

University of São Paulo  
Institute of Mathematics and Statistics  
Bachelor's in Computer Science

Breno Helfstein Moura

# Hashing Functions and Hash Tables

## A practical approach

São Paulo  
November de 2019

# Hashing Functions and Hash Tables

## A practical approach

Final undergraduate thesis for subject  
MAC0499 – Trabalho de Formatura Supervisionado.

Supervisor: Prof. Dr. Jose Coelho de Pina Junior  
[ Cosupervisor: Prof. Dr. Nina S. T. Hirata ]

São Paulo  
November de 2019

# Abstract

During this undergraduate thesis I will explain and learn more about one of the most fascinating and used idea in Computer Science, Hashing and Hash tables. I divided this thesis in three main parts:

- Hash Functions
- Hash Tables
- Applications

**TODO:** Remember to explain multiplicative hashing

During the first part I explain why hashing functions are an important idea in Computer Science, summarize some of the ideas Donald Knuth present on his book (The Art of Computer programming, Vol. 3) and use some metrics to evaluate what is a good hash function.

**TODO:** Replace collision resolution with open addressing.

During the second part I will explain one the most used data structures in computer programming, hash tables. I will explain what constitute a hash table, show some of the classic implementations of this data structure and explain some of the most used collision-resolution strategies. It is nice to observe here that although hash tables is a simple concept, there is still debates regarding this subject with no clear consensus on what is a state of the art hash table.

**TODO:** It would be nice to explain consistent hashing

During the third, and last, part I will explain some application of hash functions in computer science problems. I will explain Rabin-Karp, a string search algorithm that uses hashing and a solution to identify isomorphisms on trees using hashing functions.

I hope this is as fun to read for you as it was for me to write!

**Keywords:** Hash functions, Hash Tables, collision-resolution, open-addressing.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Hash Functions . . . . .	1
<b>2</b>	<b>Hash Functions</b>	<b>3</b>
<b>3</b>	<b>Conclusions</b>	<b>5</b>
<b>A</b>	<b>Apendix</b>	<b>7</b>
	<b>Bibliography</b>	<b>9</b>



# Chapter 1

## Introduction

Hashing and Hash functions are among the most classic topics within computer science, yet is still one of the topics with most debate about what is state of the art. While the hash table was invented in 1953, widely discussed by Donald Knuth in his famous book *The Art of Computer Programming* (Vol.3, Chapter 6.4), there are still many tweaks that can be made to boost its performance for specific use cases. One great example is F14, an open-source memory efficient hash table by Facebook <sup>1</sup>.

An example of lack of consensus in this area are the different hash functions and hash table implementations in different languages. There is no clear consensus on how to decide the size of a hash table, what are the tradeoffs of the collision-resolution algorithms or even what defines a good hashing function. Hopefully, we got years of research on the topic to study and present a view on the subject, and that is what I am going to do throughout this undergraduate thesis.

During this introduction I will give a brief idea of what are hash functions, hash tables and why they are interesting. This will also give you a grasp of what you can expect to read about in each chapter.

### 1.1 Hash Functions

Hash functions are functions that can be used to map data of an arbitrary size to data of a fixed size [Wikipedia \(2019\)](#)

---

<sup>1</sup>F14 is open sourced: <https://engineering.fb.com/developer-tools/f14/>





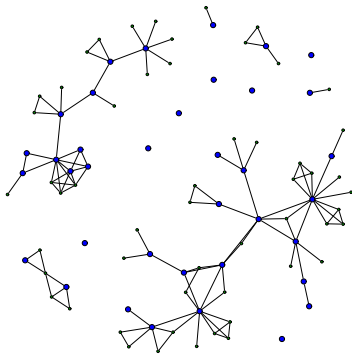
# Chapter 2

## Hash Functions

Embora neste exemplo tenhamos apenas um capítulo, entre a introdução e a conclusão de uma monografia podemos ter uma sequência de capítulos descrevendo o trabalho e os resultados. Estes podem descrever fundamentos, trabalhos relacionados, método/modelo/algoritmo proposto, experimentos realizados, resultados obtidos.

Cada capítulo pode ser organizado em seções, que por sua vez pode conter subseções.

Um exemplo de figura está na figura 2.1.



**Figure 2.1:** *Exemplo de uma figura.*



# Chapter 3

## Conclusions

Texto texto<sup>1</sup>.

---

<sup>1</sup>Exemplo de referência para página Web: [www.vision.ime.usp.br/~jmena/stuff/tese-exemplo](http://www.vision.ime.usp.br/~jmena/stuff/tese-exemplo)



## Appendix A

# Appendix

[illegible]



# Bibliography

**Wikipedia(2019)** Wikipedia. Hash function, 2019. URL [https://en.wikipedia.org/wiki/Hash\\_function](https://en.wikipedia.org/wiki/Hash_function). Citado na pág. [1](#)