

You need to implement plagiarism system. For this purpose please use [this corpus](#) it's quite small, but it quite representative for this task. And yes - you need to use knowledge obtained from this course e.g. lsh, minhashing etc. Output should be in next format: document id - plagiarized document id