

SUMS Dataset Processing Flowchart

Author: Levente Hajdu

Referenced Operations

Drops

Permanently removes files based on some criteria
Note: Drops make use of random selection to respect file system duplication

Sorts

Reorders the files based on a criteria

Splits

Divides the dataset into more jobs

