

# Web Intelligence

## EXERCISE FOR LECTURE 1

Implement a function, which takes two text strings as arguments and determines if they are near-duplicates. Use what you have learned about shinkling and implement the *Jaccard similarity*. You may also use sketches, but be aware that a good implementation will then likely take you more than the 2 + 4 hours that you should use on this exercise. What is the difference between these two implementations?

The near-duplicate function will come in handy for the mini-project that you will be starting on next week, so remember to test it! You are allowed to use any programming language that you prefer.