EARLY CANADIAN CULTURAL JOURNALS INDEX (ECCJ)

A. Background and Introduction:

This project began in the early 1980s to experiment with the application of emerging relational database technology to problems of accessing and retrieving information published in older Canadian cultural journals by using bibliographical notation with some additional searchable fields. The project was supported by a significant SSHRC Research Grant that ran for three years. At the time, computing power was limited; the project actually began on a mainframe using Honeywell ARES, an early SQL-software, and migrated to ORACLE on a desk top after the advent of 386 computers. The limited availability of computing power affected the database design and the lingering impact of that is the large number of separate fields in the database structure. Basically, a smaller amount of data distributed through a large number of fields offered a faster response time to queries than a lot of data in a few fields. I suspect (if desired) that could be changed by reorganizing the data; in today's computing, the difference in response time would be negligible. But because (to me) there seemed little to gain, I have left the structure as originally designed.

The other thing that is worth noting is that the database not only functions as an index, allowing various avenues of access to the information it holds, but it also acts as a tool to construct (or re-construct) detailed tables-of-contents (including illustrations, advertisements, etc.) to individual issues and whole runs of journals. This capacity expands the analytical usefulness of the data collection.

I think I mentioned a couple early publications on the project, but I repeat them here FYI:

"Early Canadian Cultural Journals Databank Project: A Preliminary Report" Bulletin of The Bibliographical Society of Canada n.s. No. 27 (Nov. 1986), 9-12.

"Opportunities and Pitfalls: Observations on the Application of Database Software to Bibliographical Activity," Papers of the Bibliographical Society of Canada, 28 (1989): 13-24.

Both these can be found online at the Bibliographical Society of Canada website; look under publications for Bulletin and for Papers.

B. Basic Statistics:

The database covers 203 magazines/journals published in Nineteenth-Century Canada and consists of 137, 943 records/entries of 17 fields each. Not all runs are completed; only what we could find extant is included. If you are interested, I have a small "control" database that records what we have seen, where it exists, what has been entered, and what is missing.

C. Database Field Structure:

The raw data that I am sending you is in comma delimited format, sequentially arranged in 17 fields; there is one full record/entry per line. Here is the sequence of the data using the field names I used in creating the original database structure and indicating the original field lengths (in characters). These field names are not part of the raw data file. [Incidentally, if it is of any use to you, I currently have working model of the database in Inmagic DB/Textworks software. This is a different kind of database structure, so the Dump Files do not work with most other forms of database software. Comma delimited format is more universal.]

Field Name	Char	Note
JNL	[3]	- each "Journal" was given a 3-letter code to simplify queries and (when combined with NDX) to form a unique identifier for each record
NDX	[4]	- each record of each journal has an "index" number beginning at 0001 which, when combined with the JNL code, forms a unique identifier within the database as a whole
SNAME	[15]	- author's "Surname" (for given article)
FNAME	[15]	- author's "First name(s)" or initials
TI	[65]	- "Title" of article
MAG	[60]	- literal title of "Magazine" in which article appears (ie, longer and more recognizable version of JNL)
VOL	[2]	- "Volume" of Magazine/Journal
NUM	[2]	- volume "Number" of Magazine/Journal
MO	[3]	- "Month" (3-letter code)
YR	[4]	- "Year"
PG	[9]	- "page" reference (eg., 127-215)
S1	[20]	- "Subject" term 1 (usually a broad category)
S2	[20]	- "Subject" term 2 (usually a narrower or more specific term)

GEN	[1]	- "Genre" (a 1-letter code: P = prose; V = verse; F = fiction; R = report/news; I = illustration; M = miscellaneous, including advertisements)
ORG	[1]	- "origin" of entry (if known) – (a 1-letter code: C = Canadian; O = Other)
FLINE	[65]	- "First Line" of poetry only
SP	[2 at least]	- "Spare" (unused field filled with "xx"; included to avoid re-structuring the database if additional field needed)

D. Some Notes on Types of Searches:

We formulated search queries in whatever software protocol we were using at the time. These were the ones the users seemed to be most interested in:

1. by author's surnames and names

2. by keyword in title

3. by subject term in both subject fields (ie., an "and" structure)

Note: the old CD-ROM had a number of useful help features. For example, it allowed the user to browse an alpha-list of the terms used in S1 and S2 (and also other fields). I suspect they used the SQL "List" Command to consolidate the terminology employed in a given field. This feature allowed the user to look for likely "terms" rather than our multiplying the keywords, which would have slowed response time (ie., we expected the user's own mind to provide an addition level of computing). This type of browse was also useful in the MAG field in that it allowed the user to check and see if the magazine/journal they were interested in was included in the Index.

E. Some Notes on Report Formats:

The basic Report Format we used was the standard (MLA) bibliographical citation for journals with added annotation – all in a 3-line output (with hanging indent). We found this form of output was easily scanned by the eye of the user. Here is the basic shape of the output with punctuation and notation:

SNAME, FNAME, "TI," MAG VOL: NUM (MON YR), PG.
S1: S2 Genre: GEN Origin: ORG
First Line (poetry): "FLINE"

We also had a Report Form for out-putting Tables-of-Contents (slightly enhanced) for a given issue or a run of a magazine/journal. This consisted of:

SNAME, FNAME, "TI," [GEN], PG [MON YR]

You can see that having multiple fields allowed for a great deal of flexibility in creating reports/output. It also allowed for some unusual queries with some very detailed and specific filters. For example, we could ask for lists of all the love poetry published in 1888 by specifying year, genre, and "love" in S2. This kind of unusual filtering of the data was a product of our efforts to create a highly flexible query-capacity in order to accommodate a wide variety of possible users. I once had a social historian wanting analyse advertisements; we saved him hours of time.

C. Database Field Structure:

The raw data that I am sending you is in comma delimited format, sequentially arranged in 17 fields; there is one full record/entry per line. Here is the sequence of the data using the field names I used in creating the original database structure and indicating the original field lengths (in characters). These field names are not part of the raw data file. [Incidentally, if it is of any use to you, I currently have working model of the database in Inmagic DB/Textworks software. This is a different kind of database structure, so the Dump Files do not work with most other forms of database software. Comma delimited format is more universal.]

Field Name	Char	Note
JNL	[3]	- each "Journal" was given a 3-letter code to simplify queries and (when combined with NDX) to form a unique identifier for each record
NDX	[4]	- each record of each journal has an "index" number beginning at 0001 which, when combined with the JNL code, forms a unique identifier within the database as a whole
SNAME	[15]	- author's "Surname" (for given article)
FNAME	[15]	- author's "First name(s)" or initials
TI	[65]	- "Title" of article
MAG	[60]	- literal title of "Magazine" in which article appears (ie, longer and more recognizable version of JNL)
VOL	[2]	- "Volume" of Magazine/Journal
NUM	[2]	- volume "Number" of Magazine/Journal
MO	[3]	- "Month" (3-letter code)
YR	[4]	- "Year"
PG	[9]	- "page" reference (eg., 127-215)
S1	[20]	- "Subject" term 1 (usually a broad category)
S2	[20]	- "Subject" term 2 (usually a narrower or more specific term)