

Prof. Dr. Agnès Voisard, Nicolas Lehmann

Datenbanksysteme, SoSe 2017

Übungsblatt 6

Tutor: Nicolas Lehmann

Tutorium 10

Boyan Hristov, Julian Habib

7. Juli 2017

Link zum Git Repository: <https://github.com/BoyanH/Freie-Universitaet-Berlin/tree/master/Datenbanksysteme/Solutions/homework6>

1. Aufgabe

1. SQL Befehle in relationale Algebra umwandeln und diese neu (optimiert) schreiben.

Beispiel:

Projektionen so früh wie möglich machen. Da Join Operationen sehr aufwändig sind. Wir haben 3 Tabellen. Eine mit Personen, eine mit ihre Hausaufgaben und eine mit ihre Noten. Wir wollen für eine Person die Note für die letzte Hausaufgabe (bei Datum) bekommen. Es wäre besser nachdem wir Person und Hausaufgaben Joinen, erstmal die letzte Hausaufgabe zu finden und dann die Note dafür zu finden (wieder mit Join zwischen Note und Hausaufgabe) als die drei Tabellen erst mal zu Joinen und so eine ganz aufwändige Joinoperation ausführen zu müssen anstatt von zwei kleinere.

2. Indizes

Erklärung: Wir wollen so wenig wie möglich nach einem Eintrag suchen und auch so wenig wie möglich Speicher für die Indixdateien verbrauchen. Dafür wäre eine B^+ -Baum Lösung in den meisten Fällen wesentlich effizienter als Hashfunktionsverfahren. Welches Verfahren aber für welche Situation am besten geeignet ist ist unter gute Indizes gemeint.

3. Statistische Optimierungen

Beispiel:

Mit der Zeit erfährt man, ungefähr wie viele Tupels nach einem Join-Operation entstehen. Deswegen kann man sehen, dass vielleicht $A \bowtie B$ viel mehrere Tupels produziert als $B \bowtie A$. Deswegen kann man mit der Zeit sicherstellen, dass die letztere Operation effizienter ist und diese, wo möglich, anwenden.

2. Aufgabe

a)

b) Wir wählen die Zentren $\{ (20,20), (10,80), (80,40) \}$

Abstände zwischen je Punkt und Zentrum:

1.

$$\sqrt{(22-20)^2 + (57-20)^2} \approx 37$$

$$\sqrt{(22-10)^2 + (57-80)^2} \approx 25.94$$

$$\sqrt{(22-80)^2 + (57-40)^2} \approx 60.44$$

Gehört also zu 2. Zentrum

2.

$$\sqrt{(79-20)^2 + (46-20)^2} \approx 64.47$$

$$\sqrt{(79-10)^2 + (46-80)^2} \approx 76.92$$

$$\sqrt{(79-80)^2 + (46-40)^2} \approx 6.08$$

Gehört also zu 3. Zentrum

3.

$$\sqrt{(61-20)^2 + (52-20)^2} \approx 52$$

$$\sqrt{(61-10)^2 + (52-80)^2} \approx 58.2$$

$$\sqrt{(61-80)^2 + (52-40)^2} \approx 22.47$$

Gehört also zu 3. Zentrum

4.

$$\sqrt{(14-20)^2 + (20-20)^2} \approx 6$$

$$\sqrt{(14-10)^2 + (20-80)^2} \approx 60.13$$

$$\sqrt{(14-80)^2 + (20-40)^2} \approx 68.96$$

Gehört also zu 1. Zentrum

5.

$$\sqrt{(22-20)^2 + (14-20)^2} \approx 6.324555320336759$$

$$\sqrt{(22-10)^2 + (14-80)^2} \approx 67.08203932499369$$

$$\sqrt{(22-80)^2 + (14-40)^2} \approx 63.56099432828282$$

Gehört also zu 1. Zentrum

6.

$$\sqrt{(75-20)^2 + (17-20)^2} \approx 55.08175741568164$$

$$\sqrt{(75-10)^2 + (17-80)^2} \approx 90.52071586106685$$

$$\sqrt{(75-80)^2 + (17-40)^2} \approx 23.53720459187964$$

Gehört also zu 3. Zentrum

7.
 $\sqrt{(81-20)^2 + (40-20)^2} \approx 64.19501538281614$
 $\sqrt{(81-10)^2 + (40-80)^2} \approx 81.49233092751734$
 $\sqrt{(81-80)^2 + (40-40)^2} \approx 1.0$
 Gehört also zu 3. Zentrum

8.
 $\sqrt{(2-20)^2 + (91-20)^2} \approx 73.24616030891995$
 $\sqrt{(2-10)^2 + (91-80)^2} \approx 13.601470508735444$
 $\sqrt{(2-80)^2 + (91-40)^2} \approx 93.1933474020544$
 Gehört also zu 2. Zentrum

9.
 $\sqrt{(96-20)^2 + (63-20)^2} \approx 87.32124598286491$
 $\sqrt{(96-10)^2 + (63-80)^2} \approx 87.66413177577246$
 $\sqrt{(96-80)^2 + (63-40)^2} \approx 28.0178514522438$
 Gehört also zu 3. Zentrum

10.
 $\sqrt{(4-20)^2 + (31-20)^2} \approx 19.4164878389476$
 $\sqrt{(4-10)^2 + (31-80)^2} \approx 49.36598018878993$
 $\sqrt{(4-80)^2 + (31-40)^2} \approx 76.53103945458993$
 Gehört also zu 1. Zentrum

Jetzt müssen wir die neue Zentren berechnen

1. Mittelwert von 4,5 und 10 Punkt $((\frac{14+22+4}{3}), (\frac{20+14+31}{3})) \approx (13.3, 21.7)$
 $\sqrt{(22-20)^2 + (32-20)^2} \approx 6.91 > \frac{3}{5}$

2. Mittelwert von 1 und 8 Punkt
 $((\frac{22+2}{2}), (\frac{57+91}{2})) = (12, 74)$
 $\sqrt{(10-12)^2 + (80-74)^2} \approx 6.32 > \frac{3}{5}$

3. Mittelwert von 2,3,6,7 und 9 Punkt
 $((\frac{79,61,75,81,96}{5}), (\frac{46,52,17,40,63}{5})) = (78.4, 43.6)$
 $\sqrt{(10-12)^2 + (80-74)^2} \approx 3.93 > \frac{3}{5}$

Also der Algorithmus geht weiter mit 2. Schritt.

1.
 $\sqrt{(22-13.3)^2 + (57-21.7)^2} \approx 36.35629244023653$
 $\sqrt{(22-12)^2 + (57-74)^2} \approx 19.72308292331602$
 $\sqrt{(22-78.4)^2 + (57-43.6)^2} \approx 57.969992237363634$
 Gehört also zu 2. Zentrum

2.
 $\sqrt{(79 - 13.3)^2 + (46 - 21.7)^2} \approx 70.04983940024417$
 $\sqrt{(79 - 12)^2 + (46 - 74)^2} \approx 72.61542535852834$
 $\sqrt{(79 - 78.4)^2 + (46 - 43.6)^2} \approx 2.473863375370594$
 Gehört also zu 3. Zentrum

3.
 $\sqrt{(61 - 13.3)^2 + (52 - 21.7)^2} \approx 56.50999911520084$
 $\sqrt{(61 - 12)^2 + (52 - 74)^2} \approx 53.71219600798314$
 $\sqrt{(61 - 78.4)^2 + (52 - 43.6)^2} \approx 19.321490625725545$
 Gehört also zu 3. Zentrum

4.
 $\sqrt{(14 - 13.3)^2 + (20 - 21.7)^2} \approx 1.8384776310850226$
 $\sqrt{(14 - 12)^2 + (20 - 74)^2} \approx 54.037024344425184$
 $\sqrt{(14 - 78.4)^2 + (20 - 43.6)^2} \approx 68.58804560563014$
 Gehört also zu 1. Zentrum

5.
 $\sqrt{(22 - 13.3)^2 + (14 - 21.7)^2} \approx 11.618089343777658$
 $\sqrt{(22 - 12)^2 + (14 - 74)^2} \approx 60.8276253029822$
 $\sqrt{(22 - 78.4)^2 + (14 - 43.6)^2} \approx 63.69552574553412$
 Gehört also zu 1. Zentrum

6.
 $\sqrt{(75 - 13.3)^2 + (17 - 21.7)^2} \approx 61.87875241146996$
 $\sqrt{(75 - 12)^2 + (17 - 74)^2} \approx 84.95881355103778$
 $\sqrt{(75 - 78.4)^2 + (17 - 43.6)^2} \approx 26.81641288464958$
 Gehört also zu 3. Zentrum

7.
 $\sqrt{(81 - 13.3)^2 + (40 - 21.7)^2} \approx 70.1297369166604$
 $\sqrt{(81 - 12)^2 + (40 - 74)^2} \approx 76.92203845452875$
 $\sqrt{(81 - 78.4)^2 + (40 - 43.6)^2} \approx 4.4407206622349005$
 Gehört also zu 3. Zentrum

8.
 $\sqrt{(2 - 13.3)^2 + (91 - 21.7)^2} \approx 70.21524051087485$
 $\sqrt{(2 - 12)^2 + (91 - 74)^2} \approx 19.72308292331602$
 $\sqrt{(2 - 78.4)^2 + (91 - 43.6)^2} \approx 89.90951006428631$
 Gehört also zu 2. Zentrum

9.
 $\sqrt{(96 - 13.3)^2 + (31 - 21.7)^2} \approx 83.22127131929673$
 $\sqrt{(96 - 12)^2 + (31 - 74)^2} \approx 94.36630754670864$
 $\sqrt{(96 - 78.4)^2 + (31 - 43.6)^2} \approx 21.645322820415494$
 Gehört also zu 3. Zentrum

10.
 $\sqrt{(4 - 13.3)^2 + (31 - 21.7)^2} \approx 13.152186130069786$
 $\sqrt{(4 - 12)^2 + (31 - 74)^2} \approx 43.73785545725808$
 $\sqrt{(4 - 78.4)^2 + (31 - 43.6)^2} \approx 75.45939305348274$
 Gehört also zu 1. Zentrum

Man kann hier merken, dass die Punkte in genau die selben Cluster geladen sind als letztes mal. Es gibt also keinen Sinn weiter zu rechnen, da genau die selbe Cluster Zentren entstehen werden, also der Euklidische Abstand zwischen die alte und die neue Zentren wird je 0 sein, also kleiner als $\frac{3}{5}$.

c) Wieder erstmal berechnen, zu welchen Cluster je Punkt gehört

1.
 $\sqrt{(22 - 15)^2 + (57 - 85)^2} \approx 28.861739379323623$
 $\sqrt{(22 - 85)^2 + (57 - 15)^2} \approx 75.71657678474378$
 Gehört also zu 1. Zentrum

2.
 $\sqrt{(79 - 15)^2 + (46 - 85)^2} \approx 74.94664769020693$
 $\sqrt{(79 - 85)^2 + (46 - 15)^2} \approx 31.575306807693888$
 Gehört also zu 2. Zentrum

3.
 $\sqrt{(61 - 15)^2 + (52 - 85)^2} \approx 56.61271941887264$
 $\sqrt{(61 - 85)^2 + (52 - 15)^2} \approx 44.10215414239989$
 Gehört also zu 2. Zentrum

4.
 $\sqrt{(14 - 15)^2 + (20 - 85)^2} \approx 65.00769185258002$
 $\sqrt{(14 - 85)^2 + (20 - 15)^2} \approx 71.17583859709698$
 Gehört also zu 1. Zentrum

5.
 $\sqrt{(22 - 15)^2 + (14 - 85)^2} \approx 71.34423592694787$
 $\sqrt{(22 - 85)^2 + (14 - 15)^2} \approx 63.00793600809346$
 Gehört also zu 2. Zentrum

6.
 $\sqrt{(75 - 15)^2 + (17 - 85)^2} \approx 90.68627239003708$
 $\sqrt{(75 - 85)^2 + (17 - 15)^2} \approx 10.198039027185569$
 Gehört also zu 2. Zentrum

7.
 $\sqrt{(81-15)^2 + (40-85)^2} \approx 79.88116173416608$
 $\sqrt{(81-85)^2 + (40-15)^2} \approx 25.317977802344327$
 Gehört also zu 2. Zentrum

8.
 $\sqrt{(2-15)^2 + (91-85)^2} \approx 14.317821063276353$
 $\sqrt{(2-85)^2 + (91-15)^2} \approx 112.53888216967503$
 Gehört also zu 1. Zentrum

9.
 $\sqrt{(96-15)^2 + (31-85)^2} \approx 97.3498844375277$
 $\sqrt{(96-85)^2 + (31-15)^2} \approx 19.4164878389476$
 Gehört also zu 2. Zentrum

10.
 $\sqrt{(4-15)^2 + (31-85)^2} \approx 55.10898293381942$
 $\sqrt{(4-85)^2 + (31-15)^2} \approx 82.56512580987206$
 Gehört also zu 1. Zentrum

Jetzt müssen wir die neue Zentren berechnen

1. Mittelwert von 1,4,8 und 10 Punkt $((\frac{22+14+2+4}{4}), (\frac{57+20+91+31}{4})) \approx (10.5, 49.75)$
 $\sqrt{(10.5-15)^2 + (49.75-85)^2} \approx 35.53 > \frac{2}{5}$

2. Mittelwert von 2,3,5,6,7 und 9 Punkt
 $((\frac{79+61+22+75+81+96}{6}), (\frac{46+52+14+17+40+63}{6})) = (69, 38.7)$
 $\sqrt{(15-69)^2 + (85-38.7)^2} \approx 28.6 > \frac{2}{5}$

Also der Algorithmus geht weiter mit 2. Schritt

1.
 $\sqrt{(22-10.5)^2 + (57-49.75)^2} \approx 13.594576124322524$
 $\sqrt{(22-69)^2 + (57-38.7)^2} \approx 50.436990393955895$
 Gehört also zu 1. Zentrum

2.
 $\sqrt{(79-10.5)^2 + (46-49.75)^2} \approx 68.60256919387204$
 $\sqrt{(79-69)^2 + (46-38.7)^2} \approx 12.381033882515627$
 Gehört also zu 2. Zentrum

3.
 $\sqrt{(61-10.5)^2 + (52-49.75)^2} \approx 50.55009891187158$
 $\sqrt{(61-69)^2 + (52-38.7)^2} \approx 15.520631430454106$
 Gehört also zu 2. Zentrum

4.
 $\sqrt{(14-10.5)^2 + (20-49.75)^2} \approx 29.95517484509146$
 $\sqrt{(14-69)^2 + (20-38.7)^2} \approx 58.09208207664793$
 Gehört also zu 1. Zentrum

5.

$$\sqrt{(22 - 10.5)^2 + (14 - 49.75)^2} \approx 37.554127602701676$$

$$\sqrt{(22 - 69)^2 + (14 - 38.7)^2} \approx 53.09510335238081$$

Gehört also zu 1. Zentrum

6.

$$\sqrt{(75 - 10.5)^2 + (17 - 49.75)^2} \approx 72.33818148115142$$

$$\sqrt{(75 - 69)^2 + (17 - 38.7)^2} \approx 22.51421773013666$$

Gehört also zu 2. Zentrum

7.

$$\sqrt{(81 - 10.5)^2 + (40 - 49.75)^2} \approx 71.17100884489413$$

$$\sqrt{(81 - 69)^2 + (40 - 38.7)^2} \approx 12.070211265756702$$

Gehört also zu 2. Zentrum

8.

$$\sqrt{(2 - 10.5)^2 + (91 - 49.75)^2} \approx 42.11665347579269$$

$$\sqrt{(2 - 69)^2 + (91 - 38.7)^2} \approx 84.99582342680138$$

Gehört also zu 1. Zentrum

9.

$$\sqrt{(96 - 10.5)^2 + (31 - 49.75)^2} \approx 87.5317799430584$$

$$\sqrt{(96 - 69)^2 + (31 - 38.7)^2} \approx 28.07650263120391$$

Gehört also zu 2. Zentrum

10.

$$\sqrt{(4 - 10.5)^2 + (31 - 49.75)^2} \approx 19.844709622466137$$

$$\sqrt{(4 - 69)^2 + (31 - 38.7)^2} \approx 65.4544880050253$$

Gehört also zu 1. Zentrum

Jetzt müssen wir die neue Zentren berechnen

1. Mittelwert von 1,4,5,8 und 10 Punkt $((\frac{22+14+22+2+4}{5}), (\frac{57+20+14+91+31}{5})) \approx (12.8, 42.6)$

$$\sqrt{(12.8 - 10.5)^2 + (42.6 - 49.75)^2} \approx 7.51 > \frac{2}{5}$$

2. Mittelwert von 2,3,6,7 und 9 Punkt

$$((\frac{79+61+75+81+96}{5}), (\frac{46+52+17+40+63}{5})) = (78.4, 43.6)$$

$$\sqrt{(78.4 - 69)^2 + (43.6 - 38.7)^2} \approx 10.6 > \frac{2}{5}$$

Also der Algorithmus geht weiter mit 3. Schritt

1.

$$\sqrt{(22 - 12.8)^2 + (57 - 42.6)^2} \approx 17.08800749063506$$

$$\sqrt{(22 - 78.4)^2 + (57 - 43.6)^2} \approx 57.969992237363634$$

Gehört also zu 1. Zentrum

2.

$$\sqrt{(79 - 12.8)^2 + (46 - 42.6)^2} \approx 66.28725367670621$$

$$\sqrt{(79 - 78.4)^2 + (46 - 43.6)^2} \approx 2.473863375370594$$

Gehört also zu 2. Zentrum

3.

$$\sqrt{(61 - 12.8)^2 + (52 - 42.6)^2} \approx 49.10804414757322$$

$\sqrt{(61 - 78.4)^2 + (52 - 43.6)^2} \approx 19.321490625725545$
 Gehört also zu 2. Zentrum

4.
 $\sqrt{(14 - 12.8)^2 + (20 - 42.6)^2} \approx 22.631835983852483$
 $\sqrt{(14 - 78.4)^2 + (20 - 43.6)^2} \approx 68.58804560563014$
 Gehört also zu 1. Zentrum

5.
 $\sqrt{(22 - 12.8)^2 + (14 - 42.6)^2} \approx 30.043302082161343$
 $\sqrt{(22 - 78.4)^2 + (14 - 43.6)^2} \approx 63.69552574553412$
 Gehört also zu 1. Zentrum

6.
 $\sqrt{(75 - 12.8)^2 + (17 - 42.6)^2} \approx 67.2621736193531$
 $\sqrt{(75 - 78.4)^2 + (17 - 43.6)^2} \approx 26.81641288464958$
 Gehört also zu 2. Zentrum

7.
 $\sqrt{(81 - 12.8)^2 + (40 - 42.6)^2} \approx 68.24954212300622$
 $\sqrt{(81 - 78.4)^2 + (40 - 43.6)^2} \approx 4.4407206622349005$
 Gehört also zu 2. Zentrum

8.
 $\sqrt{(2 - 12.8)^2 + (91 - 42.6)^2} \approx 49.59032163638384$
 $\sqrt{(2 - 78.4)^2 + (91 - 43.6)^2} \approx 89.90951006428631$
 Gehört also zu 1. Zentrum

9.
 $\sqrt{(96 - 12.8)^2 + (31 - 42.6)^2} \approx 84.00476176979492$
 $\sqrt{(96 - 78.4)^2 + (31 - 43.6)^2} \approx 21.645322820415494$
 Gehört also zu 2. Zentrum

10.
 $\sqrt{(4 - 12.8)^2 + (31 - 42.6)^2} \approx 14.560219778561038$
 $\sqrt{(4 - 78.4)^2 + (31 - 43.6)^2} \approx 75.45939305348274$
 Gehört also zu 1. Zentrum

Wir sind wieder zum gleichen Ergebniss für die Cluster wie im vorigen Schritt gekommen, also wir müssen nicht weiter ausführen.

3. Aufgabe

- a) Es gibt 5 Flüge, 3 davon erfolgreiche für das Passagier.
 Deswegen ist $P(Y) = \frac{3}{5}$ also (prior) Wahrscheinlichkeit $P(Y) = \frac{3}{5}$ dass ein Passagier überlebt.
- b) $P(X|Y) = \frac{P(X) \cap P(Y)}{P(Y)}$
 $P(X) = \frac{3}{5}$, da es 3 Frauen in 5 Flüge gibt
 $P(X) \cap P(Y) = \frac{2}{5}$, da es 2 Frauen die überlebt haben gibt in 5 Flüge
 Damit ist $P(X|Y) = \frac{P(X) \cap P(Y)}{P(Y)} = \frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2 \times 5}{5 \times 3} = \frac{2}{3}$

c) Analog zu A, $P(X) = \frac{4}{5}$

d) Vorbedingung $P(X) = \frac{3}{5}$, dass ein Passagier weniger als 200 für sein Ticket bezahlt hat
Wahrscheinlichkeit für Überleben haben wir schon berechnet, $\frac{3}{5}$

$$P(X) \cap P(Y) = \frac{1}{5}$$

Damit ist $P(Y|X) = \frac{P(Y) \cap P(X)}{P(X)} P(X) = \frac{\frac{1}{5}}{\frac{3}{5}} = \frac{1 \times 5}{5 \times 3} = \frac{1}{3}$ die Wahrscheinlichkeit, dass ein Passagier, der weniger als 200 für sein Ticket bezahlt hat, den Flug überlebt hat

4. Aufgabe

- a) Support gibt an, wie groß die Wahrscheinlichkeit ist, dass ein Set in einer Transaktion vorkommt.
- b) Konfidenz ist die Wahrscheinlichkeit, dass ein Set in einer Transaktion vorkommt, unter der Vorbedingung dass ein anderer Set in der Transaktion vorkommt.
- c) Candidate Set C_0

$$\text{Support Kaugummi} = \text{Sup(Kaugummi)} = \frac{3}{5} = 0.6$$

$$\text{Sup(Chips)} = \frac{4}{5} = 0.8$$

$$\text{Sup(Pizza)} = \frac{2}{5} = 0.4$$

$$\text{Sup(Milch)} = \frac{2}{5} = 0.4$$

$$\text{Sup(Eier)} = \frac{3}{5} = 0.6$$

$$\text{Sup(Bier)} = \frac{4}{5} = 0.8$$

$$\text{Level Set } L_0 \text{ Support Kaugummi} = \text{Sup(Kaugummi)} = \frac{3}{5} = 0.6$$

$$\text{Sup(Chips)} = \frac{4}{5} = 0.8$$

$$\text{Sup(Eier)} = \frac{3}{5} = 0.6$$

$$\text{Sup(Bier)} = \frac{4}{5} = 0.8$$

Candidate Set C_1

$$\text{Sup(k,c)} = \frac{2}{5}$$

$$\text{Sup(k,e)} = \frac{2}{5}$$

$$\text{Sup(k,b)} = \frac{2}{5}$$

$$\text{Sup(c,e)} = \frac{2}{5}$$

$$\text{Sup(c,b)} = \frac{2}{5}$$

$$\text{Sup(e,b)} = \frac{2}{5}$$

$$\text{Level Set } L_1 \text{ Sup(k,b)} = \frac{2}{5}$$

$$\text{Sup(c,b)} = \frac{2}{5}$$

$$\text{Candidate Set } C_2 \text{ Sup(k,b,c)} = \frac{2}{5}$$

Konfidenz

$$\text{conf(k} \rightarrow \text{b)} = \frac{\text{sup(k,b)}}{\text{sup(k)}} = 1 > 0.8$$

$$\text{conf(b} \rightarrow \text{k)} = \frac{\text{sup(b,k)}}{\text{sup(b)}} = \frac{2}{4} < 0.8$$

$$\text{conf}(c \rightarrow b) = \frac{\text{sup}(c,b)}{\text{sup}(c)} = \frac{3}{4} < 0.8$$

$$\text{conf}(b \rightarrow c) = \frac{\text{sup}(b,c)}{\text{sup}(b)} = \frac{3}{4} < 0.8$$

Es gibt folglich nur ein Assoziationsregel, der Support von mindestens 0,6 und Konfidenz von mindestens 0,8 hat, das ist $k \rightarrow b$