

Exercise 2

Sprachverarbeitung (VL + Ü)

Nils Reiter, nils.reiter@uni-koeln.de

April 11, 2023 (Summer term 2023)

Our project Gutenberg dump contains two editions of Doyle's "The Valley of Fear". We want to study how they differ (if they differ).

- Find out their id numbers.
- Extract their word frequencies.
- Inspect and compare them (manually). Do you think it's the same text?