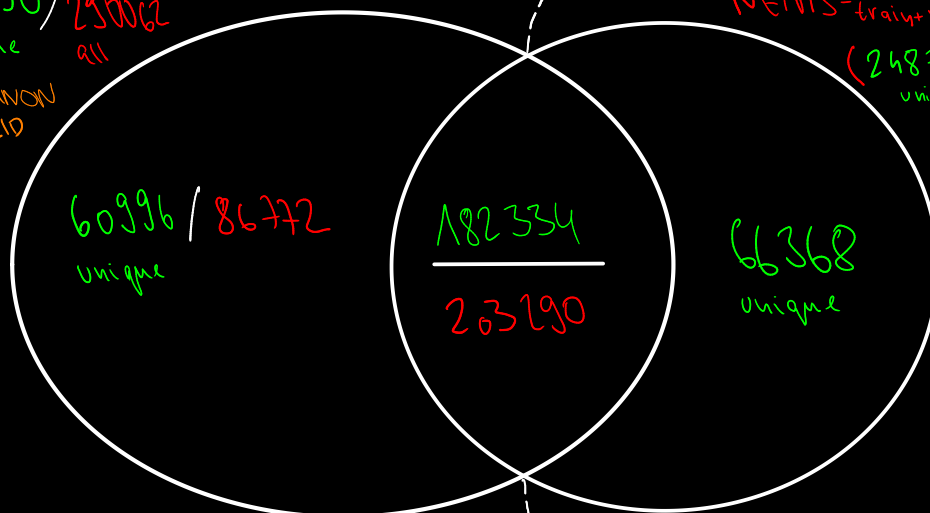


NIST20 (350618 spectra)

→ 60556 corrupted (no SMILES or no INCHIKEY
or RDX it failed to create MOLECULE
(not valid))

NIST-clean (with SMILES)
243330 / 290062
unique
all
CANON
VALID

237108 MESS
↑
NEIMS-train+valid
(248702 / 248707)
unique



With SMILES
CANON
VALID

$$A \cup B = 309698 / 356430$$