

```
1 C:\Users\User\Desktop\ergasia\.venv\Scripts\python.exe C:\Users\User\Desktop\ergasia\apomakrismo_test.py
```

```
2
3 #####
4 Running experiments for K = 5
5 #####
```

```
6
7 Loading and Preprocessing Data (Brute Force)...
8 Creating Shingles (Brute Force)...
9 Starting Brute Force Jaccard computation...
10 Finished Brute Force Jaccard computation.
11 Brute Force total time: 9394.14 sec
12
13
```

```
14 Loading and preprocessing data (LSH) with threshold=0.5, num_perm=64...
15 Generating MinHash signatures (LSH)...
16 Building LSH index...
17 Finding nearest neighbors...
18 Nearest neighbor search completed.
19
```

```
20 Final Results Table:
```

	Type	BuildTime	QueryTime	TotalTime	fraction of the true K most similar documents that are reported by
	LSH method as well			Parameters	
23	Brute-Force-Jaccard	0.00	9394.139493	9394.139493	100% K=5
24	LSH-Jaccard	3425.46	3.800000	3429.260000	7% K=5, Perm=64, threshold=0.5

```
25
26
27 #####
28 Running experiments for K = 7
29 #####
```

```
30
31 Loading and Preprocessing Data (Brute Force)...
32 Creating Shingles (Brute Force)...
33 Starting Brute Force Jaccard computation...
34 Finished Brute Force Jaccard computation.
35 Brute Force total time: 9362.30 sec
36
```

```

37
38 Loading and preprocessing data (LSH) with threshold=0.5, num_perm=64...
39 Generating MinHash signatures (LSH)...
40 Building LSH index...
41 Finding nearest neighbors...
42 Nearest neighbor search completed.
43
44 Final Results Table:
45
46      Type      BuildTime      QueryTime      TotalTime      fraction of the true K most similar documents that are reported by LSH
47 method as well
48 Brute-Force-Jaccard      0.00      9362.29586      9362.29586      100%
49
50 LSH-Jaccard      3420.37      3.79000      3424.15000      6% K=7, Perm=64, threshold=0.5
51 #####
52 Running experiments for K = 9
53 #####
54
55 Loading and Preprocessing Data (Brute Force)...
56 Creating Shingles (Brute Force)...
57 Starting Brute Force Jaccard computation...
58 Finished Brute Force Jaccard computation.
59 Brute Force total time: 9362.73 sec
60
61
62 Loading and preprocessing data (LSH) with threshold=0.5, num_perm=64...
63 Generating MinHash signatures (LSH)...
64 Building LSH index...
65 Finding nearest neighbors...
66 Nearest neighbor search completed.
67
68 Final Results Table:
69
70      Type      BuildTime      QueryTime      TotalTime      fraction of the true K most similar documents that are reported by
71 LSH method as well      Parameters
72 Brute-Force-Jaccard      0.00      9362.730903      9362.730903

```

[illegible]