

PD Dr. rer. nat. habil. Sven Groppe

## Übungen zur Vorlesung

# Mobile und Verteilte Datenbanken

WS 2011/2012

## Übung 11 – Parallele Semantic Web Datenbank

### Aufgabe 1:

Sei  $H$  eine endliche Menge von Hash-Funktionen mit Signatur  $U \rightarrow \{0, 1, \dots, m-1\}$ .  $H$  wird universell genannt, falls für jedes Paar von verschiedenen Schlüsseln  $x, y \in U$  die Anzahl der Hash-Funktionen  $h \in H$  mit  $h(x) = h(y)$  genau  $|H|/m$  ist. Mit anderen Worten, die Chance einer Kollision zwischen  $x$  und  $y$  mit  $x \neq y$  ist genau  $1/m$  bei einer zufällig gewählten Hash-Funktion  $h \in H$ , also genau so groß wie die Chance einer Kollision bei Hash-Funktionen, die alle Elemente aus  $U$  zufällig auf  $\{0, 1, \dots, m-1\}$  abbilden. Bemerken Sie, dass für die letztgenannte Art von Hash-Funktionen eine Tabelle der Größe  $|U|$  notwendig wäre, um die Abbildungen von  $U$  auf  $\{0, 1, \dots, m-1\}$  zu verwalten.

Die folgende Klasse  $H_{prim}$  von Hash-Funktionen ist universell:

Sei  $m$  eine Primzahl. Wir zerlegen einen Schlüssel  $x$  in  $r+1$  Unterkomponenten, so dass  $x = \langle x_0, x_1, \dots, x_r \rangle$ . Die einzige Bedingung ist, dass der maximale Wert einer Unterkomponente kleiner als  $m$  sein muss. Sei nun weiterhin  $a = \langle a_0, a_1, \dots, a_r \rangle$  eine Sequenz von  $r+1$  Elementen, die zufällig aus  $\{0, 1, \dots, m-1\}$  gewählt sind. Dann sei eine entsprechende Hash-Funktion  $h_a \in H_{prim}$  durch  $h_a(x) = (\sum_{i=0}^r a_i * x_i) \bmod m$  definiert. Bemerken Sie, dass wir durch Zahlentheorie beweisen können, dass  $H_{prim}$  universell ist.

a) Berechnen Sie die Hash-Werte von 3, 20 und 82 bezüglich der Hash-Funktion  $h_{\langle 5, 8, 9 \rangle}(x) = (\sum_{i=0}^r a_i * x_i) \bmod 13$ . Eine Unterkomponente  $x_i$  habe dabei den Wertebereich  $\{0, \dots, 7\}$  und werde durch  $x_i = (x \div 8^i) \bmod 8$  berechnet.

b) Zeigen Sie: Falls wir  $a_i > 0$  fordern würden für jedes  $a_i \in \{a_0, a_1, \dots, a_r\}$ , dann ist  $H_{prim}$  nicht mehr universell.

### Aufgabe 2:

Füge die fehlenden Informationen in dem folgenden Histogramm-Index hinzu und berechne das Histogramm mit zwei Intervallen für die Variable  $?v$  und dem Schlüssel  $(5, ?v, ?o)$ :

### B+-tree for fast histogram computation (SPO collation order)

Key:  
 #triples:  
 #distinct subjects:  
 #distinct predicates:  
 #distinct objects:  
 subject } of first triple  
 predicate } different from  
 object } previous triple

Key:  
 #triples:  
 #distinct subjects:  
 #distinct predicates:  
 #distinct objects:  
 subject } of first triple  
 predicate } different from  
 object } previous triple

Leaf nodes

(1, 2, 3), (2, 2, 5), (3, 3, 3), (4, 2, 4)	(5, 1, 1), (5, 2, 6), (5, 3, 5), (5, 3, 6)	(5, 3, 7), (5, 4, 1), (5, 4, 5), (5, 4, 7)	(5, 5, 7), (5, 5, 8), (5, 6, 1), (5, 6, 2)	(5, 7, 1), (5, 7, 3), (5, 8, 3), (5, 8, 4)	(5, 8, 5), (5, 9, 4), (5, 9, 7), (8, 6, 3)
--	--	--	--	--	--

