

Quarto Meeting PRIN 2017

Febbraio 22, 2022

Unipa



- **Persone**
 - **Strutturati**
 - R. Giancarlo, G. Lo Bosco, S.E. Rombo
 - **Dottorandi**
 - D. Amato, **M. Bonomo**, G. Grimaudo

**Neighborhood based approaches for the
prediction of lncRNA-Disease
associations from tripartite graphs**

M. Bonomo, A La Placa, S.E. Rombo

Presentazione a breve



Tasks T1 e T3

- Learned Indexes- **Amato**, Lo Bosco, Giancarlo
- PhD Thesis. A Tour of Static Sorted Sets Dictionaries: From Specific to Generic with an Experimental Performance Analysis
- Learned Indexes in Small Model Space [AixIA21]
 - Two new models, competitive in space with best ones
 - Complement quite nicely the PGM Index across the Memory Hierachy



Tasks T1 e T3

- Learned Indexes- **Amato**, Lo Bosco, Giancarlo
- Uniform vs Standard Binary Search in Learned Indexes (preprint)
 - A quantitative justification of why Standard “Branchy” Binary Search is to be preferred to Uniform “Branch-free” in the Search stage in Learned Indexes.
- On the Suitability of Neural Networks in the Learned Indexes Context (manuscript in preparation)
 - Neural Nets have a big **PROMOTIONAL ROLE** in the initial proposal by Kraska et al., but ... quantification shows that they may be inadequate- **Better design needed**
 - Same scenario for Learned Bloom Filters? We are finding out (see next)

Tasks T1 e T3

- Learned Indexes- Amato, Lo Bosco, Giancarlo
 - From Specific to Generic Learned Dictionaries
 - Given a generic sorted set dictionary, other than Binary Search, how can we make it learned ?
 - A new paradigm of Learned Indexing and two case studies
 - Uniform Partition of the Universe into intervals: The “Demaine Data Structure” -Boosting Effect, Competitive with best proposals
 - Non-Uniform partition of the Universe into intervals: PGM index- Very Little Boosting Effect (the final stage of Binary Search essential for performance)

Tasks T1 e T3

- Learned Bloom Filters- Fumagalli, Raimondi, Giancarlo, Malchiodi, Frasca
 - On the role of Classifiers in Learned Bloom Filters [ICPRAMM22]
 - Marco farà cenno di cosa è stato fatto, ad ora.
 - Dependence of Learned Bloom Filters on Query Distributions
 - The performance, in particular false positive rate, of Bloom Filters is independent of query distribution.
 - Such a statement does not hold for Learned Bloom Filters- We are studying such a dependence.



Tasks T3-T4

- K-mers Dictionaries in Small Space for Alignment Freeness- Small Space makes large scale experiments possible
 - The power of word-frequency-based alignment-free functions: a comprehensive large-scale experimental analysis- Giuseppe Cattaneo, Ferraro Petrillo, Chiara Romualdi, Raffaele Giancarlo [Bioinformatics 22]



Task T2-T4

- . Topological ranks reveal functional knowledge encoded in biological networks: a comparative analysis- Bonomo, Rombo, Giancarlo
 - A new Paradigm to extract knowledge from Biological Networks, keeping only a part of the Network
 - A new way for lossy compression of Biological Networks, based on preserving important knowledge rather than on encoding of similar parts of a network- **to be better explored**
 - Preprint already available



Task T4

. Disk Storage of Compressed k-mer Dictionaries,
with or without Random Access in Main Memory-
Giancarlo, Grimaudo

- How to store succinctly k-mer dictionaries on disk-
Apparently a solved problem
 - A full characterization of the various scenarios and choices possible: **which, when and for what purpose**
 - Experiments under way, with some surprises
 - Maybe one day we will even finish this...