
 - Festgelegte Anzahl von Startpunkten

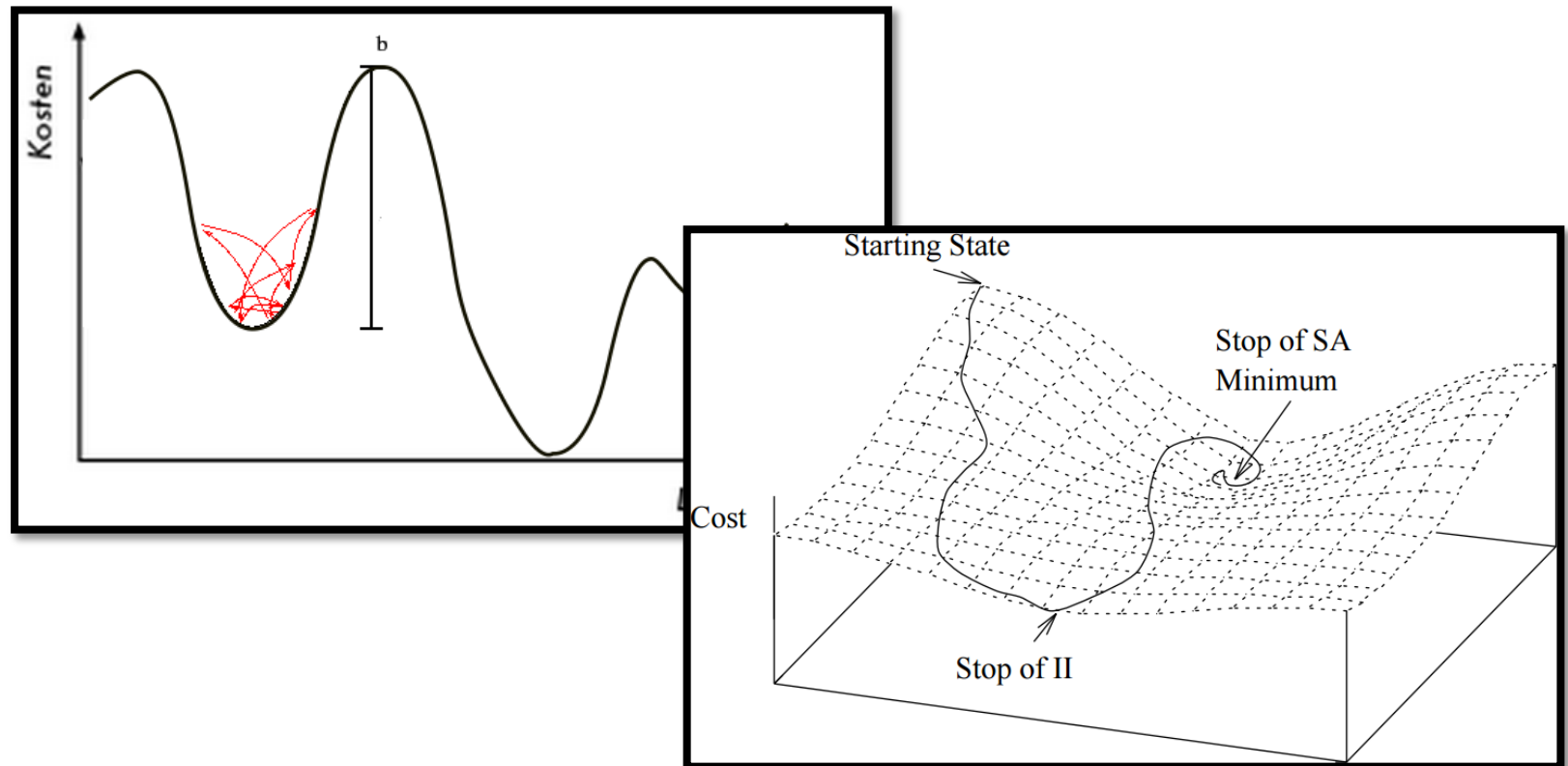
Iterative Improvement

```
Input : Verbundanfrage Q
Output : Anfrageplan für Q
minCost  $\leftarrow \infty$ 
while  $\neg$ Abbruchbedingug erfüllt do
    newPlan  $\leftarrow$  randomSolution(Q)
    while  $\neg$ localMinimum(newPlan) do
        neighbour  $\leftarrow$  randomNeighbour(newPlan)
        if cost(neighbor) < cost(newPlan) then
            newPlan  $\leftarrow$  neighbor
        end
    end
    if cost(newPlan) < minCost then
        optPlan  $\leftarrow$  newPlan;
        minCost  $\leftarrow$  cost(optPlan)
    end
end
return optPlan
```

Simulated Annealing (Abkühlung)

- **Abwandlung des Iterative Improvement**
 - Auch schlechtere Lösungen können zugelassen werden
 - Abhängig von der Annahmewahrscheinlichkeit („Temperatur“ T)
 - Vergleich mit zufälligem Wert zwischen 0 ... 1
 - Temperatur nimmt mit der Zeit ab
 - System kommt zum „erliegen“
- **Aufgrund von Laufzeit (Temperatur) kein Einzug in DBMS**

Simulated Annealing



Simulated Annealing

```

Input : Verbundanfrage Q
Output : Anfrageplan für Q
minCost  $\leftarrow \infty$ 
plan  $\leftarrow$  randomSolution(Q)
while  $\neg$ frozen do
    while  $\neg$ equilibrium(plan) do
        newPlan  $\leftarrow$  randomNeighbour(plan)
         $\Delta$ cost  $\leftarrow$  cost(newPlan) – cost(plan)
        if  $\Delta$ cost  $\leq$  0 then
            plan  $\leftarrow$  newPlan
        else
            with probability  $e^{\Delta\text{cost}/T}$  do
                plan  $\leftarrow$  newPlan
            end
        end
        if cost(plan) < minCost then
            newPlan  $\leftarrow$  neighbor
        end
    end
    if cost(newPlan) < minCost then
        optPlan  $\leftarrow$  newPlan;
        minCost  $\leftarrow$  cost(optPlan)
    end
    T  $\leftarrow$  reduce(T)
end
return optPlan

```

Weitere Algorithmen

- **Two-Phase Optimization**
 - Iterative Improvement für Suche von lokale Minima
 - Bestes lokales Minima wird ausgewählt
 - Simulated Annealing für Suche von besseren Lösungen
 - Reduktion der Anfangstemperatur
- **Toured Simulated Annealing**
 - Mehrere Simulated Annealing Suchen
 - Reduktion der Anfangstemperatur

Genetische Algorithmen

- **Basiert auf der biologischen Evolution**
 - Lösungen werden von Generation zu Generation weitergegeben
 - Mehrere Lösungen werden erarbeitet (population)
 - Lösungen werden als String (chromosomes) dargestellt
 - Bestehend aus Zeichen (genes)
 - Encoding
 - Max- oder Minimierungsfunktion stellt Eignung (fitness) dar
 - Eignung als optimale Lösung sollte sich mit jeder Generation verbessern

Genetische Algorithmen

- **Bestandteile des Algorithmus**
 - 0. Generation wird zufällig erstellt
 - Jede weitere Generation entsteht durch:
 - Selection
 - Kreuzung (Crossover)
 - Mutation
 - Abbruch wenn eine Generation bestimmte Kriterien erfüllt

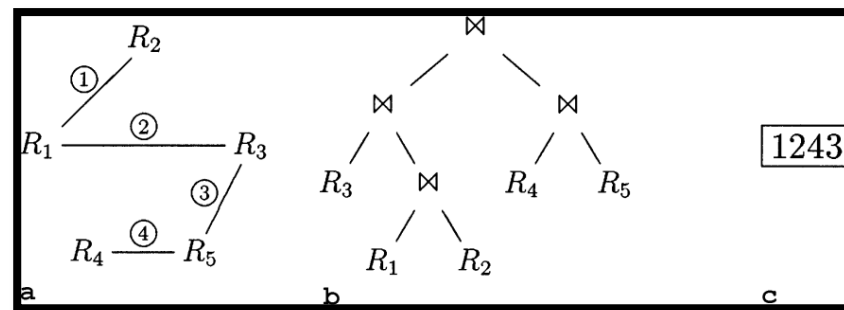
Genetische Algorithmen

■ Encoding

- Geordnete Liste
 - Linksortierter Baum

$$\left(\left((r_1 \bowtie r_4) \bowtie r_3 \right) \bowtie r_2 \right) \bowtie r_5 \Rightarrow 14325$$

- Buschiger Baum



Genetische Algorithmen

■ Encoding

■ Ordinalzahl

■ Linksortierter Baum

- $[1,2,3,4,5] \rightarrow 1 - 1$

- $[2,3,4,5] \rightarrow 4 - 3$

- $[2,3,5] \rightarrow 3 - 2$

- ...

- $\Rightarrow 13211$

■ Buschiger Baum

- $[1,2,3,4,5] \rightarrow 1 + 2 - 12$

- $[12,3,4,5] \rightarrow 3 + 12 - 21$

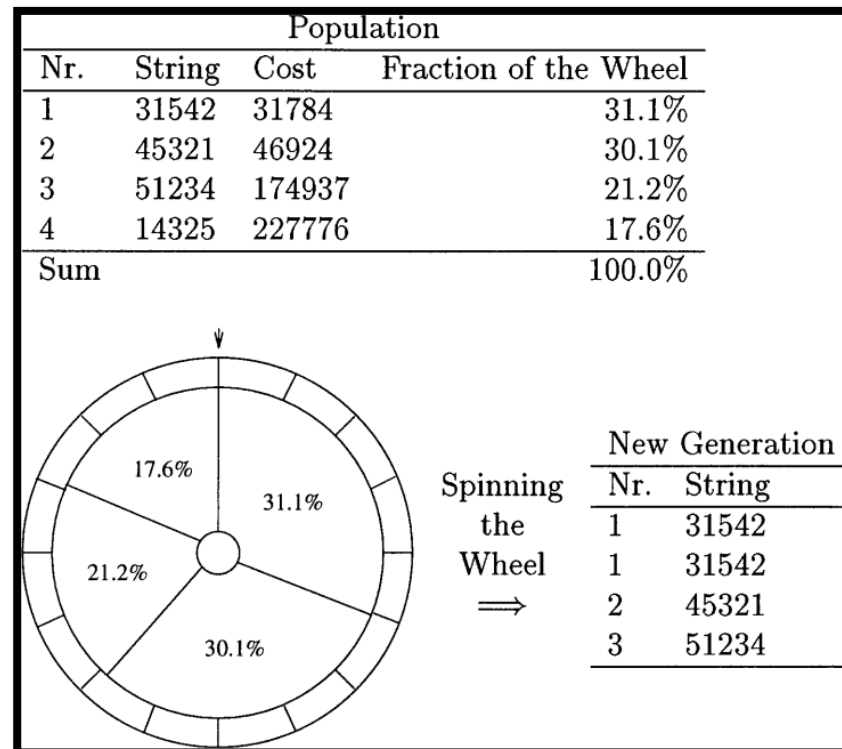
- $[123,4,5] \rightarrow 4 + 5 - 23$

- ...

- $\Rightarrow 12\ 21\ 23\ 12$

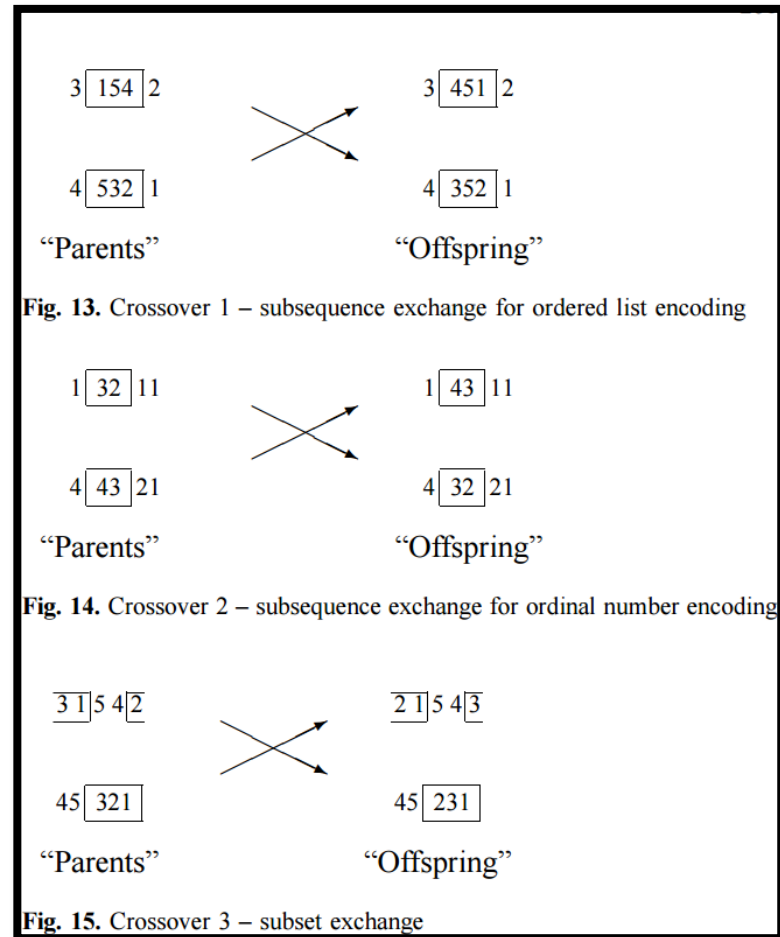
Genetische Algorithmen

■ Selection



Genetische Algorithmen

■ Kreuzung



Genetische Algorithmen

- **Derzeit nur in PostgreSQL (GEQO)**
 - Optional
 - Verwendet immer gleichen Seed um Evolutionen zu „Wiederholen“
 - Gute Möglichkeit zur Parallelisierung

Quellen

- **Datenbanken – Implementierungstechniken**
 - Gunter Saake, Andreas Heuer, Kai-Uwe Sattler [2011]
- **Heuristic and Randomized Optimization for the Join Ordering Problem**
 - Steinbrunn, Moerkotte, Kemper [The VLDB Journal Bd. 3, Nr. 6, S. 191–208, 1997]
- <http://www-db.in.tum.de/teaching/ws1415/queryopt/session08.pdf>
 - [2015]
- <https://www.db.informatik.uni-kassel.de/Lehre/SS2014/SeminarDB/DB-ImplTech1213.pdf>
 - [2015]