UniPi past and ongoing activity round-up
[Tasks T₁, T₂, T₃, T₄]

~

Paolo FERRAGINA
Two new PhD students

Antonio Boffa
PhD Student
LinkedIn · Website

Francesco Tosoni
PhD Student
LinkedIn
Our Theory Achievements


Submitted to *Theoretical Computer Science*


Submitted to *Theoretical Computer Science*

Our Algorithm Engineering Achievements

Compressed and Learned Data Structures

Software & Datasets

**PGM-index**
An optimal learned data structure that enables fast point and range searches in arrays of billions of items using orders of magnitude less space than traditional indexes.

[GitHub](#) • [Website](#)

**LA-vector**
Compressed learned bitvector supporting efficient rank and select queries.

[GitHub](#)

Pizza & Chili
Datasets for compressed indexes and test collections bechmarking

[GitHub](#) • [Website](#)

FM-index v2
A full-text index data structure that combines compression and indexing by encapsulating in a single compressed file both the original file plus some indexing information.

[Website](#)
New features:

- Python API
- External memory support
- Multi-dimensional indexing

The Piecewise Geometric Model index (PGM-index) is a data structure that enables fast lookup, predecessor, range searches and updates in arrays of billions of items using orders of magnitude less space than traditional indexes while providing the same worst-case query time guarantees.
PGM featured on Hacker News and social media

PGM Indexes: Learned indexes that match B-tree performance with 83x less space (unipi.it)

583 points by hbrundage 45 days ago | hide | past | favorite | 123 comments

Mark Papadakis
@markpapadakis
pgm.di.unipi.it very excited about this. If the claims are true, this is going to be very disruptive and "game changing", even.
8:52 AM - Jan 25, 2021

Kelly Sommers
@kellabyte
Piecewise Geometric Model index (PGM-index) is a data structure enables fast lookup, predecessor, range searches & updates in billions using orders of magnitude less space than traditional indexes while providing the same worst-case query time guarantees.
pgm.di.unipi.it

Peter Zaitsev
@PeterZaitsev
The PGM-index - New smaller faster index structure for databases pgm.di.unipi.it
2:06 PM - Jan 26, 2021

Miguel Ángel Pa...
@miguelinias3
The Piecewise Geometric Model index (PGM-index) looks extremely nice pgm.di.unipi.it On the page you can find links to the paper and a presentation.
11:34 AM - Jan 25, 2021

Ian Maurer
@imaurer
New Find: PyGM is a Python library that enables fast query operations on sorted number lists w/ a tiny memory overhead. Internally, it uses PGM-index, a state-of-the-art learned data structure that scales to billions of elements in a few tens of MBs.

Stefan Keller
@sfbilller
Piecewise Geometric Model index (PGM) for databases: Plugins validate and benchmark this promising new database index and on a real #DBMS like #PostgreSQL. @Jcc Paolo Ferragina and Giorgio Vinciiguerra @Unipi

Database Intern...
@thereal databass
Fascinating! Check it out, another learned index data structure; this time without neural nets, The PGM-index. Presentation:
youtube.com/watch?v=gCkJ29...
Paper: vidb.org/ovidb/vidb13/p1...
10:28 AM - Jan 25, 2021

Vinciiguerra/PyGM
@gmail.com
1:15 PM - Jan 25, 2021
The Algorithm Engineering 2020 Challenge

〜

Stats

13 participants
1 online
152 submissions
50 successful
94 with errors
8 didn't compile

Leaderboard

Which other initiatives for the next year?
The LA-vector

Next talk by Antonio Boffa
Currently investigated problems

- Study the combination of repetitiveness (LZ-like) and approximate linearity in the data (PGM-like)
  - New results with G. Manzini [T1,T3] + possible Patent @ UniPi
- Test PGM-index and other data structures in a real DB-scenario
  - Collaboration with Stratos Idreos @ Harvard [T1]
- Study new/engineered compressed and learned indexes for strings with applications to key-value stores [T1,T3]
- Study lossless compression of Neural Networks
  - New results with many people... [T2]
- Comparison between PPM and NN in chars prediction in texts
  - Collaboration with all other units [T4]

Details on the Talks + SW Libraries + Papers of this last year are available on the Web site of the project