# Text Technologies for Data Science Assignment 2

s1140740

## 1 Introduction

The purpose of this report is to analyse performance of algorithms used to compute similarities between documents.

## 2 Algorithm `brute.py`

This algorithm computes similarities between documents naively in $O(n^2)$ time, where $n$ is the total number of documents. For every document it computes the similarity with every document we have already seen. It uses cosine similarity function with tf.idf weights.

## 3 Algorithm `index.py`

This algorithm uses an inverted index to compute similarities between documents. It does so in $O(n^2)$ as well.

## 4 Algorithm `best.py`

## 5 Conclusion