# Homework 2:

* Attached to this document there are the files *dataset\_x.txt*, (x=1,10,25) containing each x\*100 news articles. Write a Python program that, using shingles and minhashing tecniques, detects whether two articles are similar or not.
* Use as hash functions *h(x)=(a\*x+b) mod c* where *a*  and *b* are randomly chosen integers and *c* is a prime number greater (not too much) than the maximum value of *x*.
* For now do not use the LSH technique