

# Assignment 7

Huffman Coding

DESIGN Document

## Description of the program

---

For your assignment, you will be creating a program that attempts to identify the most likely authors for an anonymous sample of text given a large database of texts with known authors. Modern-day stylometry usually is performed using machine learning, achieving high identification accuracies by learning over time, but implementing this from scratch would take an extraordinary effort. We instead will settle and be content with using an algorithm that's commonly used in machine learning to identify authors of anonymous samples of text, albeit less accurately. What is this algorithm? It is the k-nearest neighbors algorithm.

## Files to be included

---

### Source and header files:

1. **bf.h:**
2. **bf.c: Contains the implementation of the Bloom filter ADT.**
3. **Bv.h**
4. **bv.c**

### Additional Files:

1. **Makefile:** formats all source code, including the header files.
2. **README.md:** Description of how to use my program and Makefile. It also includes any command-line option that my program accepts. Any false positives reported by scan-build should be documented and explained here as well. Note down any known bugs or errors in this file as well for the graders.

3. **DESIGN.pdf (*This file*):** The design document describes the preliminary design and design process for my program with sufficient detail for potential replication
4. **WRITEUP.pdf:** Analysis and description of the produced program, as well as graphs generated and/or information, gathered from outputs

Pseudocode/Structure: