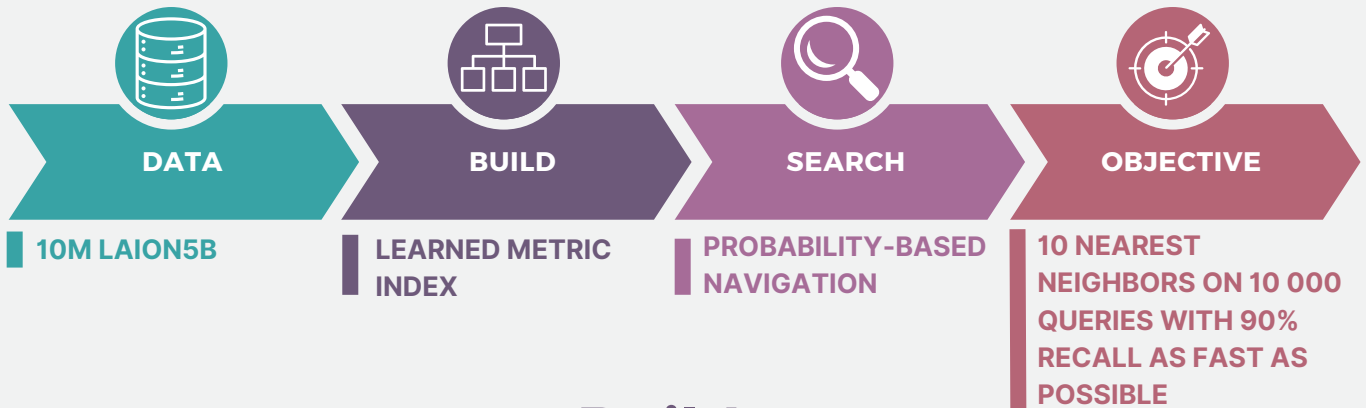# SISAP 2023 Indexing Challenge
Learned Metric Index

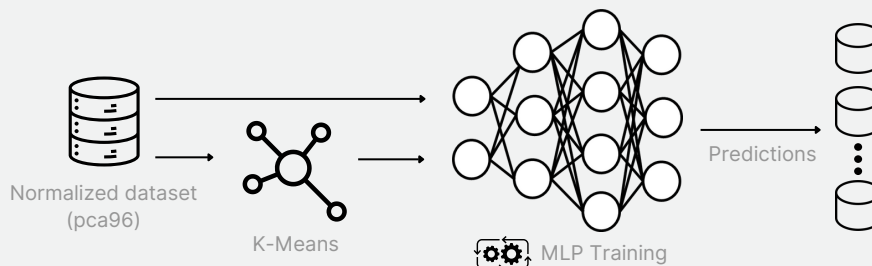T. Slanináková, D. Procházka, M. Antol, J. Oľha, V. Dohnal

github.com/LearnedMetricIndex

Masaryk University, Brno, Czech Republic

DATA	BUILD	SEARCH	OBJECTIVE
10M LAION5B	LEARNED METRIC INDEX	PROBABILITY-BASED NAVIGATION	10 NEAREST NEIGHBORS ON 10 000 QUERIES WITH 90% RECALL AS FAST AS POSSIBLE

Build

Normalized dataset (pca96) → K-Means → MLP Training → Predictions

Search

Queries (m=10 000, d=768)

PCA 96

NAVIGATION — Probability distribution — Categories=122

BUCKET-LEVEL SEQUENTIAL SEARCH
Mutual distances — Final distances
Cat 122
Cat 1
m (subset)
k=10
N. objects in cat
Distances — k
ANNs — k

Performance

Threshold filtering 9.7%
Navigation 0.5%
Sorting 16.1%
Computing distances in bucket-level search 72.6%

Time to process 10 000 queries (s)

- Distance computation
- Sorting
- Overall time

750
500
250
0

start 06/23 — submission 07/23 — now 09/23
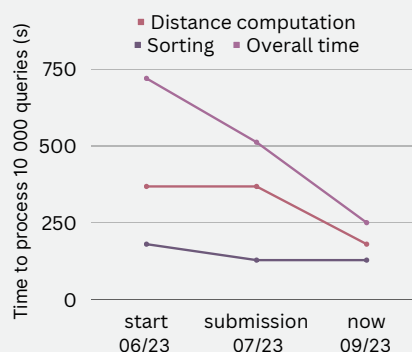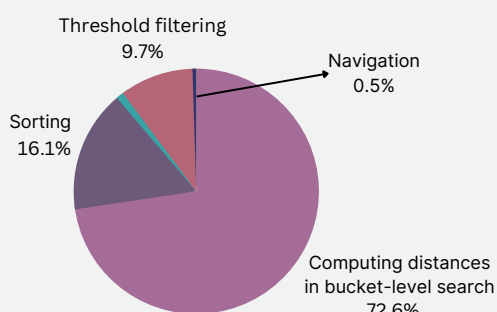
Future work

- Method being continually improved
- Single MLP in submission, now hierarchy
- Speed-up in build time by parallel MLP training
- Implementation of bucket object filtering
- Parallelization of bucket visits