

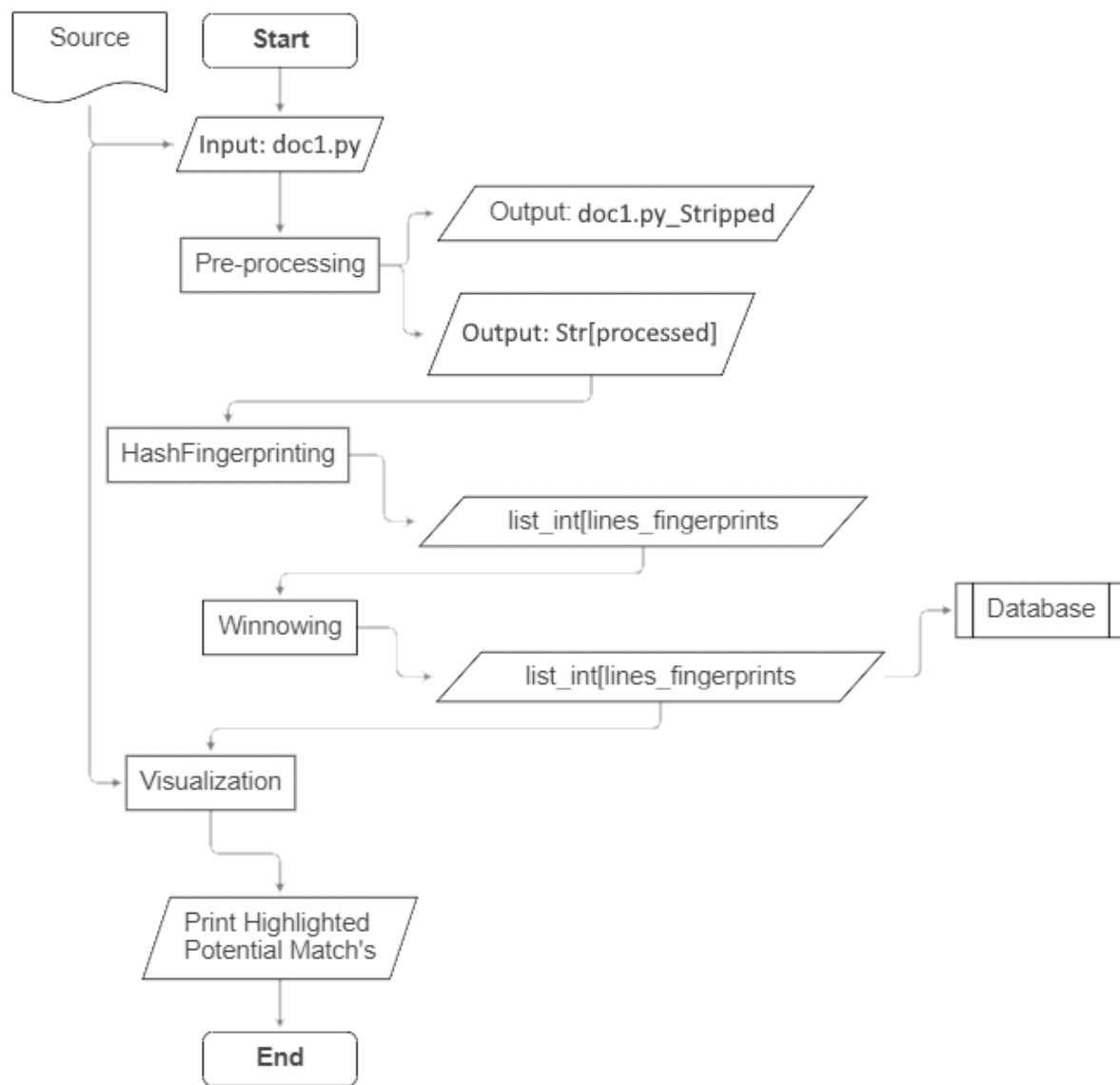
Preprocessing - Chase Jamieson

Hashing/Fingerprinting & Visual - Tracy Hotchkiss & Vinh Duong

Winnowing - Trevor Holland

Pipeline of Plagiarism Detector

Pipeline Interface



Sample of Visualization

input file compared to doc1 is:

```
def numberFunction():
```

```
    a = 1 + 2
```

```
    b = a + a
```

```
    c = b + 1
```

```
    return c
```

Percentage matched is: 14.925373134328357

Contracts

Preprocessing	
Input: source file	Example1 ('databaseFile1.py')
Output: stripped source file	Example2 (databaseFile1.py_Stripped)
Output: processed source file	Example3 (Str[processed])

Example1

```
def numberFunction():  
    a = 1 + 2  
    b = a + a  
    c = b + 1  
    return c  
  
def stringFunction(input1, input2):  
    string1 = "abcd"  
    string2 = "cda"  
    string3 = string1 + string2  
    return string2  
  
def doesNothing():  
    a = 1  
    b = "b"  
    c = True  
  
def arrayAddition(a):  
    a = a + [1]  
  
numberFunction()  
stringFunction("a", "b")  
doesNothing()  
arrayAddition([4, 3, 2])
```

Example2

```
def numberFunction():  
    a = 1 + 2  
    b = a + a  
    c = b + 1  
    return c  
  
def stringFunction(input1, input2):  
    string1 = "abcd"  
    string2 = "cda"  
    string3 = string1 + string2  
    return string2  
  
def doesNothing():  
    a = 1  
    b = "b"  
    c = True  
  
def arrayAddition(a):  
    a = a + [1]  
  
numberFunction()  
stringFunction("a", "b")  
doesNothing()  
arrayAddition([4, 3, 2])
```

Example3

```
def fun1():  
    var1 = 1 + 2  
    var2 = var1 + var1  
    var3 = var2 + 1  
    return var3  
  
def fun2(var4, input2):  
    var5 = "abcd"  
    var6 = "cda"  
    var7 = var5 + var6  
    return var6  
  
def fun3():  
    var1 = 1  
    var2 = "var2"  
    var3 = True  
  
def fun4(var1):  
    var1 = var1 + [1]  
  
fun1()  
fun2("var1", "var2")  
fun3()  
fun4([4, 3, 2])
```

Contracts

HashFingerprinting	
Input: processed source file	Example3 ('str[processed]')
Output: fingerprints source file	Example4 (list_int[lines_fingerprints])

Example4

```
[([1], 1440), ([1], 1382), ([1], 1265), ([1], 1008), ([1], 1553), ([1], 1514),
```

Contracts

Winnowing	
Input: fingerprints source file	Example4 (list_int[lines_fingerprints])
Output: list lines and fingerprints	Example5 (list_int[lines_fingerprints])

Example5

```
[([1], 1440), ([1], 1382), ([1], 1265), ([1], 1008), ([1], 692), ([1, 2], 632),
```

Contracts

Visual	
Input: source file	Example1 ('databaseFile1.py')
Input: list lines and fingerprints	Example5 (list_int[lines_fingerprints])
Output: Print Highlighted Potential Matches with their match percentage	Example6

Sample of Example6

```
input file compared to doc1 is:
```

```
def numberFunction():
```

```
    a = 1 + 2
```

```
    b = a + a
```

```
    c = b + 1
```

```
    return c
```

```
Percentage matched is: 14.925373134328357
```