

A Crash Course on Data Compression

0. Welcome

Giulio Ermanno Pibiri

ISTI-CNR, giulio.ermanno.pibiri@isti.cnr.it



@giulio_pibiri



@jermmp

What is this course?

- This is an **introductory course on Data Compression**.
- A blend of Algorithms, Data Structures, and Programming (C++).
- 5 Modules + 3 Lab Sessions (approx. 20h):
 1. **Introduction**
 2. **Integer Codes**
 3. **List Compressors**
 4. **Statistical Compressors**
 5. **Dictionary-based Compressors**



[Course Web page](#)

Resources

- Slides/code at the course Web page.
- Books:
 1. Robert Sedgewick and Kevin Wayne. 2011. *Algorithms*. Four-th edition. Addison-Wesley Professional, ISBN 0-321-57351-X
 2. David Salomon. 2007. *Variable-Length Codes for Data Compression*. Springer Science & Business Media, ISBN 978-1-84628-959-0.
 3. Alistair Moffat and Andrew Turpin. 2002. *Compression and Coding Algorithms*. Springer Science & Business Media, ISBN 978-1-4615-0935-6.
 4. Gonzalo Navarro. 2016. *Compact Data Structures*. Cambridge University Press, ISBN 978-1-107-15238-0.
- Survey papers:
 1. G. E. P. and Rossano Venturini. 2020. *Techniques for Inverted Index Compression*. ACM Computing Surveys. 53, 6, Article 125 (November 2021), 36 pages. <https://doi.org/10.1145/3415148>
 2. Alistair Moffat. 2019. *Huffman Coding*. ACM Computing Surveys. 52, 4, Article 85 (July 2020), 35 pages. <https://doi.org/10.1145/3342555>

Prerequisites

- Math and CS: computational complexity (Big-Oh notation), recursion, basic algebra, elementary data structures (e.g., arrays, trees, and lists).
- Programming: built-in types, loops, functions, arrays, objects, logical and bit-wise operators.
- Computer with C++ environment (compiler, STL, text editor + terminal).

- To start with C++:



[tutorialspoint](https://www.tutorialspoint.com/cplusplus/)



[cplusplus](https://en.cppreference.com/)

Exam Modality

- Your choice:
 1. Implement (correctly!) a data compression algorithm and discuss it.
 2. Study a research paper and discuss it.
 3. Contribute to the course: prepare some slides and submit a pull request.
- To be agreed with me.

Room and Schedule

- When:
11-15/04/2022, 14:00-18:00
- Where:
Aula Riunioni del Dipartimento di Ingegneria dell'Informazione
via G. Caruso 16, Pisa