

# Assignment 07

## Digital Libraries and Foundations of Information Retrieval

Winter semester 2022

1542011 Franka Brunen, 1365848 Andreas Schneider

**Task 1:**

Language models

4+4+3+4 Points

- (a)  $d_1 : P('Mosel', 'Marx', 'Porta', 'Trier', 'Dom') = \frac{1}{5} * \frac{1}{5} * \frac{1}{5} * \frac{1}{5} * \frac{1}{5}$   
 $d_2 : P('Trier', 'Porta', 'Porta') = \frac{1}{3} * \frac{2}{3} * \frac{2}{3}$   
 $d_3 : P('Trier', 'Mosel', 'Marx', 'Trier', 'Mosel', 'Porta', 'Dom', 'Mosel') = \frac{2}{8} * \frac{3}{8} * \frac{1}{8} * \frac{2}{8} * \frac{3}{8} * \frac{1}{8} * \frac{1}{8} * \frac{3}{8}$
- (b) (1)  $P(q|d_1) \propto \log(\frac{1}{5})$   
 $P(q|d_2) \propto \log(\frac{1}{3})$   
 $P(q|d_3) \propto \log(\frac{2}{8})$   
 (2)  $P(q|d_1) \propto \log(\frac{1}{5}) * \log(\frac{1}{5})$   
 $P(q|d_2) \propto \log(\frac{1}{3}) * \log(\frac{2}{3})$   
 $P(q|d_3) \propto \log(\frac{2}{8}) * \log(\frac{1}{8})$
- (c)  $P_c('Mosel') = \frac{2}{16}$   
 $P_c('Marx') = \frac{2}{16}$   
 $P_c('Porta') = \frac{4}{16}$   
 $P_c('Trier') = \frac{4}{16}$   
 $P_c('Dom') = \frac{2}{16}$
- (d)  $d_1 : P('Mosel', 'Marx', 'Porta', 'Trier', 'Dom') = \frac{1}{5} * \frac{2}{16} + \frac{1}{5} * \frac{1}{5} + \frac{1}{5} * \frac{4}{16} + \frac{1}{5} \dots$   
 $d_2 : P('Trier', 'Porta', 'Porta') = \frac{1}{3} * \frac{4}{16} + \frac{2}{3} * \frac{4}{16} + \frac{2}{3} * \frac{4}{16}$   
 $d_3 : P('Trier', 'Mosel', 'Marx', 'Trier', 'Mosel', 'Porta', 'Dom', 'Mosel') = \frac{2}{8} * \frac{4}{16} + \frac{3}{8} * \frac{2}{16} + \frac{1}{8} * \frac{2}{8} * \frac{4}{16} + \frac{3}{8} * \frac{1}{8} * \frac{1}{8} * \frac{3}{8} * \frac{2}{16} \dots$