

Digital Libraries WS 2018/2019

Übungsblatt 7

Aaron Winziers - 1176638; Michael Wolz - 1195270

16. Dezember 2018

Aufgabe 1

a)

Worthäufigkeiten d_1 :

- Stadt: 1
- Land: 1
- Fluss: 1
- Trier: 1
- Mosel: 1
- $\sum_k \#Vorkommen w_k = 5$

$$P(Stadt) = P(Land) = P(Fluss) = P(Trier) = P(Mosel) = \frac{1}{5} = 0.2$$

Worthäufigkeiten d_2 :

- Fluss: 2
- Trier: 1
- $\sum_k \#Vorkommen w_k = 3$

$$P(Fluss) = \frac{2}{3} = 0.66$$

$$P(Trier) = \frac{1}{3} = 0.33$$

Worthäufigkeiten d_3 :

- Stadt: 3
- Fluss: 1

- Trier: 2
- Mosel: 1
- Leben: 1
- $\sum_k \#Vorkommen w_k = 8$

$$P(Stadt) = \frac{3}{8} = 0.375$$

$$P(Fluss) = \frac{1}{8} = 0.125$$

$$P(Trier) = \frac{2}{8} = 0.25$$

$$P(Mosel) = \frac{1}{8} = 0.125$$

$$P(Leben) = \frac{1}{8} = 0.125$$

b)

$q = Trier$

$$P(q|d_1) \propto \log 0.2 \approx -2.3$$

$$P(q|d_2) \propto \log 0.33 \approx -1.6$$

$$P(q|d_3) \propto \log 0.25 = -2$$

$q = Trier, Fluss$

$$P(q|d_1) \propto \log 0.2 + \log 0.2 \approx -3.9$$

$$P(q|d_2) \propto \log 0.33 + \log 0.66 \approx -2$$

$$P(q|d_3) \propto \log 0.25 + \log 0.125 \approx -4.1$$

c)

Worthäufigkeiten

- Stadt: 4
- Fluss: 4
- Trier: 4
- Mosel: 2
- Leben: 1
- Gesamtlänge T der Kollektion: 16

$$P_c(Stadt) = \frac{4}{8} = 0.5$$

$$P_c(Fluss) = \frac{4}{8} = 0.5$$

$$P_c(Trier) = \frac{4}{8} = 0.5$$

$$P_c(Mosel) = \frac{2}{8} = 0.25$$

$$P_c(Leben) = \frac{1}{8} = 0.125$$

d)

$$P_{LM}(Trier|d_1) = 0.5 * 0.2 + 0.5 * 0.5 = 0.35$$

$$P_{LM}(Trier|d_2) = 0.5 * 0.33 + 0.5 * 0.5 = 0.415$$

$$P_{LM}(Trier|d_3) = 0.5 * 0.25 + 0.5 * 0.5 = 0.375$$

$$P_{LM}(Fluss|d_1) = 0.5 * 0 + 0.5 * 0.5 = 0.25$$

$$P_{LM}(Fluss|d_2) = 0.5 * 0.66 + 0.5 * 0.5 = 0.58$$

$$P_{LM}(Fluss|d_3) = 0.5 * 0.125 + 0.5 * 0.5 = 0.3125$$

$q = Trier$

$$P_{LM}(q|d_1) \propto \log 0.35 \approx -1.5$$

$$P_{LM}(q|d_2) \propto \log 0.415 \approx -1.27$$

$$P_{LM}(q|d_3) \propto \log 0.375 \approx -1.42$$

$q = Trier, Fluss$

$$P_{LM}(q|d_1) \propto \log 0.35 + \log 0.25 \approx -3.5$$

$$P_{LM}(q|d_2) \propto \log 0.415 + \log 0.58 \approx -2.1$$

$$P_{LM}(q|d_3) \propto \log 0.375 + \log 0.3125 \approx -3.1$$

Aufgabe 2

a)

- $idf_t(Vater) = \log(\frac{7}{2}) = 1.8017$
- $idf_t(Mutter) = \log(\frac{7}{5}) = 0.485$
- $idf_t(Koenigin) = \log(\frac{7}{2}) = 1.807$
- $idf_t(Zwerge) = \log(\frac{7}{1}) = 2.807$
- $idf_t(Koenigstochter) = \log(\frac{7}{2}) = 1.807$

- $idf_t(Wolf) = \log(\frac{7}{2}) = 1.807$
- $idf_t(Gold) = \log(\frac{7}{3}) = 1.222$
- $idf_t(Haus) = \log(\frac{7}{7}) = 0$

$$idf_t = \begin{pmatrix} 1.802 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.485 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.807 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.807 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.807 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.807 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.222 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} tf = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 & 1 & 0 \\ 2 & 2 & 3 & 2 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 8 & 1 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & 6 \\ 2 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 5 & 1 & 3 & 4 & 1 & 1 \end{pmatrix}$$

$$idf_t \times tf = \begin{pmatrix} 0 & 9.009 & 0 & 0 & 0 & 1.802 & 0 \\ 0.970 & 0.970 & 1.455 & 0.970 & 0 & 0 & 1.455 \\ 0 & 0 & 0 & 0 & 14.456 & 1.807 & 0 \\ 0 & 0 & 0 & 0 & 11.228 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.807 & 1.807 & 0 \\ 0 & 0 & 0 & 10.842 & 0 & 0 & 10.842 \\ 2.444 & 0 & 1.222 & 0 & 1.222 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

b)

$$q^T = (0, 1, 0, 0, 0, 0, 0, 1)$$

$$\|q\| = \sqrt{2}$$

$$\|d_1\| = 2.629$$

$$\|d_2\| = 9.061$$

$$\|d_3\| = 1.900$$

$$\|d_4\| = 10.855$$

$$\|d_5\| = 18.434$$

$$\|d_6\| = 3.127$$

$$\|d_7\| = 10.939$$

$$sim(d_1, q) = \frac{0*0+0.970*1+0*0+0*0+0*0+0*0+2.444*0+0*1}{\|q\|\|d_1\|} = 0.261$$

$$sim(d_2, q) = 0.261$$

$$sim(d_3, q) = 0.391$$

$$sim(d_4, q) = 0.261$$

$$sim(d_5, q) = 0$$

$$sim(d_6, q) = 0$$

$$sim(d_7, q) = 0.391$$