

U.S. Federal Institute of Museum and Library Services
National Leadership Grant

**Realizing the Vision of Networked Access to
Library Resources**

*An Applied Research and Demonstration Project to
Establish and Operate a Z39.50 Interoperability Testbed*

**SQL Data Analysis Procedures to
Create Aggregate and Candidate
Record Groups on Sample of
Decomposed MARC Records**

Phase 1 Testing

Jung-Won Yoon
<jw_yoon@hotmail.com>
Z-Interop Research Assistant
&

William E. Moen, Ph.D.
<wemoen@unt.edu>
Principal Investigator

School of Library and Information Sciences
Texas Center for Digital Knowledge
University of North Texas
Denton, TX 76203

October 14, 2001
Revised January 1, 2002

Table of Contents

- 1. Introduction**
- 2. Overview of Data Analysis Procedure**
- 3. Detailed Data Analysis Procedure**

Appendix A: Sample Reports from Analysis Procedures

SQL Data Analysis Procedures to Create Aggregate and Candidate Record Groups on Sample of Decomposed MARC Records

1. Introduction

This document describes the data analysis procedures developed to create the Aggregate and Candidate Record Groups using SQL statements. This is the preliminary version of these procedures tested and validated on a sample of decomposed MARC records. (For a description of how the MARC records were decomposed see the Z-Interop document, **Decomposing MARC 21 Records for Analysis** A subsequent version may be necessary as we move to the procedures for the entire file of decomposed records.

These analysis procedures were guided by the logic and guidelines contained in the Z-Interop document, **Analysis Logic and Procedures for Creating a Test Dataset of MARC 21 Records for the Z39.50 Interoperability Testbed**. For details on the approach and purpose of the analyses described below, see this document.

2. Overview of Data Analysis Procedure

SQL statements were developed to process and analyze the decomposed MARC 21 records. Before working on the complete set of decomposed records from the 400,000 MARC 21 records, the following SQL statements were tested on 9,366 decomposed records from sample of 152 MARC 21 records. The decomposed records were loaded into a MYSQL database running on a Sun Ultra machine for processing.

Normalization procedures were performed on the decomposed records prior to the SQL analysis. The normalization procedures are described in the Z-Interop document, **Data Normalization Procedures on Decomposed MARC 21 Records**.

The procedure for developing SQL statements can be categorized as four: Create Database, Create Table, Load Data, and Create Queries.

Task	Output
Create Database	Database: test_z
Create Table	Table: SAMPLE1 SEARCHTERM*
Load Data	Inserted data from 9,366 decomposed MARC records to SAMPLE1
Create Queries	<ul style="list-style-type: none"> • Frequency list (frequency.txt)** • Alphabetical list (alphabet.txt) ** • Aggregate group report A (aggregate.txt) ** • Aggregate group report B (baggregate.txt) ** • Candidate group for title keyword (title.txt) ** • Candidate group for author keyword (author.txt) ** • Candidate group for subject keyword (subject.txt) ** • Candidate group for any keyword (any.txt) **

* SEARCHTERM table will be used until the form interface supporting the single term search is developed.

** The output files are under /usr/local/mysql/var/test_z

3. Detailed Data Analysis Procedure

1) Create Database

Database that will contain tables is created. The database name is "test_z".

```
CREATE DATABASE test_z;
```

2) Create Table

- The table that will contain decomposed records is created. Field names and the data type for each field should be determined. The table name is "SAMPLE1".

```
CREATE TABLE SAMPLE1 (  
    S_CODE INTEGER NOT NULL AUTO_INCREMENT,  
    S_OCLC INTEGER,  
    S_FIELD CHAR(3),  
    S_FLDPOSINREC INTEGER,  
    S_FIELDOCC INTEGER,  
    S_SUBFLDCODE CHAR(1),  
    S_SUBFLDPOS INTEGER,  
    S_SUBFLDOCC INTEGER,  
    S_WORDPOSIINFLD INTEGER,  
    S_WORD VARCHAR(20),  
    PRIMARY KEY (S_CODE));
```

- The table that will contain search terms is created. This table will be used until the form interface supporting the single term search is developed. The table name is "SEARCHTERM"

```
CREATE TABLE SEARCHTERM(  
    W_CODE INTEGER NOT NULL AUTO_INCREMENT,  
    W_WORD VARCHAR(20),  
    PRIMARY KEY (W_CODE));
```

3) Load Data

9,366 decomposed records are loaded to the SAMPLE 1 table.

```
LOAD DATA INFILE 'sample_norm'  
    INTO TABLE SAMPLE1 (S_OCLC, S_FIELD, S_FLDPOSINREC,  
        S_FIELDOCC, S_SUBFLDCODE, S_SUBFLDPOS, S_SUBFLDOCC,  
        S_WORDPOSIINFLD, S_WORD);
```

4) Create Queries

The eight queries are developed for several reports based on the requirements in the Z-Interop document, ***Analysis Logic and Procedures for Creating a Test Dataset of MARC 21 Records for the Z39.50 Interoperability Testbed.*** Refer to that document for logic and justification of each report.

For creating the candidate groups for specific types of searches (e.g., title, author, subject), the choice of areas of the MARC 21 record to interrogate is in line with the recommendations for indexing MARC 21 records to support searches defined in the Bath and U.S. National Z39.50 profiles. For more information, see the Z-Interop document, ***Indexing Guidelines to Support Z39.50 Profile Searches.***

- **Frequency list**

The output is a count of all the words in the records in ascending order. This report will be used for selecting terms for subsequent analysis and their use in test searches. For sample of report, see Appendix A.

```
SELECT COUNT(S_WORD), S_WORD INTO OUTFILE 'frequency.txt'
  FROM SAMPLE1
  GROUP BY S_WORD
  ORDER BY 1;
```

- **Alphabetical list**

The output is a list of all the words from the frequency count in alphabetical order. This report will be used for 1) checking to ensure that the frequency count is being done appropriately, and 2) selecting words for truncated word searches. For sample of report, see Appendix A.

```
SELECT COUNT(S_WORD), S_WORD INTO OUTFILE 'alphabet.txt'
  FROM SAMPLE1
  GROUP BY S_WORD
  ORDER BY 2;
```

- **Aggregate group report A**

The output is a list of OCLC record numbers in which a search term appears. This report will be used for identifying all possible records in which a search term occurs to check against the list of records returned from test searches. For sample of report, see Appendix A.

```
SELECT SAMPLE1.S_OCLC, S_WORD INTO OUTFILE 'aggregate.txt'
  FROM SAMPLE1, SEARCHTERM
  WHERE (SAMPLE1.S_WORD = SEARCHTERM.W_WORD)
  GROUP BY S_WORD, S_OCLC;
```

- **Aggregate group report B**

The output is a list of OCLC record numbers, fields and subfields in which a search term occurs. This report will be used for 1) providing data upon which the Candidate Record Group are derived, and 2) analyzing test search results when records not in the Candidate Record Groups are returned from a search. For sample of report, see Appendix A.

```
SELECT SAMPLE1.S_OCLC, S_FIELD, S_SUBFLDCODE, S_WORD INTO OUTFILE
  'baggregate.txt'
  FROM SAMPLE1, SEARCHTERM
  WHERE (SAMPLE1.S_WORD = SEARCHTERM.W_WORD)
  ORDER BY 4, 1;
```

- **Candidate group for title keyword**

The output is a list of the record numbers in which a term appears in selected fields/subfields for title keyword search. This report will be used to establish the benchmark of the SIRSI reference implementation for the title keyword search. Refer to **Indexing Guidelines to Support Z39.50 Profile Searches** for field information. For sample of report, see Appendix A.

```
SELECT S_OCLC, S_FIELD, S_SUBFLDCODE, S_WORD INTO OUTFILE
  'title1.txt'
  FROM SAMPLE1, SEARCHTERM
  WHERE (SAMPLE1.S_WORD = SEARCHTERM.W_WORD) AND
  ((S_FIELD = '100' AND S_SUBFLDCODE IN ('f', 'g', 'k', 'l', 'n',
  'p', 't')))
```

```

        OR (S_FIELD = '110' AND S_SUBFLDCODE IN ('p', 't'))
        OR (S_FIELD = '111' AND S_SUBFLDCODE IN ('d', 'f', 'g',
'k', 'l', 'n', 'p', 't'))
    OR (S_FIELD = '130' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't'))
    OR (S_FIELD = '210' AND S_SUBFLDCODE IN ('a', 'b'))
    OR (S_FIELD = '211' AND S_SUBFLDCODE = 'a')
    OR (S_FIELD = '212' AND S_SUBFLDCODE = 'a')
    OR (S_FIELD = '214' AND S_SUBFLDCODE = 'a')
    OR (S_FIELD = '222' AND S_SUBFLDCODE IN ('a', 'b'))
    OR (S_FIELD = '240' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's'))
    OR (S_FIELD = '241' AND S_SUBFLDCODE IN ('a', 'h'))
    OR (S_FIELD = '242' AND S_SUBFLDCODE IN ('a', 'b', 'h', 'n',
'y', 'p'))
    OR (S_FIELD = '243' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's'))
        OR (S_FIELD = '245' AND S_SUBFLDCODE IN ('a', 'b', 'f',
'g', 'h', 'k', 'n', 'p', 's'))
        OR (S_FIELD = '246' AND S_SUBFLDCODE IN ('a', 'b', 'f',
'g', 'h', 'n', 'p'))
        OR (S_FIELD = '247' AND S_SUBFLDCODE IN ('a', 'b', 'f',
'g', 'h', 'n', 'p'))
        OR (S_FIELD = '400' AND S_SUBFLDCODE IN ('n', 'p', 't',
'v'))
    OR (S_FIELD = '410' AND S_SUBFLDCODE IN ('d', 'f', 'g',
'k', 'l', 'n', 'p', 't', 'v'))
    OR (S_FIELD = '411' AND S_SUBFLDCODE IN ('p', 't', 'v'))
    OR (S_FIELD = '440' AND S_SUBFLDCODE IN ('a', 'n', 'p'))
    OR (S_FIELD = '490' AND S_SUBFLDCODE IN ('a', 'v'))
    OR (S_FIELD = '505' AND S_SUBFLDCODE IN ('a', 'g', 't'))
    OR (S_FIELD = '534' AND S_SUBFLDCODE = 't')
    OR (S_FIELD = '700' AND S_SUBFLDCODE IN ('m', 'n', 'o',
'p', 'r', 's', 't'))
        OR (S_FIELD = '710' AND S_SUBFLDCODE IN ('d', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't'))
        OR (S_FIELD = '711' AND S_SUBFLDCODE IN ('m', 'n', 'o',
'p', 'q', 'r', 's', 't'))
        OR (S_FIELD = '730' AND S_SUBFLDCODE IN ('a', 'd', 'f',
'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't'))
        OR (S_FIELD = '740' AND S_SUBFLDCODE IN ('a', 'h', 'n',
'p'))
    OR (S_FIELD = '760' AND S_SUBFLDCODE IN ('s', 't'))
    OR (S_FIELD = '762' AND S_SUBFLDCODE IN ('s', 't'))
    OR (S_FIELD = '765' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
    OR (S_FIELD = '767' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
    OR (S_FIELD = '770' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
    OR (S_FIELD = '772' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
    OR (S_FIELD = '773' AND S_SUBFLDCODE IN ('c', 'k', 'p',
's', 't'))
    OR (S_FIELD = '774' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))

```

```

        OR (S_FIELD = '775' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '776' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '777' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '780' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '785' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '786' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '787' AND S_SUBFLDCODE IN ('c', 'k', 's',
't'))
        OR (S_FIELD = '800' AND S_SUBFLDCODE IN ('f', 'g', 'h',
'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'v'))
        OR (S_FIELD = '810' AND S_SUBFLDCODE IN ('f', 'g', 'h',
'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'v'))
        OR (S_FIELD = '811' AND S_SUBFLDCODE IN ('f', 'g', 'h',
'k', 'l', 'n', 'p', 's', 't', 'v'))
    OR (S_FIELD = '830' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'v'))
    ORDER BY 4, 1;

```

- **Candidate group for author keyword**

The output is a list of the record numbers in which a term appears in selected fields/subfields for author keyword search. This report will be used to establish the benchmark of the SIRSI reference implementation for the author keyword search. Refer to **Indexing Guidelines to Support Z39.50 Profile Searches** for field information. For sample of report, see Appendix A.

```

SELECT S_OCLC, S_FIELD, S_SUBFLDCODE, S_WORD INTO OUTFILE
'author.txt'
FROM SAMPLE, WORDTEST
WHERE (SAMPLE.S_WORD = WORDTEST.W_WORD) AND
((S_FIELD = '100' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd', 'e',
'g', 'j', 'k', 'q', 'u', '4'))
OR (S_FIELD = '110' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd',
'e', 'g', 'k', 'n', 'u', '4'))
OR (S_FIELD = '111' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'g', 'n', 'q', 'u', '4'))
OR (S_FIELD = '242' AND S_SUBFLDCODE = 'c')
OR (S_FIELD = '245' AND S_SUBFLDCODE = 'c')
OR (S_FIELD = '400' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'k', 'u', '4'))
OR (S_FIELD = '410' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'k', 'n', 'u'))
OR (S_FIELD = '411' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'g', 'n', 'q', 'u', '4'))
OR (S_FIELD = '505' AND S_SUBFLDCODE = 'r')
OR (S_FIELD = '508' AND S_SUBFLDCODE = 'a')
OR (S_FIELD = '511' AND S_SUBFLDCODE = 'a')
OR (S_FIELD = '534' AND S_SUBFLDCODE = 'a')
OR (S_FIELD = '700' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'j', 'k', 'q', 'u', '4'))
OR (S_FIELD = '705' AND S_SUBFLDCODE = 'a')

```

```

        OR (S_FIELD = '710' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'k', 'n', 'u', '4'))
        OR (S_FIELD = '711' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'g', 'n', 'q', 'u', '4'))
        OR (S_FIELD = '715' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '720' AND S_SUBFLDCODE IN ('a', 'e'))
        OR (S_FIELD = '760' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '762' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '765' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '767' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '770' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '772' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '773' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '774' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '775' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '776' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '777' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '780' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '785' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '786' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '787' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '800' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'j', 'k', 'q', 'u', '4'))
        OR (S_FIELD = '810' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'k', 'n', 'u', '4'))
        OR (S_FIELD = '811' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'g', 'n', 'q', 'u', '4'))
ORDER BY 4, 1;

```

- **Candidate group for subject keyword**

The output is a list of the record numbers in which a term appears in selected fields/subfields for subject keyword search. This report will be used to establish the benchmark of the SIRSI reference implementation for the subject keyword search. Refer to **Indexing Guidelines to Support Z39.50 Profile Searches** for field information. For sample of report, see Appendix A.

```

SELECT S_OCLC, S_FIELD, S_SUBFLDCODE, S_WORD INTO OUTFILE
'subject1.txt'
FROM SAMPLE, WORDTEST
WHERE (SAMPLE.S_WORD = WORDTEST.W_WORD) AND
((S_FIELD = '600' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd', 'e',
'f', 'g', 'h', 'j', 'k', 'l', 'm', 'n', 'o', 'p', 'q', 'r', 's',
't', 'u', 'v', 'x', 'y', 'z', '2', '4'))
OR (S_FIELD = '610' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd',
'e', 'f', 'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't',
'u', 'v', 'x', 'y', 'z', '2', '4'))
OR (S_FIELD = '611' AND S_SUBFLDCODE IN ('a', 'c', 'd', 'e',
'f', 'g', 'h', 'k', 'l', 'n', 'p', 'q', 's', 't', 'u', 'v', 'x',
'y', 'z', '2', '4'))
OR (S_FIELD = '630' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'v', 'x', 'y',
'z', '2'))
OR (S_FIELD = '650' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd',
'e', 'v', 'x', 'y', 'z', '2'))
OR (S_FIELD = '651' AND S_SUBFLDCODE IN ('a', 'v', 'x', 'y', 'z',
'2'))

```

```

        OR (S_FIELD = '653' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '654' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'v',
'y', 'z', '2'))
        OR (S_FIELD = '655' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'v',
'x', 'y', 'z', '2'))
        OR (S_FIELD = '656' AND S_SUBFLDCODE IN ('a', 'k', 'v', 'x',
'y', 'z', '2'))
        OR (S_FIELD = '657' AND S_SUBFLDCODE IN ('a', 'v', 'x', 'y',
'z', '2'))
        OR (S_FIELD = '658' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd',
'2'))
        OR (S_FIELD = '752' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd'))
ORDER BY 4, 1;

```

- **Candidate group for any keyword**

The output is a list of the record numbers in which a term appears in selected fields/subfields for any keyword search. This report will be used to establish the benchmark of the SIRSI reference implementation for the any keyword search. For sample of report, see Appendix A.

```

SELECT S_OCLC, S_FIELD, S_SUBFLDCODE, S_WORD INTO OUTFILE 'any.txt'
FROM SAMPLE, WORDTEST
WHERE (SAMPLE.S_WORD = WORDTEST.W_WORD) AND
((S_FIELD = '100' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd', 'e',
'f', 'g', 'j', 'k', 'l', 'n', 'p', 'q', 't', 'u', '4'))
OR (S_FIELD = '110' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd', 'e',
'g', 'k', 'n', 'p', 't', 'u', '4'))
OR (S_FIELD = '111' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'f', 'g', 'k', 'l', 'n', 'p', 'q', 't', 'u', '4'))
OR (S_FIELD = '130' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't'))
OR (S_FIELD = '210' AND S_SUBFLDCODE IN ('a', 'b'))
OR (S_FIELD = '211' AND S_SUBFLDCODE = 'a')
OR (S_FIELD = '212' AND S_SUBFLDCODE = 'a')
OR (S_FIELD = '214' AND S_SUBFLDCODE = 'a')
OR (S_FIELD = '222' AND S_SUBFLDCODE IN ('a', 'b'))
OR (S_FIELD = '240' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's'))
OR (S_FIELD = '241' AND S_SUBFLDCODE IN ('a', 'h'))
OR (S_FIELD = '242' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'h',
'n', 'y', 'p'))
OR (S_FIELD = '243' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's'))
OR (S_FIELD = '245' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'f', 'g', 'h', 'k', 'n', 'p', 's'))
OR (S_FIELD = '246' AND S_SUBFLDCODE IN ('a', 'b', 'f',
'g', 'h', 'n', 'p'))
OR (S_FIELD = '247' AND S_SUBFLDCODE IN ('a', 'b', 'f',
'g', 'h', 'n', 'p'))
OR (S_FIELD = '400' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'k', 'n', 'p', 't', 'u', 'v', '4'))
OR (S_FIELD = '410' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'f', 'g', 'k', 'l', 'n', 'p', 't', 'u', 'v'))
OR (S_FIELD = '411' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'g', 'n', 'p', 'q', 't', 'v', '4', 'u'))
OR (S_FIELD = '440' AND S_SUBFLDCODE IN ('a', 'n', 'p'))
OR (S_FIELD = '490' AND S_SUBFLDCODE IN ('a', 'v'))

```

```

        OR (S_FIELD = '505' AND S_SUBFLDCODE IN ('a', 'g', 't',
'r'))
        OR (S_FIELD = '508' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '511' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '534' AND S_SUBFLDCODE IN ('a', 't'))
        OR (S_FIELD = '600' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'f', 'g', 'h', 'j', 'k', 'l', 'm', 'n', 'o', 'p',
'q', 'r', 's', 't', 'u', 'v', 'x', 'y', 'z', '2', '4'))
    OR (S_FIELD = '610' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd',
'e', 'f', 'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't',
'u', 'v', 'x', 'y', 'z', '2', '4'))
        OR (S_FIELD = '611' AND S_SUBFLDCODE IN ('a', 'c', 'd', 'e',
'f', 'g', 'h', 'k', 'l', 'n', 'p', 'q', 's', 't', 'u', 'v', 'x',
'y', 'z', '2', '4'))
        OR (S_FIELD = '630' AND S_SUBFLDCODE IN ('a', 'd', 'f', 'g',
'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'v', 'x', 'y',
'z', '2'))
        OR (S_FIELD = '650' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'd',
'e', 'v', 'x', 'y', 'z', '2'))
        OR (S_FIELD = '651' AND S_SUBFLDCODE IN ('a', 'v', 'x', 'y',
'z', '2'))
        OR (S_FIELD = '653' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '654' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'v',
'y', 'z', '2'))
        OR (S_FIELD = '655' AND S_SUBFLDCODE IN ('a', 'b', 'c', 'v',
'x', 'y', 'z', '2'))
        OR (S_FIELD = '656' AND S_SUBFLDCODE IN ('a', 'k', 'v', 'x',
'y', 'z', '2'))
        OR (S_FIELD = '657' AND S_SUBFLDCODE IN ('a', 'v', 'x', 'y',
'z', '2'))
        OR (S_FIELD = '658' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', '2'))
        OR (S_FIELD = '700' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'g', 'j', 'k', 'm', 'n', 'o', 'p', 'q', 'r', 's',
't', 'u', '4'))
        OR (S_FIELD = '705' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '710' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'f', 'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r',
's', 't', 'u', '4'))
        OR (S_FIELD = '711' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'g', 'm', 'n', 'o', 'p', 'q', 'r', 's', 't', 'u', '4'))
        OR (S_FIELD = '715' AND S_SUBFLDCODE = 'a')
        OR (S_FIELD = '720' AND S_SUBFLDCODE IN ('a', 'e'))
        OR (S_FIELD = '730' AND S_SUBFLDCODE IN ('a', 'd', 'f',
'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't'))
        OR (S_FIELD = '740' AND S_SUBFLDCODE IN ('a', 'h', 'n',
'p'))
        OR (S_FIELD = '752' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd'))
        OR (S_FIELD = '760' AND S_SUBFLDCODE IN ('a', 's', 't'))
        OR (S_FIELD = '762' AND S_SUBFLDCODE IN ('a', 's', 't'))
        OR (S_FIELD = '765' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '767' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '770' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))

```

```
        OR (S_FIELD = '772' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '773' AND S_SUBFLDCODE IN ('a', 'c', 'k',
'p', 's', 't'))
        OR (S_FIELD = '774' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '775' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '776' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '777' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '780' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '785' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '786' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '787' AND S_SUBFLDCODE IN ('a', 'c', 'k',
's', 't'))
        OR (S_FIELD = '800' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'f', 'g', 'h', 'j', 'k', 'l', 'm', 'n', 'o', 'p',
'q', 'r', 's', 't', 'u', 'v', '4'))
        OR (S_FIELD = '810' AND S_SUBFLDCODE IN ('a', 'b', 'c',
'd', 'e', 'f', 'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r',
's', 't', 'u', 'v', '4'))
        OR (S_FIELD = '811' AND S_SUBFLDCODE IN ('a', 'c', 'd',
'e', 'f', 'g', 'h', 'k', 'l', 'n', 'p', 'q', 's', 't', 'u',
'v', '4'))
        OR (S_FIELD = '830' AND S_SUBFLDCODE IN ('a', 'd', 'f',
'g', 'h', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'v'))))
ORDER BY 4, 1;
```

Appendix A

Sample Reports from Analysis Procedures

The following provide selections from the reports generated via the SQL procedures described in this document.

Frequency Count Report

This shows a sample of the output. Listed here are the "words" that appear between 25 and 100 times in the sample records.

25	1969	30	1969	40	Bibliographical
26	A	30	J	41	to
26	an	30	with	44	[1969]
26	edited	31	illus.	45	Bibliography:
26	UKM	32	on	51	OCL
27	23	32	Press	56	University
27	joint	32	[by]	58	Illus
27	[1968]	37	United	63	York
28	Includes	38	24	76	New
29	22	39	ed	86	in

Alphabetical Order Report

This shows a sample of the output. Listed here are the entries beginning with the letter A.

26	A	1	A6	1	account
1	a-ja---	1	A64	1	acquisition
1	a-ph---	2	Abbott	3	act
1	a-tu---	1	Ability	1	activity
1	A.S	1	able	1	Actors
1	A163	2	about	3	Adams
1	A3	2	Abraham	1	additional
1	A413	2	Abrams	1	Adirondack
2	A45	1	abroad	1	Administration
1	A5	2	Academic		

Aggregate Report A

Listed here is a sample of the report for the word "Charles." This report lists the OCLC number and the term.

225	Charles
740	Charles
872	Charles
1154	Charles

Aggregate Report B

Listed here is a sample of the report for the word "Charles." This report lists the OCLC number, the field and subfields in which the term occurs, and the term.

225	245 a	Charles
740	504 a	Charles
740	600 q	Charles
872	440 a	Charles
1154	700 c	Charles

Candidate Group for Title Keyword Report

Listed here is a sample of the report for the term "College" as a title word. This report lists the OCLC number, the title-related fields and subfields.

421	245 b	College
442	245 a	College
442	490 a	College
834	505 a	College

Candidate Group for Author Keyword Report

Listed here is a sample of the report for the term "Charles" as an author word. This report lists the OCLC number, the author-related fields and subfields, and the term.

1154	700 c	Charles
------	-------	---------

Candidate Group for Subject Keyword Report

Listed here is a sample of the report for the term "Drama" as a subject word. This report lists the OCLC number, the subject-related fields and subfields, and the term.

165	600 d	Drama
165	651 y	Drama
165	655 a	Drama

Candidate Group for Any Keyword Report

Listed here is a sample of the report for the term "Edward" as a word appearing in at least an author, title, or subject related field. This report lists the OCLC number, the general keyword related fields and subfields, and the term.

759	505 a	Edward
833	490 a	Edward
833	651 x	Edward