

Parsing out Thesis call numbers

The following query can be used to parse out "Thesis" call numbers into the major components (LC Class, LC Class Number, First Cutter, Second Cutter and Year). Thesis call numbers can be very variable, so the query does not work for everything!

Most problems appear to be in getting the second cutter - any suggestions for improvement are gratefully accepted! - JL 11-30-23

-- 11-22-23: This query parses out thesis call numbers into LC class, Lc class number, internal year, first cutter and second cutter

-- Does not work well with malformed call numbers or very unusual thesis call number schemes

-- 11-30-23: updated query to work with Law theses and sometimes for Thesis Film

-- 1. Select thesis holdings records where "Thesis" is part of the call number or call number prefix

WITH parameters AS

(SELECT

'%Mann%':VARCHAR AS library_name_filter

),

recs AS (

SELECT

ll.library_name,
he.permanent_location_name,
he.holdings_hrid,
he.discovery_suppress,
he.call_number_prefix,
he.call_number

FROM folio_reporting.holdings_ext AS he

LEFT JOIN folio_reporting.locations_libraries AS ll

ON he.permanent_location_id = ll.location_id

WHERE (he.call_number ILIKE '%thesis%' OR he.call_number_prefix ILIKE '%thesis%')

AND (he.discovery_suppress = 'False' OR he.discovery_suppress IS NULL)

AND ll.library_name ILIKE (SELECT library_name_filter FROM parameters)

)

-- 2. Apply parsing formulas to call number to get LC class, LC class number, internal year, first cutter and second cutter

SELECT

recs.library_name,
recs.permanent_location_name,
recs.holdings_hrid,

```
recs.call_number_prefix,  
recs.call_number,
```

```
-- get LC class
```

```
CASE
```

```
    WHEN recs.call_number ILIKE '%film%' THEN ' - '
```

```
    WHEN recs.call_number ILIKE '%thesis%' THEN SUBSTRING (recs.call_number, 's[A-Za-z]{1,3}')  
    ELSE SUBSTRING (recs.call_number, '[A-Za-z]{1,3}') END AS lc_class,
```

```
-- get LC class number
```

```
CASE WHEN recs.call_number ILIKE '%film%' THEN NULL
```

```
ELSE
```

```
    (TRIM (trailing '.' FROM
```

```
        CASE
```

```
            WHEN SUBSTRING (recs.call_number, 1, 1) IN ('1', '2') THEN NULL
```

```
            WHEN SUBSTRING (recs.call_number, 2, 1) = '' THEN NULL
```

```
            ELSE SUBSTRING (recs.call_number, 'd{1,4}\.{0,}\d{0,}')  
        END))
```

```
END AS lc_class_number,
```

```
-- get internal year
```

```
    TRIM (CASE WHEN SUBSTRING (recs.call_number, 1, 1) IN ('1', '2')
```

```
        THEN SUBSTRING (recs.call_number, 1, 4)
```

```
        ELSE SUBSTRING (recs.call_number, 's\d{4}\s{0,1}')
```

```
        END) AS internal_year,
```

```
-- get first cutter
```

```
CASE WHEN recs.library_name NOT LIKE 'Law%'
```

```
THEN SUBSTRING (recs.call_number, 'd\s{0,1}[A-Z]\d{1,}')
```

```
ELSE SUBSTRING (recs.call_number, '\.[A-Z]\d{1,}[a-z]{0,1}')
```

```
END AS first_cutter,
```

```
-- get second cutter
```

```
CASE
```

```
    WHEN SUBSTRING (recs.call_number, '\.s{0,1}[A-Z]\d{1,}') IS NULL
```

```
OR
```

```
    TRIM (SUBSTRING (recs.call_number, '\.s{0,1}[A-Z]\d{1,}')) = TRIM (SUBSTRING (recs.call_number, 's{1}\.{0,1}[A-Z]{1}\d{1,}[a-z]{0,1}'))
```

```
THEN NULL
```

```
ELSE SUBSTRING (recs.call_number, 's{1}[A-Z]{1}\d{1,}x{0,}a{0,}b{0,}') END AS second_cutter
```

FROM recs

ORDER BY

permanent_location_name,

-- then order by LC class

lc_class,

-- then order by lc class number (using the entire CASE WHEN statement above)

CASE

WHEN SUBSTRING (recs.call_number,1,1) IN ('1','2') THEN NULL

WHEN SUBSTRING (recs.call_number,2,1) = ' ' THEN NULL

ELSE SUBSTRING (recs.call_number, 'd{1,4}\.{0,}\d{0,}')

END)::NUMERIC,

-- then order by internal year (using the entire CASE WHEN statement)

TRIM (CASE WHEN substring (recs.call_number,1,1) IN ('1','2')

THEN substring (recs.call_number,1,4)

ELSE SUBSTRING (recs.call_number, 's\d{4}\s{0,1}')

END)::INTEGER,

-- then order by first cutter, then second cutter, then the whole call number in case the cutters are blank

first_cutter,

second_cutter,

recs.call_number

;