



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Do we need data structures? The top data structure you should know for next generation computing

1920-87R-03

Aims

- *The course presents the most recent approaches to the design of data structures for dealing with the challenges of modern computing on big collections of text data. The goal of the course is the acquisition of techniques and concepts that allow to face emerging challenges in the field of computer science.*
- *Topics include:*
 - *storing and querying by succinct data structures and how to design efficient algorithms for building such succinct data structures,*
 - *Modern applications.*
 - *Data streaming and algorithms for dealing with text data*

Contents

Some of the lectures are given by experts in the field and there will be a seminar from Travis Gagie in the second week of March (9th of March).

The main lectures will be in the third and last week of March.

Introduction

- A list of novel data structures and their uses in different applications
- What is indexing of big data and what are the main queries? What is data streaming?
- Introduction to Bloom Filters and their applications to manage big data in small space

Tecniche di hashing con applicazioni al WEB

Web crawling and indexing

Web graphs

(Paolo Boldi <http://boldi.di.unimi.it/>)

MPHF (minimum perfect hashing function)

- First part: (M)PHF
 - ? The general problem (intended application: graph construction)
 - ? Interlude: hash functions
 - ? The MWHC algorithm for sparse arrays
 - ? The MWHC to build (M)PHF
- Second part: MMPHF
 - ? Introduction
 - ? First construction: LCPs
 - ? Second construction: Z-fast tries

Data structures for pattern matching (24th of March)

Suffix-trees and suffix-arrays

(Gregory Kucherov, docente del collegio di dottorato

<https://www-igm.univ-mlv.fr/~koutcher>

Detailed program

Prerequisites

Teaching form

Textbook and teaching resource

Slides from the lectures + some references to selected papers

Semester

Assessment method

It is requested to give a presentation on a topic presented during the course

Office hours
