

---

## Batch locating duplicate bibliographic records

- **Article Type:** Q&A
  - **Product:** Aleph
  - **Product Version:** 20, 21, 22, 23
- 

### Question

Is there an Aleph service to locate duplicate bibliographic records in a database? If not, is there SQL?

It should work like the check done by /exlibris/aleph/unn\_n/xxx01/tab/tab\_locate = check record by record on CAT module, but for the entire database. Can manage-26 be used for this?

---

### Answer

manage-26 detects similar headings rather than similar bibliographic records.

Article "[Check for Duplicate Record in Cataloging](#)" contains some very good information, discussing the "locate" function in detail, but does not discuss batch processing.

The manage-36 could be used. You would export the entire database (using the print-03 service) in Aleph sequential format and then use that as input to manage-36. The duplicates would be output in the output #3 "multi-match" file.

This manage-36 would produce a \*very\* large output file #2 (unique matches) -- which could be ignored -- but may cause the job to take a while to run.

The advantage to this is that you can use the existing match logic which you have set up for matching incoming records (rather than the multiple SQL's with OCLC#, ISSN, etc., described below)....

if the records have come from the same source, such as OCLC, something simpler may suffice.

The following SQL could be used to locate records with duplicate OCLC numbers:

```
select /*+ DYNAMIC_SAMPLING(2) ALL_ROWS */ substr (z11_rec_key,1,85), z11_doc_number from
z11 where z11_rec_key like '035%' and substr (z11_rec_key,1,85) in (select substr
(z11_rec_key,1,85) from z11 group by substr (z11_rec_key,1,85) having count(*) > 1) order
by substr (z11_rec_key,1,85), z11_doc_number;
```

Run against a usm01 z11 table, this SQL retrieved 112 rows.

This SQL uses the z11 035 index. If the OCLC number is in a different z11 index, that should be used. Also, other standard numbers, such as ISBN, LC card number, etc., can be used in addition or instead.

---

## Additional Information

[manage-36 is probably better than the following, but I am leaving it here, for the sake of completeness.]

The matching which needs to occur in detecting duplicates is very much like the logic used in matching records for Union View or a Union Catalog -- as discussed in Article 000007308. Probably this more sophisticated "union" logic will not usually be necessary but, there may be cases where it is....

If the z120 (Union View) table has been built for this xxx01 library, the following SQL could be used to locate duplicates:

```
SQL> select z120_rec_key_1, z120_rec_key from z120 where z120_same_no_lines > '001' order
by z120_rec_key_1, z120_rec_key;
```

-- though, in this case, the duplicates likely represent records from different libraries or different sources and "de-duplication" may not be desirable.

**Category:** Cataloging (500)

---

- **Article last edited:** 9-Aug-2018