Near-duplicate page detection reports:

My code for testing it: (you can find it on `LSH.py`)

Code:

```
if __name__ == '__main__':
    docs = []
    title_to_id = {}
    with open(project_root+'/Logic/core/LSHFakeData.json') as f:
        docs = json.load(f)
    summaries = []
    for doc in docs:
        temp = doc['summaries']
        summary = ''
        for t in temp:
            summary += ' ' + t
        summaries.append(summary)
        title_to_id[len(summaries) - 1] = doc['title']
    docs = []
    with open(project_root+'/data/IMDB_Crawled.json') as f:
        docs = json.load(f)
    for doc in docs:
        temp = doc['summaries']
        summary = ''
        for t in temp:
            summary += ' ' + t
        if summary == '':
            continue
        summaries.append(summary)
        title_to_id[len(summaries) - 1] = doc['title']

    num_hashes = 625
    min_hash_lsh = MinHashLSH(summaries, num_hashes)
    buckets = min_hash_lsh.perform_lsh()

    print_buckets = []
    print("Buckets:")
    for bucket_id, bucket in buckets.items():
        if len(bucket) > 1 and bucket not in print_buckets:
            print_buckets.append(bucket)
    print_buckets.sort(key=lambda x: x[0])
    for i, bucket in enumerate(print_buckets):
        print(f"Bucket {i+1}:\t", end='')
        for j, doc_idx in enumerate(bucket):
            print(f'{title_to_id[doc_idx]}(index={doc_idx})', end=' ')
            if j != len(bucket) - 1:
                print('- ', end='')
        print()

    min_hash_lsh.jaccard_similarity_test(buckets, summaries)
```

Outputs:

```
Buckets:
Bucket 1:       test1(index=0) - test2(index=1)
Bucket 2:       test7(index=6) - test8(index=7)
Bucket 3:       test13(index=12) - test14(index=13)
Bucket 4:       test15(index=14) - test16(index=15)
Bucket 5:       test17(index=16) - test18(index=17)
Bucket 6:       test19(index=18) - test20(index=19)
Bucket 7:       Apocalypse Now(index=75) - The Post(index=1273)
Bucket 8:       M(index=120) - Wild Strawberries(index=221)
Bucket 9:       Opening Night(index=791) - The Aura(index=1484)
Bucket 10:      The Batman Part II(index=1252) - We Live in
Time(index=2149)
Bucket 11:      The Hunchback of Notre Dame(index=2118) - Return to
Oz(index=2330)
```