

Histogramme in der Anfrageoptimierung

Datenbanken Implementierungstechniken SS15

Chrstian Weckel

06.07.15

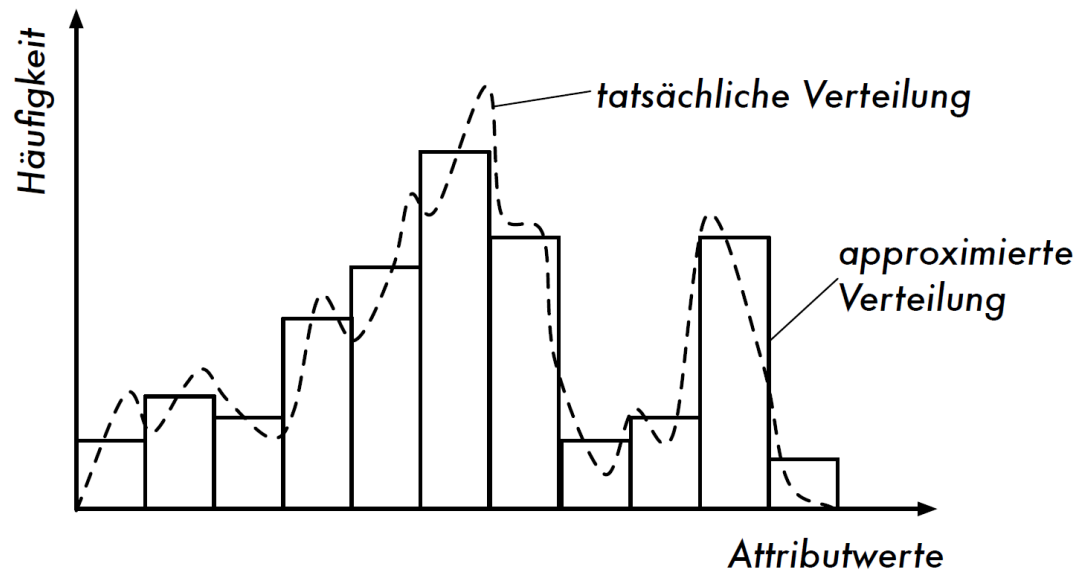
Gliederung

1. Motivation
2. Histogramme
3. Vorteile
4. Freiheitsgrade
5. Histogrammarten
6. Konstruktion und Pflege

1 Motivation

- ▶ Optimierung einer Datenbank
- ▶ Kostenabschätzung
 - ▶ nimmt Gleichverteilung der Quellwerte an
 - ▶ fehlerhafte Kostenabschätzung
 - ▶ reale Datenverteilung muss geschätzt werden

2 Histogramm



- ▶ Partitionierung der Daten in Buckets
- ▶ Approximation der Quellwerte pro Bucket

3 Vorteile

- ▶ kann mit ungleich verteilten Quellwerten umgehen
- ▶ sehr genaue Annäherung
- ▶ geringer Speicherbedarf
- ▶ optimier- und anpassbar
- ▶ besonders geeignet bei Bereichsanfragen

4 Freiheitsgrade

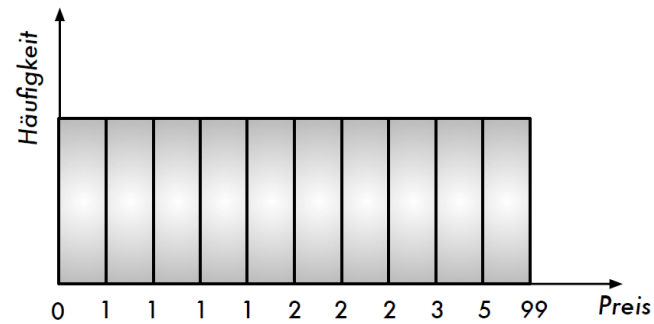
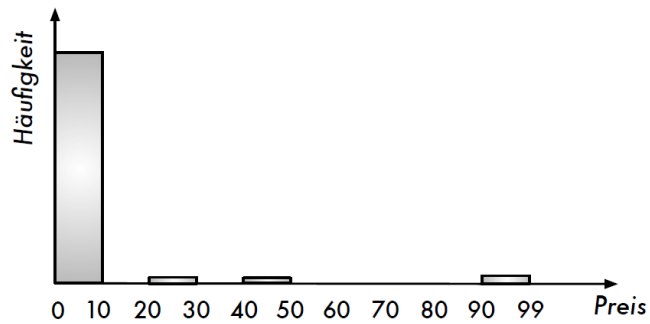
- ▶ Anzahl der Elemente
- ▶ Sortierparameter
- ▶ Quellparameter
- ▶ Approximation der Werte

5 Histogrammarten

- ▶ übliche Histogramme:
 - ▶ Equi-sum
 - ▶ Equi-width
 - ▶ Equi-depth
 - ▶ V-optimal
 - ▶ Spline-basiert
- ▶ Histogramme mit Vermeidung von Gruppierung abweichender Quellwerte:
 - ▶ Maxdiff
 - ▶ Compressed
 - ▶ End-biased

5 Histogrammarten: Beispiel

Preis	1	2	3	5	10	30	50	99
Anzahl	100	75	30	20	15	3	2	5



6 Konstruktion und Pflege in DBMS

- ▶ statischer Ansatz:
 - ▶ einmalig erstellt
 - ▶ bei Anpassung Neuerstellung notwendig
- ▶ dynamischer Ansatz:
 - ▶ passt sich Datenänderungen an
 - ▶ verwaltet Stichprobe der Relation
 - ▶ Änderungen in Datenbank wird auf Stichprobe abgebildet
 - ▶ nutzt Anfrageergebnisse zur Anpassung