

*Prof. Dr. Vladimir Shikhman*  
*Professur für Wirtschaftsmathematik*  
*Technische Universität Chemnitz*

*Übungsleiter: David Müller*  
*david.mueller@mathematik.tu-chemnitz.de*

## **Mathematical Foundation of Big Data Analytics (SS 2019)** **Recommendation System II**

**Ex. 1** A holiday portal hires you, in order to improve travel suggestions for their users. You have access on a database, which contains a bunch of user reviews. The relevant extracts are listed below:

- (R1) ...spent one week at Lake Garda, which is beautifully surrounded by hills. There were a lot of sport and wellness offers I tried, but most of the time I relaxed at the beach and enjoyed the sun. ...
- (R2) ...had astonishing beach holidays at the sea. Weather and hotel have been perfect. ...
- (R3) ...the only things I want to do on holidays is relaxing at the water, while catching some sun. Our balcony provided a nice view of the lake ...
- (R4) ... We had a short hiking trip to the hills and mountains of Austria. ...
- (R5) ...made a mountainbike tour through the hills, because we like the quiet environment. Fortunately we had also time for a day in the beautiful spa area, ...
- (R6) ...I love the mountains and snow. The slopes were perfectly prepared for skiing ...

- a) Build a term-document matrix out of the reviews, where the relevant terms are lake, sea, beach, hills, mountains.
- b) Measure similarity of users via cosine measure.  
Recall that the cosine similarity measure of two vectors  $x, y \in \mathbb{R}^n$ , is given by

$$\frac{x^T y}{\|x\| \|y\|}.$$

Are the results satisfactory?

- c) Calculate the best rank 2 approximation of the term-document matrix.
- d) Compare the cosine similarities based on c) with the original ones and interpret the results.