BOOK MANAGER – FINAL PROJECT

Probabilistic Methods for Computer Engineering
2015-2016

Ana Tavares 76629
Diogo Ferreira 76425

BLOOM FILTER

BLOOMFILTER.M

Process of Creation and Insertion

- Initialize
- Insert (calls Insert.m)

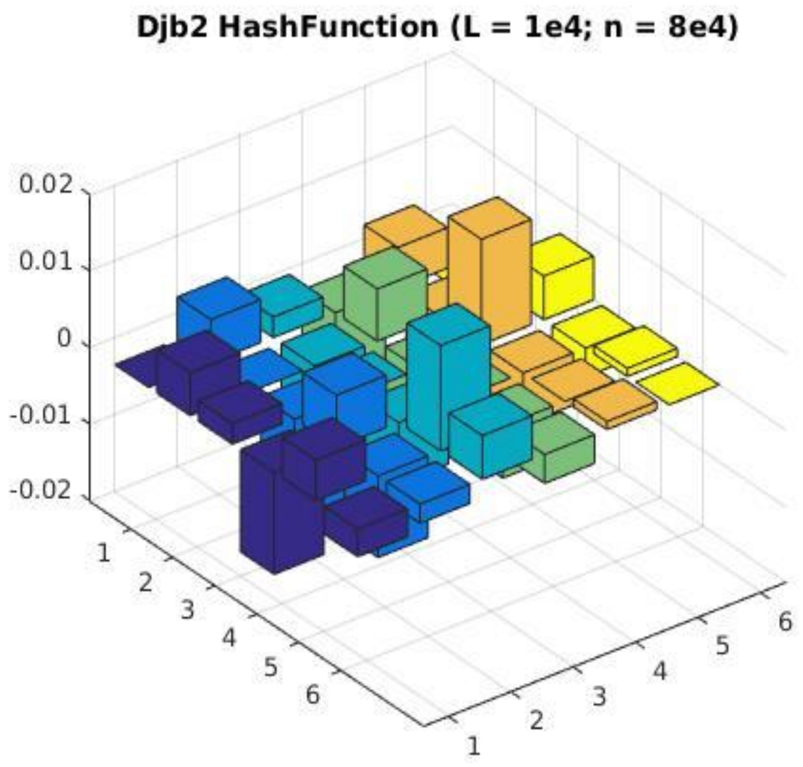
ISMEMBER.M

Assesses whether an element belongs to a set through the values returned by

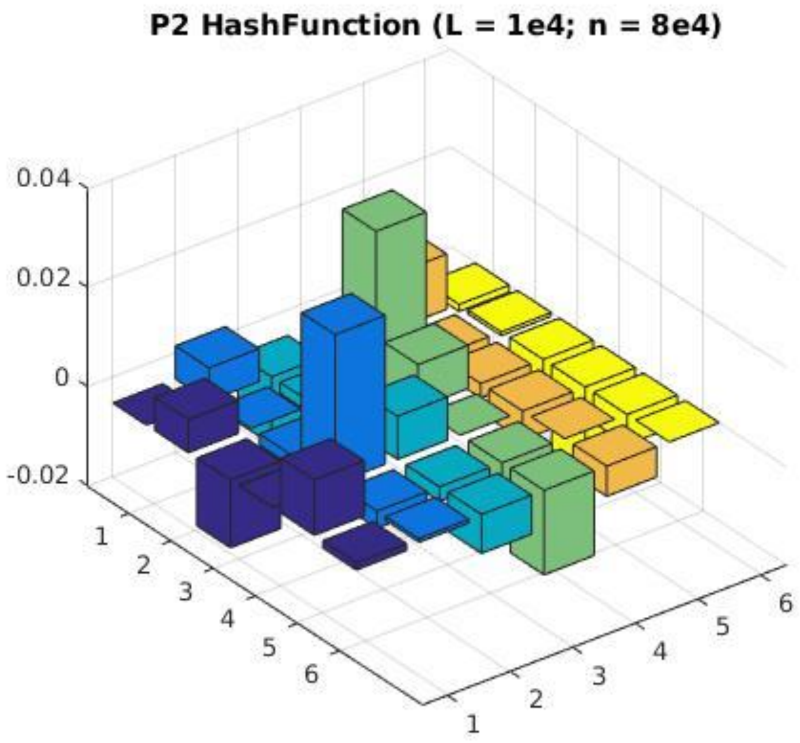
Hash Function

HASH FUNCTIONS

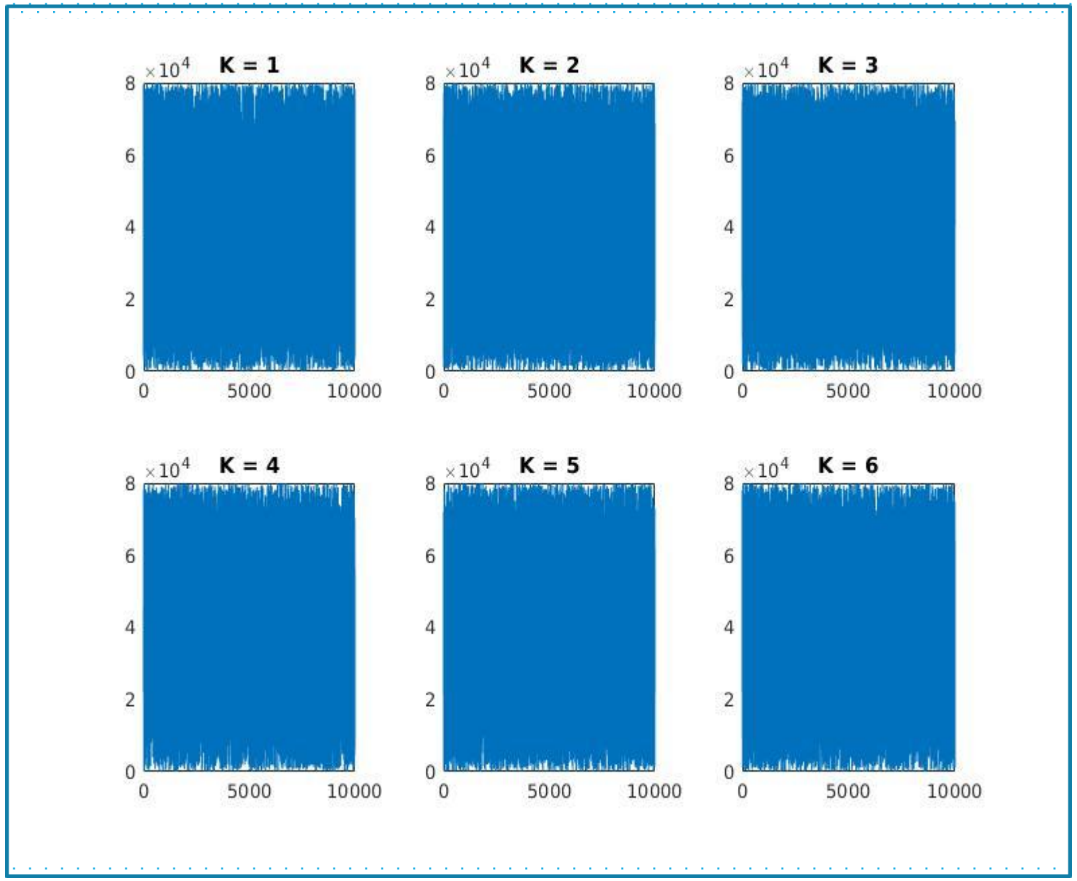
Dbj2



P2



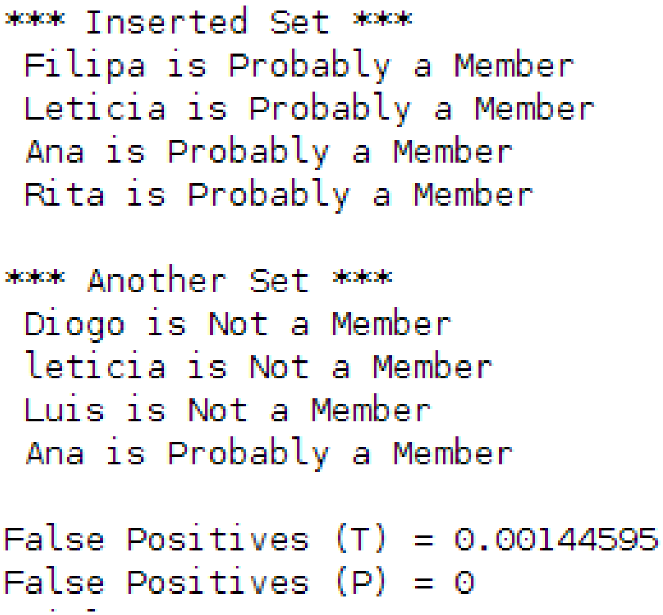
DBJ2



- Created by Dan Bernstein
- Mystery behind the constants (33 and 5318)
  - Similarities with Linear Congruential Generator(LCG)

TESTS

BasicTest.m

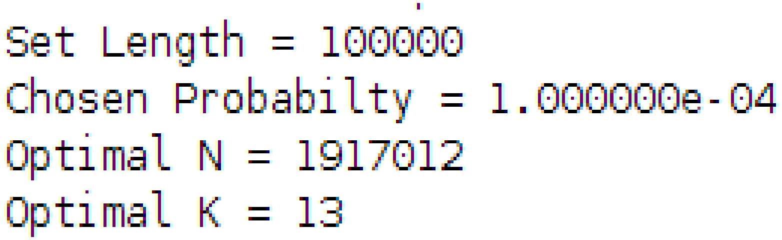


```
*** Inserted Set ***
 Filipa is Probably a Member
 Leticia is Probably a Member
 Ana is Probably a Member
 Rita is Probably a Member

*** Another Set ***
 Diogo is Not a Member
 leticia is Not a Member
 Luis is Not a Member
 Ana is Probably a Member

False Positives (T) = 0.00144595
False Positives (P) = 0
```

TestWithOptimalN.m



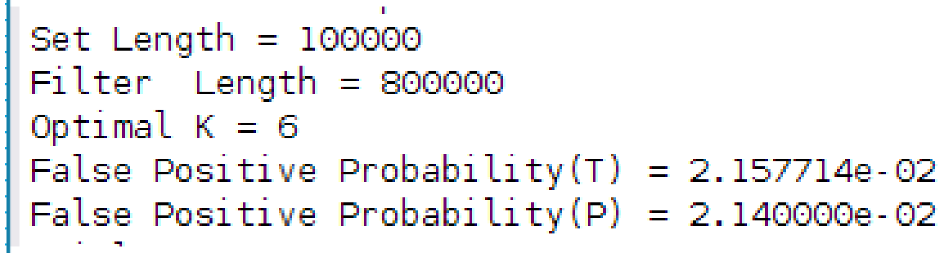
```
Set Length = 100000
Chosen Probabilty = 1.000000e-04
Optimal N = 1917012
Optimal K = 13
```

$$n = \frac{L.\log(\frac{1}{p})}{\log(2)^2} \quad ;$$

n = Bloom Filter Length
L = Set Length
p = False Positive Probability

TESTS

TestWithOptimalK.m



```
Set Length = 100000
Filter  Length = 800000
Optimal K = 6
False Positive Probability(T) = 2.157714e-02
False Positive Probability(P) = 2.140000e-02
```

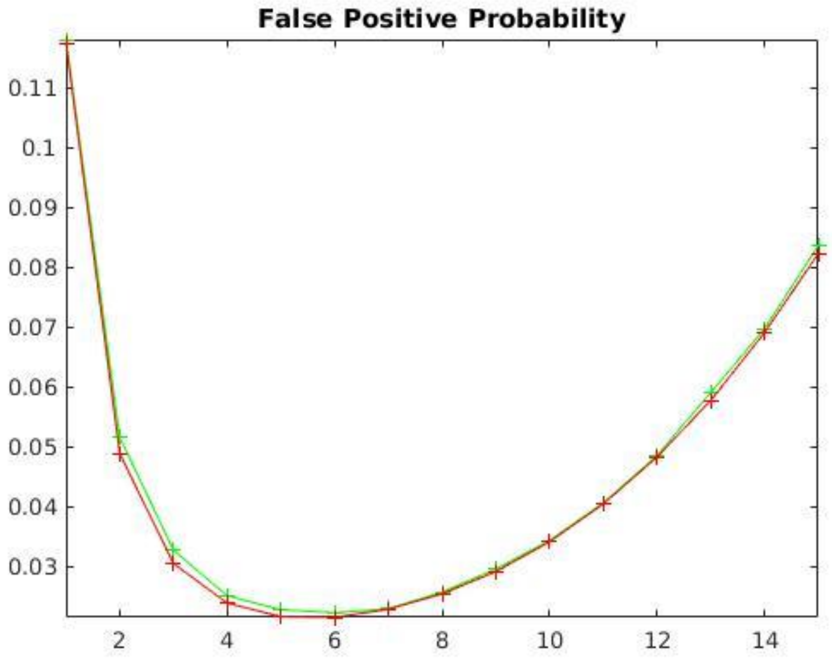
$$k = \frac{n}{L}.\ln 2 \quad ;$$

n = Bloom Filter Length

L = Set Length

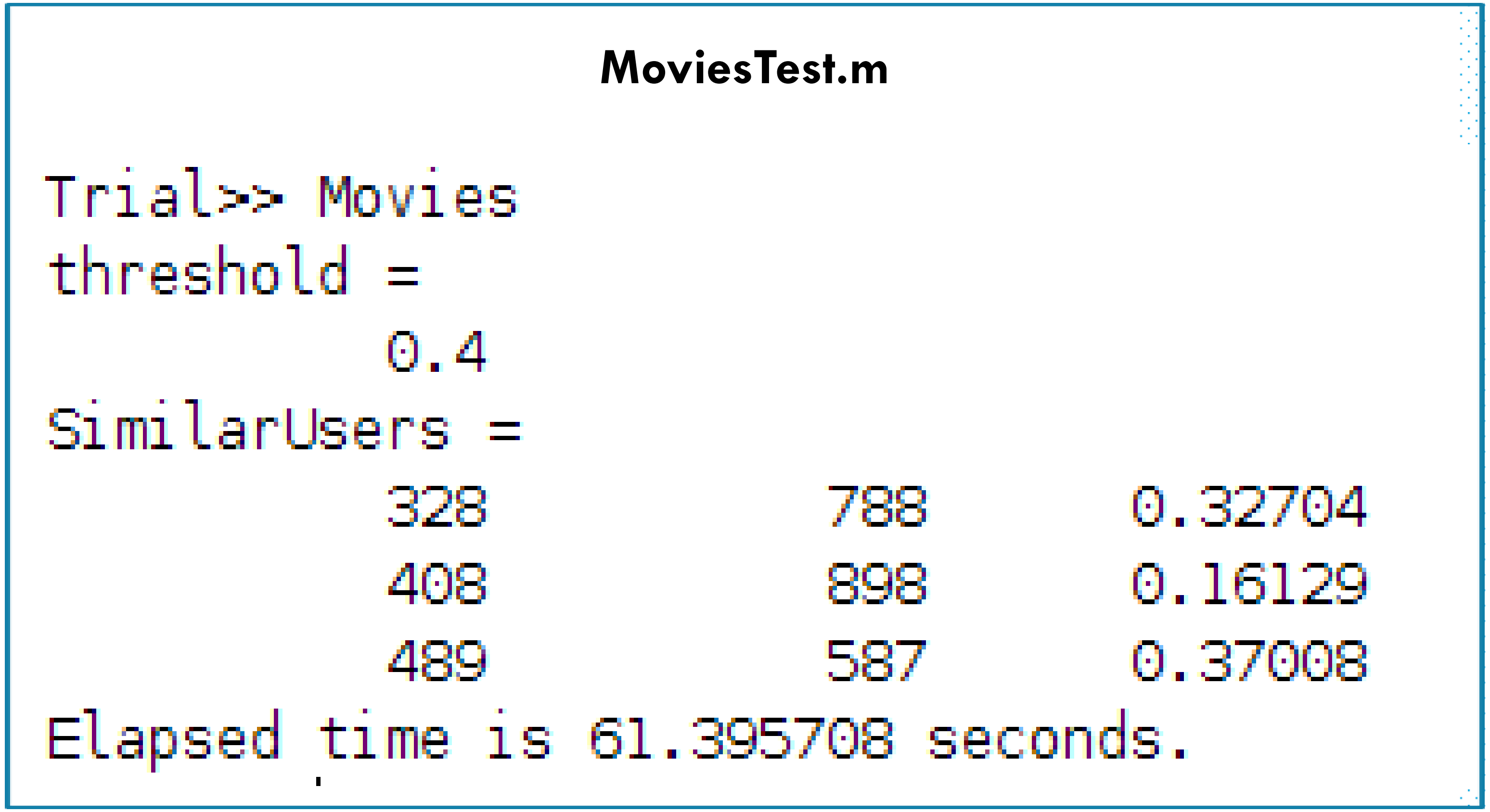
k = Hash Functions Number

RandomStringTest.m



JACCARD SIMILARITIES

TEST – JACCARD DISTANCE



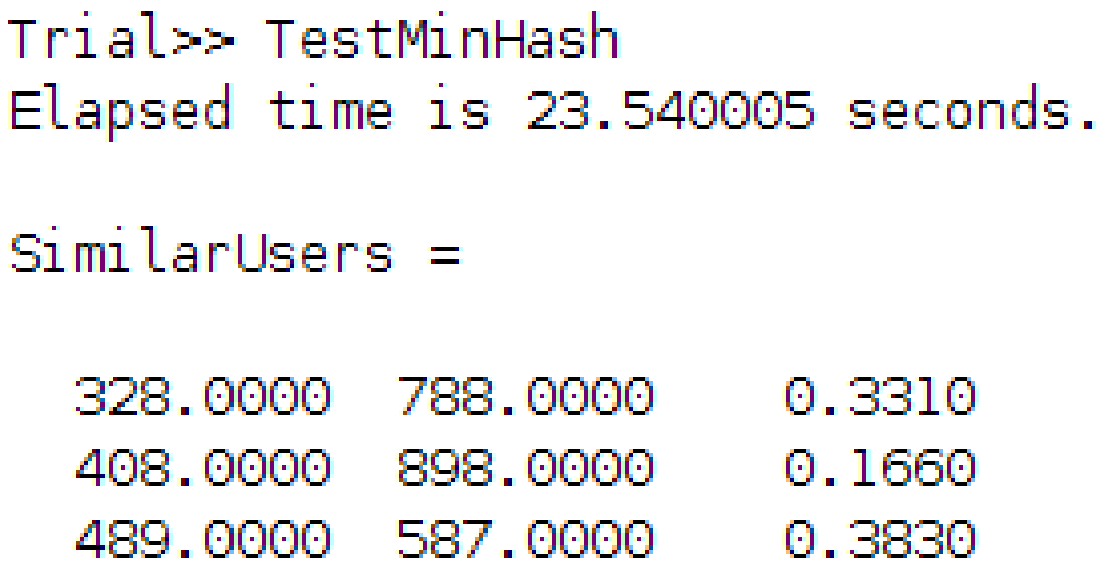
MoviesTest.m

```
Trial>> Movies
threshold =
         0.4
SimilarUsers =
         328           788        0.32704
         408           898        0.16129
         489           587        0.37008
Elapsed time is 61.395708 seconds.
```

MINHASH

TESTS

TestMinHash.m



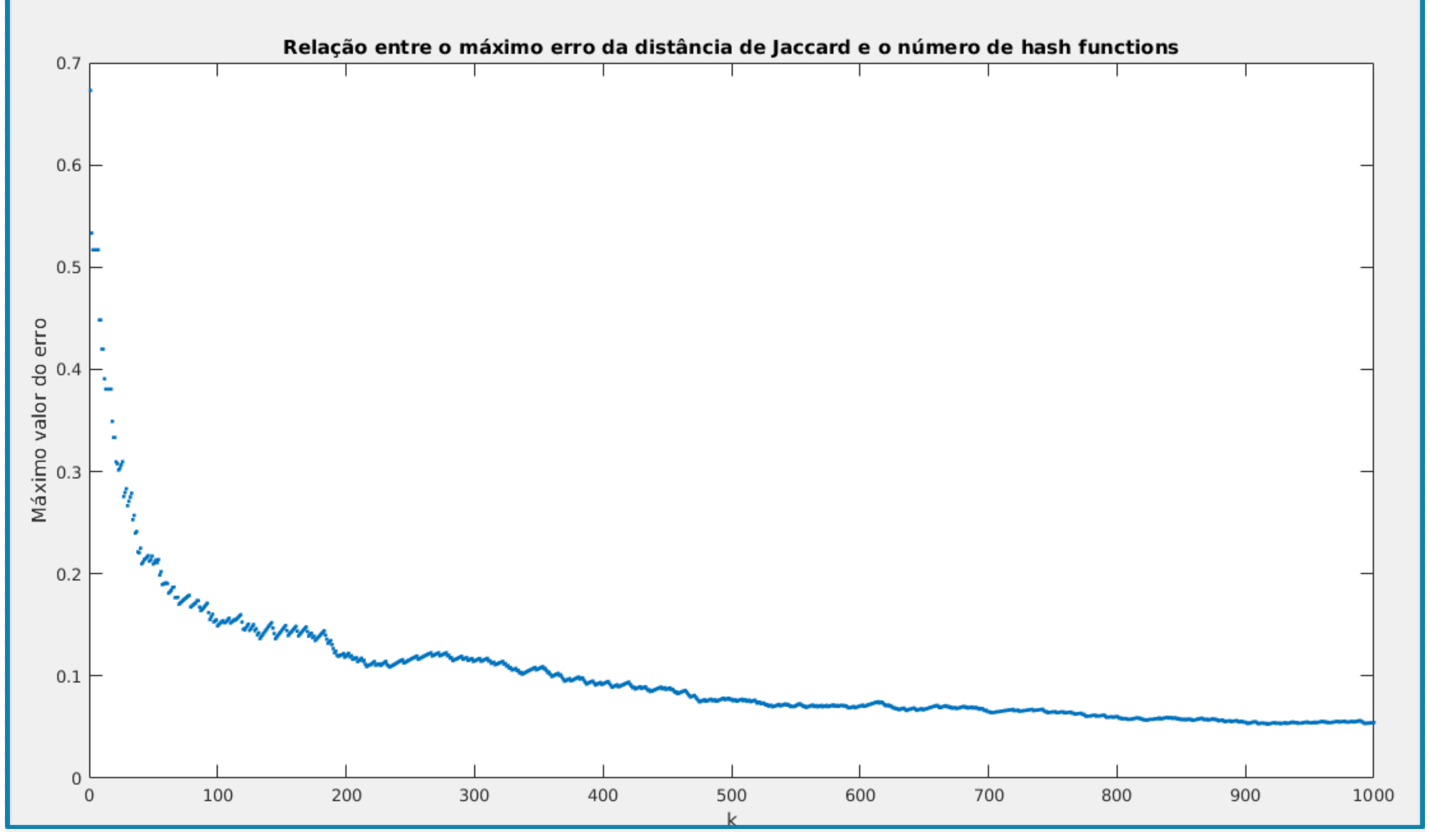
```
Trial>> TestMinHash
Elapsed time is 23.540005 seconds.

SimilarUsers =

  328.0000  788.0000    0.3310
  408.0000  898.0000    0.1660
  489.0000  587.0000    0.3830
```

TESTS

FindKMinhash.m



APPLICATION

DATASET

Book Crossing Freiburg University

- BX-Users
- BX-Books
- BX-Books-Ratings

READ AND PROCESS DATA

- ReadData.m
- usersBooks.m

BLOOM FILTER

- Evaluate the existence or the nonexistence of a book

MINHASH

- Search Users (just 1k) with equal or similar ratings (Threshold = 0,5)

GRAPHIC USER INTERFACE

- Search Bar
- Book Names List
- User ID List
- Similar Users Window

LET’S TRY?