# How to remove records from deduplilcation

Created By: Knut Bøckman
Created on: 8/08/2018

Question:

How do I remove deduplication from record sets that are already deduped?

Answer:

To avoid deduplication of records, you can use t=99 in the dedup section of the Normalization rules, or you can set you pipes to end after stage "Persistence" (if you want neither dedup or FRBR) or stage "FRBR without dedup" (if you want FRBR). But this only keeps new records from going through the deduplication process, it does not reverse the process for records that are already deduped. In order to achieve this, you need to ensure that the data elements that creates a match between records are unique for each record, so they are separated insted of match. The solution is to configure you NR set so that the pnx/dedup/c.. and pnx/dedup/f.. fields are populated with a unique value for each record, instead of the standard values (ISBN, Title, Author, etc., that are there to find matches). The safest way to ensure uniqueness is to populate all theses fields (or the morst significant of them) with PNX value control/ recordid. Make sure PNX/dedup value of subfield "t" is set to 1 or 2, Then deploy the normalization set and run a no-harvest pipe through the dedup stage - the records will go through the deduplication process, and since none of them are matching, they will not be part of any dedup_mrg group. Run indexing and hotswap to complete the process.

Set PNX/dedup value of subfield  "t" to 99 to skip the dedup process for future pipe runs.

Report Abuse