

A Microfilm Information Retrieval System for Newspaper Libraries

Kenneth Janda and David Gordon

Northwestern University, Evanston, Illinois 60201

■ Advances in information technology promise significant changes in the operating environment of tomorrow's newspapers. One specific development in information technology deserves consideration for immediate use by moderate-sized newspaper libraries. Newspaper clippings, and index codes that describe those clippings, can be stored on 16mm film and searched rapidly and automatically with the MIRACODE information retrieval system. The system was applied to 1,000 clippings on file in the Medill School of Journalism. Operational problems in preparing index codes and coding the clippings are discussed. The retrieval capabilities of the system are outlined in examples of newspaper library research strategies.

RECENT developments in information technology offer new approaches to the old problems of indexing, storing, searching, and retrieving newspaper clippings—the most important and most unwieldy holding of newspaper libraries. Photographic reduction of newspaper clippings on microfilm has long offered an inexpensive solution to the storage problem itself, but conventional microfilm technology does not provide for effective search and retrieval of clip-

pings stored on film, even when assisted by a good index to the contents of the films.

At a considerably greater expense, modern computer technology allows for rapid retrieval of clippings indexed according to information selected for keypunching. Computers by themselves, however, do not solve either the indexing or the storage problems. Until the use of computers is more widespread in the typesetting process, even a computerized indexing process will involve the major extra step of keypunching the key words for the index. For those newspapers which use tape to drive their Linotypes, it becomes a simpler matter to feed that tape into a computer programmed to select the index entries.

For those who can afford it, the storage, indexing, and search-retrieval problems can be solved with specially designed integrated computer-and-microfiche systems, similar to the ones being developed for *The New York Times* (1) and planned for the *Toronto Globe and Mail* (2). But these solutions are not yet perfected and may not be within the financial resources of most newspapers. In the meantime, another approach promises a simpler and less expensive alternative.

One specific development in information technology provides a relatively low cost but highly effective solution to the information storage and retrieval needs of the more numerous moderate-sized

newspaper libraries. This technology uses microfilm to store *both* the clippings *and* the index codes that describe the clippings. An electric eye in the microfilm reader scans the codes on the film and controls retrieval and display of desired clippings. This technique is employed in equipment manufactured by the Eastman Kodak Company and marketed under the name, MIRACODE, an acronym for MICROFILM RETRIEVAL ACCESS CODE. The MIRACODE system is presently utilized in a worldwide study of political parties at Northwestern University, where the equipment is used to search magazines of 16mm film containing more than 40,000 pages of coded information photographed from some 2,500 documents on party politics in 36 countries (3).

Success with the system on the party politics projects has led to its application to other research topics. We have recently concluded a pilot study of the MIRACODE technology applied to clipping files assembled by Northwestern students in the Medill School of Journalism. The MIRACODE system is far less sophisticated than the proposed information bank of *The New York Times*. It is also far less expensive and much less complicated to operate. On all three counts, it may well be more in line with the current needs of many newspapers and their staffs. This system—or any system—should be viewed as one facet of the much larger problem of how to provide and effectively retrieve all the information which should be available to those who disseminate information to the public.

The MIRACODE system deserves attention especially for its advantages over certain drawbacks in contemporary computer technology: Its use of microfilm storage enables the system to handle large amounts of textual material without abstracting or endless keypunching. It allows direct man-machine interaction with browsing capabilities that are impossible with computer systems operating in a batch-processing mode. Furthermore, its relatively low purchase price (less than \$30,000 for the basic system) permits its consideration as a system for tasks which could not justify the expense of renting

and operating a suitable computer. Finally, it has some of the same powerful searching capabilities of a computer, employing Boolean logic on machine-readable optical codes.

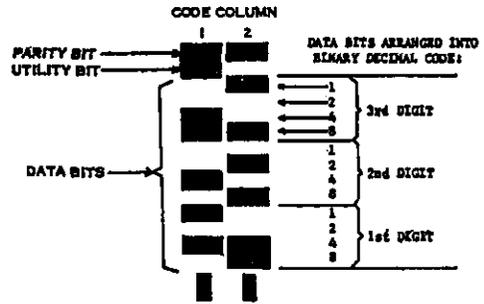
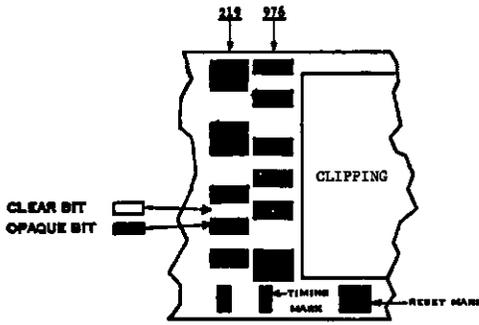
For papers which have run out of morgue space, or which have been looking for a better way to handle the materials they clip and file, the MIRACODE system may offer a manageable first step into the information retrieval field. This paper describes the methodology in detail and presents the results of our pilot study.

The Miracode System

The advantages of microfilm for recording and storing large files of material have long been recognized, but no method has heretofore been provided for effective retrieval of information once recorded on film. To be sure, an index can be prepared to show the content and location of information on film, and the appropriate reel can be selected and run through a microfilm reader to zero in on the desired frame by visually checking sequence numbers on the film. But this method of retrieval is too crude and cumbersome for many research purposes, which require an ability to retrieve material swiftly and automatically according to logical connections among subjects being discussed, for example, retrieving only discussions of Subjects A and B that do not mention Subject C. Previously, such logical searching capability was available only through digital computer technology. The MIRACODE system, however, incorporates electronic circuitry to detect logical combinations among machine-readable code numbers associated with the input material.

The basic components of the MIRACODE system are a special 16mm microfilm camera and a microfilm reader. Material is prepared for the MIRACODE system by coding the items (in our case, newspaper clippings) according to some structured scheme and keypunching all codes for a given clipping on a matching punchcard. In our study, for example, each clipping was tagged with three-digit

Figure 1. Coding for Miracode Film. Two, Three-Digit Codes Recorded in Binary Form



Binary Decimal Code for Column 1:

$$\left. \begin{array}{l} \text{3rd Digit} \quad 8 + 4 - 3 = 9 \\ \text{2nd Digit} \quad 4 \quad - 3 = 1 \\ \text{1st Digit} \quad 4 + 1 - 3 = 2 \end{array} \right\} 219$$

Binary Decimal Code for Column 2:

$$\left. \begin{array}{l} \text{3rd Digit} \quad 8 + 1 - 3 = 6 \\ \text{2nd Digit} \quad 8 + 2 - 3 = 7 \\ \text{1st Digit} \quad 8 + 4 - 3 = 9 \end{array} \right\} 976$$

Note that all binary decimal code patterns are "excess three." The parity bit is required if the data bits plus the utility bit is an even number.

codes identifying its date of publication; the newspaper of origin; source of authorship; main, secondary, and tertiary topics; and the main, secondary, and tertiary individuals or organizations named in the story.

The camera is connected directly to a modified IBM 026 keypunch. At the microfilming stage, the clippings are photographed immediately after the corresponding punchcards have passed under the "read" station of the keypunch. Code numbers read by the keypunch machine are transmitted to the camera where they are translated into a binary pattern of clear and opaque rectangles recorded on film immediately in front of the corresponding clippings.

Each vertical column of rectangles on the film contains one three-digit code number, which can be deciphered as

shown in Figure 1. The clippings and the codes are reproduced on the film in accordance with the diagram in Figure 2. The codes can be searched by the MIRACODE reader to locate a specific code regardless of the column in which it appears, or the machine can be instructed to search for codes only if they exist in certain columns and not in others. In keeping with comparable computer terminology, the first method of coding and searching is called "free format" and the second is "fixed format." Technical considerations dictated coding the Medill newspaper clippings in fixed format (4).

At the retrieval stage, the binary codes on the film are sensed by an optical scanning device on the MIRACODE reader, which reads the codes flashing by the scanning head at the normal film transport speed of 10 feet per second. The reader or "retrieval station" has the capability of testing for logical relationships among as many as 15 different three-digit codes as the film passes the optical scanner. The available logic for MIRACODE searches consists of a full set of relations: AND, OR, NOT, GREATER THAN, LESS THAN, and EQUAL TO. A specific code is searched by pressing down the appropriate keys at the MIRACODE keyboard. The keyboard is modular in design, allowing from one to a maximum of 15 banks of keys to operate at a retrieval station. A typical keyboard configuration involves six banks of keys which permit testing for logical relations among six

three-digit codes. These codes may be interpreted in a manner that permits encoding and retrieval of *alphabetic* information (for example, proper names) as well as *numeric* information.

A search command is communicated to the reader by pressing the SEARCH button, which starts the film transport. When the machine senses the appropriate relationship among the numbers entered on the keyboard, the film immediately comes to a halt and backs up several frames to display the image retrieved by the search command. The operator has the opportunity to examine the page being projected for its relevance to his request. Should a hard copy be desired, a black-on-white print can be made from the projected image in 25 seconds.

If the image retrieved does not satisfy the user, the search can be continued by pressing the SEARCH button again. The film will advance and stop to project the next image on the film that satisfies the search command. Rewinding occurs automatically when the end of the film is reached. If desired, the retrieval station can be set to operate automatically, printing each page on the film that satisfies a given search command.

While seated at the MIRACODE reader-retrieval station, the user can interact with his data files by changing his search command to alter the character and amount of information retrieved. He can determine in advance of retrieval of specific clips how many "hits" he will get for any given command through the operation of an optional device called a "response monitor," which reads the film and tallies the number of satisfied conditions without actually stopping to display the retrieved images. This tally is instantaneously displayed on an electronic "scoreboard" as the film is read. To decrease the number of "hits" and to make his search more selective, the user can enter additional codes—depending on the number of banks of keys available on the unit. To increase the number of "hits," he can relax the search command by turning off the small toggle switch associated with each bank of keys, thus removing codes from the search.

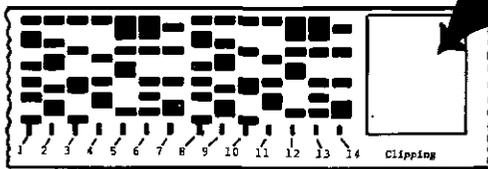
The Miracode System and the Medill Newspaper Morgue

The Medill School of Journalism maintains a file of clippings from the campus newspaper, *The Daily Northwestern*, and the local semiweekly, *The Evanston Review*. These papers are clipped by the beat reporters in the general reporting course, and the clippings are conventionally filed in a series of topical folders that have developed over time to accommodate news items from these two papers. A total of 1,021 items (clipped by the class in the 1969 Winter Quarter) constituted the material for the MIRACODE pilot project.

In previous quarters the students were requested to file each clipping in the appropriate file folder according to its main subject. This time they were asked to mount each clipping on an 8.5" × 11" sheet of paper and to fill in the information requested at the top, where spaces were provided for index codes. These sheets were collected approximately once a week by a student who keypunched the codes into cards and microfilmed the clippings (5). No attempt was made in this pilot study to have the microfilming keep pace with the clipping so the students could use the retrieval capabilities of the MIRACODE system in their own reporting activities. We were content at this stage to take our time in trying out procedures and testing the usefulness of the coding categories and retrieval concepts. This report will be limited to an evaluation of these dimensions of the project.

Obviously, it is crucial that the clippings be properly coded for effective retrieval. Constructing useful coding categories and training people in their proper application constitute difficult but not insurmountable problems. Relatively little time was spent in formal training as part of this project, but most class members picked up the rudiments very quickly (assisted by individual conferences where check-coding turned up repeated errors). No data was collected on the accuracy of the coding work, in large part because check-coding turned up a decreasing

Figure 2. Miracode Film Format.



Glass' resignation provides Cats' goal

By **FRED TAYLOR**

The unexpected resignation of Northwestern head coach Larry Glass has seemingly given the Wildcats additional incentive to close out the season as winners.

Although losses in four of its last five Big 10 games have eliminated Northwestern from title contention, the Wildcats are a team with a definite goal.

"We want to give coach Glass a going-away present," is the way NU forward Don Adams puts it. "We also want to give Brad Snyder a coming home present.

BRAD SNYDER has been an assistant basketball coach at Northwestern since 1961, and he is one of the coaching prospects being given "heavy consideration" by Athletic Director Tippy Dye.

| Column | Item | Entry |
|--------|----------------------------|--|
| 1 | Date (3 digits) | 217 (Month & Day) |
| 2 | Author (3 digits) | Code for Fred Taylor |
| 3 | Main topic (3 digits) | 961 |
| 4- 6 | Main person's name | Glass, L. (1st 7 letters plus initials) |
| 7 | Secondary topic (3 digits) | 962 |
| 8 | Tertiary topic (3 digits) | 960 |
| 9-11 | Secondary person's name | Snyder, B. (1st 7 letters plus initials) |
| 12-14 | Tertiary person's name | Adams, D. (1st 7 letters plus initials) |

number of errors as the students gained familiarity with the process. One problem that did show up repeatedly, however, was the tendency of students to code clippings on general events into categories defined by their own parochial "beat" perspectives. For example, clippings of an incident involving black-white violence and minor injuries at a fraternity house were originally coded into primary topics ranging from "police matters" through "racial relations" to "student health service." Most of these differences were caught and coded into consistent topical categories before being encoded on the microfilm.

Codes were developed for the month and date of each clipping, for its authorship, and for the three main topics of the story. The date of publication was easily coded in the three-digit structure of MIRACODE by having the first digit stand for the month and the next two digits represent the day. This coding convention works fine through September (Month 9), and it can handle Months 10, 11, and 12 by activating the extra "utility bit"

accompanying each coding column (see Figure 2). With the utility bit "off" (or on "0"), the date code "121" would stand for "January 21"; with the bit "on" (or on "1"), the same code would stand for "November 21." For our purposes, we did not find it necessary to code the year since all clippings fell within a three-month period in 1969. It is unlikely that the year would be coded in any working system, for the clippings could be organized and indexed on film magazines according to dates covered within a year. The span of time accommodated by one 100-foot film magazine would depend on the volume of clippings and the depth of the indexing codes applied to the clippings. For our system, we allowed 14 columns of three-digit codes per item, which required 2-1/8 inches of film for the codes plus the clipping. Thus, we produced about 180 feet of film for some 10 weeks of clippings from two newspapers, and we needed two magazines to hold our film.

No specific code was developed to differentiate between the newspapers in

this project, but we built the differentiation into the authorship codes. These codes were set up to allow separate coding for each byline, plus differentiation by types of clippings from each paper. The authorship codes are summarized in Table 1, along with the usage statistics, which we obtained as a by-product of keypunching the codes for input to the MIRACODE microfilmer. The punchcards were later fed into the university computer to determine exact counts of the codes used to index the clippings—which, incidentally, produced for the first time

Table 1. Author Codes.

| Codes | General Authorship Category | No. of Clippings |
|---------|---|------------------|
| 001-017 | By-lined Reporters— Review | 83 |
| 041-072 | By-lined Reporters— Daily | 206 |
| 200,216 | Unby-lined Local Story—Review | 411 |
| 201,217 | Unby-lined Local Story—Daily | 236 |
| 208,209 | Story from Medill Carbon—Daily & Review | 10 |
| 300,302 | Photo and Caption— Review | 34 |
| 301,303 | Photo and Caption— Daily | 24 |
| 400,410 | Editorial—Review | 3 |
| 401,420 | Editorial—Daily | 2 |
| 450 | Letter to the Editor— Review | 6 |
| 451 | Letter to the Editor— Daily | 3 |
| | Total* | 1,018 |

* Three items which were punched erroneously into non-existent authorship codes were dropped from this tabulation, rather than sorted out, repunched, and refilmed. These errors were not discovered until after both reels of microfilm had been processed; in an on-going operation, such items would be corrected and refilmed easily. Here, this was impossible because filming operations had already been completed.

a complete profile of the items clipped by the reporting class and constitutes, in itself, a summary of the information in the newspaper morgue for 1969.

Clippings were filmed in the order of primary topic designation in this project, rather than by date of publication or by type of clipping. This procedure was followed simply because of its greater convenience in processing the clippings, and because the search capability would not be materially affected with only two magazines. However, where more than one or two reels of microfilm are involved, we strongly suggest that the clippings be microfilmed chronologically, and, if possible, segregated onto separate reels for each publication involved. Once again looking toward the development of a separate index, this procedure should make the selection of the proper film reel most simple and direct.

The topic codes for this project were based on the filing system that had developed in the Medill morgue during previous years. No major effort was made to refine the codes for the various morgue classifications, although some relatively minor changes and additions were made during the quarter at the suggestion of various students to simplify their coding task. Each clipping was coded according to its primary topic (and, if applicable, its secondary or tertiary topics) as selected from a full set of codes which are outlined by major categories in Table 2 and reproduced in Table 3.

The specific coding scheme used in our project is not a central issue in the MIRACODE methodology. Indexing codes will necessarily vary from paper to paper—especially if they are to be applied to preexisting morgues which cannot be easily restructured to fit any "model" code. The categories used here can serve as an example of how a code can be developed to fit an already-existing morgue system and how some minor restructuring can be done without disrupting the entire operation. It should be noted that, while the topic codes in this pilot project were restricted to three digits, the MIRACODE equipment can be modified to allow codes of six or even nine digits,

should the material require such elaborate codes (6).

The system proved flexible enough to permit the encoding of the actual *names* of individuals and organizations in the news as primary, secondary, or tertiary subjects. The coding form allotted 9 spaces to each name: 7 spaces for the first 7 characters of the last name and 2 spaces for initials (7). This alphabetic information was keypunched into the card along with the rest of the numeric codes and then encoded on the film by adopting a convention of letting numbers stand for letters of the alphabet, similar to the telephone exchanges on a push-button phone. Technical considerations dictated that we use the following coding scheme:

| Letter | Code | Letter | Code | Letter | Code |
|--------|------|--------|------|--------|------|
| A | 1 | J | 1 | S | 0 |
| B | 2 | K | 2 | T | 0 |
| C | 3 | L | 3 | U | 0 |
| D | 4 | M | 4 | V | 8 |
| E | 5 | N | 5 | W | 8 |
| F | 6 | O | 6 | X | 8 |
| G | 7 | P | 7 | Y | 8 |
| H | 8 | Q | 8 | Z | 8 |
| I | 9 | R | 9 | | |

Because the same number stands for more than one letter of the alphabet, this coding convention does not provide for fool-proof retrieval by proper names and some "noise" or irrelevant "hits" will undoubtedly be produced by chance when the system is used to search for clippings about a certain person. But we felt that this search capability would be so valuable for journalistic research that the user of the system would gladly tolerate some "noise" in the system in order to retrieve clippings by searching names of people and organizations in the news.

The largest change in the coding process made during the project came in connection with the use of the names of athletes. At the suggestion of one of our sports reporters, we decided to allow encoding of an athletic team name rather than an individual's name in some cases. This change simplified the logical retrieval of items dealing with teams lacking any well-known individual members (for instance, the Purdue wrestling team;

by contrast, stories about the Purdue basketball team were invariably coded to refer to **MOUNT, R.** or **GILLIAM, H.**, in addition to **PURDUE**).

Morgue Research Strategies with Miracode

The MIRACODE system gives us the ability to search the file to retrieve clippings according to any single code or any combinations of codes on the film. For example, the system has the power to search for a by-lined clipping by a particular reporter on a specific subject for a certain date. The system also has the logical capability to search for all the clippings on a given topic that appeared

Table 2. Subject Codes.

| Codes | General Subject Category* | No. Times Used† |
|---------|--|-----------------|
| 001-017 | Politics | 118 |
| 100-199 | Local Government | 233 |
| 200-290 | Evanston Schools | 237 |
| 300-383 | Northwestern University | 742 |
| 401-444 | Community Services, Neighborhood Organizations | 35 |
| 460-473 | Churches | 8 |
| 500-550 | Community Relations | 38 |
| 600-637 | Business | 8 |
| 700-730 | Fine Arts | 72 |
| 800-852 | Geographic & Organizational Divisions | 22 |
| 900-901 | Circuit Court-Local | 7 |
| 910-919 | Health | 68 |
| 950-975 | Sports (ETHS‡ & NU) | 253 |
| 999 | Obituaries | 4 |
| | Total | 1,845 |

* For a more detailed breakdown, see Appendix.

† "Number of Times Used" will add up to more than the number of clippings because most clippings were coded for more than one topic.

‡ Evanston Township High School.

either *before* or *after* a certain date, or even *between* two dates. Assuming that the file could be properly organized to focus the search on one film magazine containing about 500 clippings, the reporter could locate and view his clippings in a matter of seconds. Even if several magazines were involved in the search, the time for searching thousands of clippings would be calculated in minutes. Once the desired clippings were located, a black-on-white copy could be produced in a matter of seconds for a cost of about \$0.08 per copy, thereby enabling the reporter to have his own copy of the clipping while the integrity of the clipping file is maintained for the use of others. The rather substantial cost of a retrieval station—around \$15,000, depending on the type and number of keyboards needed—restricts the number of readers that a paper can afford to acquire. It is likely that the library staff would play a key role in receiving search requests which they would fill by delivering copies to the reporter.

Two typical examples of possible morgue strategy with the MIRACODE system will be traced out for hypothetical reporting assignments. The simpler of the two might well be assigned to a beginning reporter, or to a general assignment or police reporter recently promoted to a new beat. This would involve background on all recent developments on a particular subject, for instance, steps taken by the Evanston municipal government toward electronic data processing. A quick search of the Morgue Codes under **Local Government, General** shows that category No. 111, **City Data Processing**, is relevant. The retrieval station would then be requested to search all microfilm containing clippings between the desired dates, for topic code No. 111, which could appear in either Column 3, 7, or 8. Depending on the number of keyboards available on the particular MIRACODE retrieval station, this could take from one to three passes of the film. It would not be necessary to search for any other codes, since proper coding procedures would require that any data processing clips previously

Table 3. Coding Categories* and Frequency of Use by Primary, Secondary, and Tertiary Importance to Story

| | <i>Prim.</i> | <i>Sec.</i> | <i>Tert.</i> |
|--|--------------|-------------|--------------|
| POLITICS (001-099) | | | |
| 001 U. S. Senate locally | 1 | | |
| 002 Congressmen locally | 5 | | |
| 003 State Senate locally | | 3 | |
| 004 State House locally | 4 | | 1 |
| 006 GOP Organizations | 5 | 2 | |
| 007 Evanston Young Republicans | 2 | 1 | |
| 008 GOP/Dems on NU Campus | 10 | | 1 |
| 011 Democratic Organizations | 3 | 5 | 3 |
| 012 League of Women Voters | 2 | 1 | 1 |
| 013 General Election 1968 | | 1 | 2 |
| 014 Republicans 1969 | | 10 | 3 |
| 015 Democrats 1969 | 1 | 8 | 3 |
| 016 Politics—General 1969 | | 8 | 6 |
| 017 Aldermanic Elections | 23 | 3 | |
| LOCAL GOVERNMENT, GENERAL (100-129) | | | |
| 100 Evanston Township | | | 2 |
| 101 Evanston City Clerk | | 1 | |
| 102 Mayor | 2 | 8 | 1 |
| 105 City Council, General | 11 | 12 | 10 |
| 106 City Council Agenda | | 1 | |
| 107 City Council, NU | | 3 | |
| 108 Administration Committee | 5 | 6 | 1 |
| 109 Budget-Finance | 1 | | |
| 114 Public Works Department | | | 1 |
| 116 Evanston Civic Center | 1 | | |
| 128 ABM's 1969 | 28 | 1 | 1 |
| 129 Con-Con 1969 | | 2 | |
| PLANNING (130-154) | | | |
| 130 Planning and Development Committee | 9 | 5 | |
| 131 Planning and Development, Passed or Denied | 1 | | |
| 132 Planning and Development, Pending | | 2 | |
| 134 Zoning Board of Appeals | 2 | | 2 |
| 135 Zoning Amendment Committee | 1 | | 2 |
| 136 Plan Commission | 3 | 1 | |

* In order to save space in this table, we have reproduced only those coding categories that were used at least once as a primary, secondary, or tertiary code for a clipping. Those that were not used at all were not listed here. This accounts for most of the gaps in the numbering.

| | <i>Prim.</i> | <i>Sec.</i> | <i>Tert.</i> |
|--|--------------|-------------|--------------|
| 137 Planning and Conservation Department | | 1 | 2 |
| 139 Pollution: Air, Lake | | 1 | |
| 140 Building Department | | 1 | |
| 142 Sign Ordinance Controversy | 5 | 1 | |
| PARKS AND RECREATION (160-169) | | | |
| 160 Parks and Recreation Board and Superintendent | 4 | 2 | 2 |
| 161 Parks and Recreation Activities | 5 | 6 | |
| 162 Evanston Youth Commission | 1 | 2 | 1 |
| 164 Nichols Lighted School Center | 2 | | |
| 165 Evanston Student Union | 1 | | |
| 166 Outreach Workers | | 1 | |
| 168 Evanston Youth Conference | 2 | | |
| POLICE-FIRE DEPARTMENTS (170-199) | | | |
| 170 Police Department 1969 | 1 | 2 | 6 |
| 171 Police/Fire Reports | 1 | | 1 |
| 172 Police—Administration, Personnel | 2 | 1 | 1 |
| 174 Police—Community Relations | 2 | 1 | |
| 175 Police—Traffic | 5 | | |
| 176 Police—Burglaries and Thefts | 6 | 1 | |
| 177 Police—Assault, Battery, Felonies | 3 | 5 | |
| 179 Police—Gambling, Narcotics, Misc. | 3 | 2 | 1 |
| 181 Police—Juvenile | 4 | 3 | |
| 182 Police—Statistics, Overall | | 1 | |
| 183 Police—All other stories | 1 | 1 | 3 |
| 190 Fire Department Administration, Personnel and Training | 1 | 1 | |
| 191 Firefighters Association Local 742 | 2 | 2 | |
| 193 Fire Department Stories | 4 | 1 | |
| EVANSTON SCHOOLS (200-299) | | | |
| 200 District 65 Caucus and Elections | 10 | | |
| 201 District 65 Board of Education | 10 | 12 | 1 |
| 202 District 65 Superintendent and Administrative Staff | 7 | 6 | 3 |
| 203 District 65 Statistics, Budget '68-'69 | 3 | 2 | 2 |
| 204 District 65 Elementary Schools | | 3 | |

coded into more general morgue categories would have been recoded and re-filmed when the **City Data Processing** file was set up (8).

A more sophisticated morgue strategy would be required if the assignment were to write a story on an Evanston zoning controversy involving a high rise apartment building proposed by the James Brothers (a development corporation rather than the notorious outlaws). This morgue strategy would require the logical capabilities of the system to retrieve only those clippings on specific topics which include the name of a certain person. The MIRACODE system possesses the necessary capabilities making possible such retrieval commands as **Planning and Development, Passed or Denied** (Code No. 131) in Columns 3, 7, or 8 combined with **James** in Columns 4-6, 9-11, or 12-14. Any other combinations of *subject and name, appropriate to a particular morgue filing system, can be retrieved in a similar fashion* (for example, **Expo and Drapeau** or **Convention and Daley**).

Every clipping on the zoning controversy contains the name of at least one of the James Brothers as one of the three main persons mentioned in the story. A quick check of the statistical analysis of the clipping file compiled in this project shows that, of the eight possibly relevant topic categories listed under **Planning** (Topic Categories 130-143), there were stories actually coded with references (primary, secondary, or tertiary) to six (Categories 130, 131, 132, 135, 136, and 137—see Table 3 for the specific category listings). In addition, there were clippings coded for Category No. 105, **City Council, General**, which is the only other place where a story on this controversy could logically be filed (9).

The strategy, then, is to search for all clippings coded 105, 130, 131, 132, 135, 136, and 137 which contain a reference to the James Brothers in Columns 4-6, 9-11, or 12-14 (primary, secondary, or tertiary references to people in the story). Had a printed index of names been prepared for these clippings, it would have shown six entries under **James** (10).

Because of our limited number of clippings, and the fact that only two reels of film were used in this pilot project, we in fact discarded the combined topic-name search strategy in checking out this hypothetical search. Instead we simply asked the station to retrieve all clippings which mentioned someone named James among the three main people in the story. We found that there were four clippings which related to the James brothers zoning controversy during the period of this project. All referred to E. James as the primary person mentioned in the story, and one of these four cited his brother, A. James, as the secondary person. So much for five of the six references. The sixth was a reference to one Mike James, who was involved in a totally different subject and was irrelevant to our hypothetical search. Had we in fact continued our retrieval strategy of topic-name combinations, this reference would never have shown up, because the subject code for the Mike James reference had to do with a campus speaker rather than zoning.

To recapitulate on the ideal search strategy in this situation:

▶ An index of names coded into the microfilm in the morgue would be consulted to determine the total number of possible references to the person or people under consideration.

▶ If that number is relatively small, the quickest retrieval strategy would be to ignore the possible topic codes and simply check out all the possible clippings, as was done here.

▶ If that number is large, further selectivity should be employed in the search by combining possible topic-codes with the desired name—for example, **Planning and Development, Passed or Denied**, combined with James.

Obviously, the more specific the topic codes, the easier and more selective that retrieval will be. But this is a problem which must be solved for all automated retrieval systems, and is a major point at which the librarian is vital to the smooth functioning of the entire process.

One other problem might well be

Table 3 (contd.)

| | <i>Prim.</i> | <i>Sec.</i> | <i>Tert.</i> |
|--|--------------|-------------|--------------|
| 205 District 65 Junior High Schools | | 1 | 6 |
| 206 District 65 PTA's | 1 | | 1 |
| 207 District 65 Outlooks | | 1 | |
| 208 District 65 Integration Plan | | 3 | 1 |
| 209 District 65 Lab School | 3 | 1 | |
| 211 District 65—Dual Enrollment | 6 | | |
| 231 Parent Teacher Council of Evanston | | 1 | |
| 240 ETHS—1969 | 1 | 1 | 2 |
| 241 District 202 Board of Education—Caucus and Elections | 9 | 3 | 1 |
| 242 District 202 Superintendent and Administrative Staff | 5 | 8 | 13 |
| 243 District 202 High School Community Relations | 16 | 19 | 8 |
| 244 District 202 Construction and Operation | 3 | 1 | 2 |
| 245 District 202 Courses, New Curriculum | 5 | 4 | 4 |
| 246 District 202 Student Activities | 19 | 10 | 2 |
| 247 District 202 PTA | | | 1 |
| 249 District 202 Miscellaneous | 1 | 3 | 6 |
| 270 North Shore Junior College | | 2 | |
| 273 National College of Education | | 1 | 2 |
| 275 Miscellaneous Schools | | 1 | |
| NORTHWESTERN UNIVERSITY (300-399) | | | |
| 300 Northwestern University | 1 | 14 | 2 |
| 301 President and Board of Trustees | 9 | 3 | 2 |
| 302 Development and Planning | 2 | 2 | |
| 303 First Plan for the '70's | 3 | 1 | |
| 304 Administration—Business | 4 | 7 | 1 |
| 305 Administration—Construction | 7 | 4 | 4 |
| 306 Buildings and Grounds—Security | 5 | | 1 |
| 307 NU Housing | 25 | 12 | 3 |
| 309 Administration—Academic—Dean of Faculties | 2 | 3 | 1 |
| 311 Admissions and Financial Aid | 2 | 1 | |
| 312 Vice President for Student Affairs | 5 | 11 | 14 |
| 313 Dean of Students Staff | 3 | 11 | 9 |



| | <i>Prim.</i> | <i>Sec.</i> | <i>Tert.</i> |
|--|--------------|-------------|--------------|
| 314 Student Activities | 1 | 24 | 8 |
| 315 FCC Pot Party Hearing | | 3 | |
| 316 Council on Undergraduate Life | 8 | 7 | 2 |
| 317 University Discipline Committee | | 6 | 2 |
| 318 Parietal Hours | 9 | 10 | 2 |
| 319 Racial | 9 | 7 | 1 |
| 321 Negro Takeover, '68 | | 1 | 2 |
| 323 Student Senate | 12 | 11 | 7 |
| 324 Associated Women Students (AWS) | 19 | 9 | 2 |
| 325 Interfraternity Council | 12 | 4 | 3 |
| 326 MOC and WOC | | 1 | |
| 327 MRHA | 4 | 2 | 6 |
| 328 NU Student Living Groups 1969 | 27 | 16 | 5 |
| 329 Deferred Rush | 10 | 5 | 1 |
| 330 Greek Organizations (Including Panhellenic) | 5 | 3 | 2 |
| 331 Non-Greeks | 1 | 1 | 2 |
| 332 Clubs, miscellaneous | 3 | 2 | 1 |
| 333 Students for a Democratic Society (SDS) | | 13 | 4 |
| 335 Chaplain, Campus Religious Groups | | | 1 |
| 338 Film Society | 2 | | |
| 339 Orgy of the Arts | 8 | | |
| 343 Placement Center | | 2 | 1 |
| 344 Board of Publications | | 1 | |
| 345 Daily Northwestern—Stories about | | | 1 |
| 346 Recruitment on Campus (Including Dow) | 16 | 6 | |
| 347 Symposium | 29 | | |
| 348 NROTC | 4 | | |
| 350 NU Vietnam | | 2 | 2 |
| 351 WAA—MU | 4 | | |
| 354 Alumni Office | | 1 | 1 |
| 355 Alumni News | 4 | | |
| 356 Faculty, General Faculty Committee, Faculty Club | 2 | 2 | 1 |
| 357 Faculty Senate | | | 2 |
| 359 NU—City | | | 1 |
| 360 National Groups on Campus | | 10 | 4 |
| 362 Class Councils '68-'69 | | | 1 |
| 367 Curriculum Changes—Proposed and Actual—SAR, etc. | 11 | | |
| 368 FMO and Other Black Student Activities | 3 | 1 | |
| NU SCHOOLS: | | | |
| 370 Arts and Sciences, no physical science | 24 | 9 | 7 |
| 371 Arts and Sciences, physical sciences | 7 | 11 | 3 |

mentioned here. The stories which gave our students and our check-coders the biggest headaches were the general roundup stories on an event, or an election race, in which eight or nine names were mentioned with virtually equal prominence. In such cases, it was extremely difficult to single out the three primary people mentioned, and some sort of arbitrary approach (incumbency, chronology of mention, etc.) was adopted. In practice, however, the retrieval of such stories would probably not be greatly hampered, since they are more likely to be searched under a topical code rather than by using the names of relatively not so prominent individuals. Certainly, the problem is no worse than the question of where to file and how to retrieve similar stories manually.

Conclusion

As noted earlier, the MIRACODE system provides a potentially useful alternative to the current dilemma of operating with a hand-sorted file of clippings, or jumping into computer operations with both feet. It avoids the necessity of extensive keypunching to record printed information in machine-readable form for on-line retrieval. This system, we believe, can serve as a transition from manual operations to more sophisticated methods of information storage and retrieval for the newspaper morgue of the future. In the process, it can bring more order and, hopefully, greater efficiency to the morgue.

Specifically, there are two major advantages we see in this system which are worth mentioning. One is the elimination of duplicate clipping operations. The multiple topic coding available in this system (and the cross-indexing possible for a topic-proper name combination or for other combinations where logical relations are important) make it unnecessary to clip more than one copy of a story. For those newspaper libraries which are currently employing up to half a dozen people to clip and mark different papers, and are saving multiple copies of the same story for cross-filing

purposes, the savings of this system in terms of both human effort and storage space are worth investigating.

Secondly, and perhaps most important, one of the major weaknesses of most newspaper libraries could easily be overcome through this system. That weakness is an absolute ignorance on the part of any one individual—or group of individuals—about the exact holdings of the morgue. Because of the availability of the punched cards as a by-product of the coding process, an up-to-date index can be produced relatively easily through the use of a program set up to print out the key words in the topic categories, the various proper names, and the number and chronology of the stories in the various categories (11). This index can also serve as a guide to the proper reel of microfilm, thus cutting search time down considerably—comparable to the search time now required in the most orderly of clipping file morgues.

The capability of continually updating an index is perhaps the most important benefit of the MIRACODE process. Both librarians and reporters can be kept informed of all currently available clippings on topics, people, or a combination of topics and people in the news. The system can be indexed to display all coded information, and has the retrieval capability to search for all categories of coded information, or for any individual categories. This combination of rapid indexing capability with the rapid and logically selective retrieval of microfilmed material should greatly increase the efficiency with which newspaper morgues can be used.

Perhaps equally important, the MIRACODE system requires only a relatively small amount of advance preparation time for implementation, and is easily adapted to existing filing systems. These are vital considerations if a newspaper morgue is bulging at the seams, and the only answer to the space problem is to discard and/or microfilm a large number of older clippings. This system is fairly easily understood and used, as indicated by the rapidity with which the coding operations were grasped by the college

Table 3 (contd.)

| | Prim. | Sec. | Tert. |
|---|-------|------|-------|
| 372 African Studies Center | 2 | | |
| 373 Center for Urban Affairs | | 2 | |
| 374 Business | 9 | | |
| 376 Education | 1 | | |
| 377 Graduate School | 1 | | |
| 378 Journalism | 1 | 3 | 1 |
| 379 Music | 24 | 2 | |
| 380 Speech | 25 | 8 | 2 |
| 381 Technological Institute —General | 2 | | |
| 383 Technological Institute —Engineering | 4 | | 1 |
| UNITED COMMUNITY SERVICES, UNITED FUND (400-429) | | | |
| 401 United Community Services | 1 | 14 | 1 |
| 402 United Fund | | | 1 |
| 404 Arden Shore | 1 | | |
| 406 Camp Fire Girls | 1 | | |
| 407 Child Care Center | 1 | | |
| 415 Visiting Nurse Associa- tion | 1 | | |
| 417 YMCA | 5 | | |
| 418 YWCA | 6 | 1 | |
| NEIGHBORHOOD ORGANIZATIONS (430-459) | | | |
| 430 Neighborhood Organiza- tions—General | | | 1 |
| 444 West End Neighbors | | | 1 |
| CHURCHES (460-489) | | | |
| 461 Churches/Ministers, General | | | 1 |
| 465 Baptist Churches | | | 1 |
| 467 First Congregational Church | 3 | | |
| 469 Rock River Conference, United Methodist | | 1 | |
| 472 Synagogues | | 1 | |
| 473 Secondary Churches | | 1 | |
| COMMUNITY RELATIONS (550-599) | | | |
| 500 Human Services Com- mittee | | 4 | 2 |
| 503 Fair Housing Ordinance | | | 1 |
| 504 Fair Housing Review Board. | 1 | | |
| 505 West Side Services Center | 8 | 3 | 4 |
| 520 Evanston Human Rela- tions Commission | 5 | 1 | 1 |
| 522 Evanston Human Rela- tions Council | 1 | | |



| | <i>Prim.</i> | <i>Sec.</i> | <i>Tert.</i> |
|--|--------------|-------------|--------------|
| 531 FOCUS—Federation of Citizens for an Unsegregated Society | | 1 | |
| 533 SCLC—Southern Christian Leadership Conference | | | 1 |
| 537 Urban League | | 1 | |
| 550 Evanston Racial—General | | 2 | 2 |
| BUSINESS (600-699) | | | |
| 600 Business—General | | 1 | 2 |
| 601 Chamber of Commerce | | 1 | |
| 602 Other Local Business Organizations | | 2 | |
| 603 Banks | | 2 | |
| FINE ARTS (700-799) | | | |
| 701 Evanston Fine Arts | 1 | 4 | 1 |
| 702 Evanston Art Center | 8 | | |
| 703 Community Concerts, Drama Club | 4 | | |
| 704 Evanston Symphony Orchestra | 1 | | |
| 705 Theatre 65 | 7 | 1 | |
| 706 Evanston Public Library | 10 | | |
| 708 Small galleries and clubs | 13 | | |
| 709 Lyric Opera Guild | 1 | | |
| 710 Evanston Woman's Club | 17 | | |
| 730 Evanston Historical Society | 4 | | |
| GEOGRAPHIC AND ORGANIZATIONAL DIVISIONS (800-899) | | | |
| 805 WCTU (Women's Christian Temperance Union) | 2 | | |
| 815 Cook County Offices in Evanston | | | 2 |
| 827 State Offices in Evanston or with Evanston Personnel | | | 1 |
| 840 Federal Offices in Evanston | | 2 | |
| 841 Office of Economic Opportunity | 3 | 1 | 1 |
| 842 Draft, Selective Service | | 1 | |
| 844 Navy Recruiting Office | 4 | | |
| 846 Air Force Recruiting Office | | 1 | |
| 847 Evanston Naval Reserve Recruiting Office | 1 | | |
| 852 Peace Groups | | | 3 |
| CIRCUIT COURT LOCALLY (900-909) | | | |
| 900 Circuit Court Locally | 4 | | |
| 901 Evanston Branch of Circuit Court, District 2 | 2 | 1 | |

juniors and seniors involved in this project. (Ease of retrieval depends on a general understanding of possible topic categories, and remains to be field tested.)

The one major additional manpower cost in this system, compared to straight manual indexing procedures, will be for keypunching operations. However, this should be offset by a reduction in staff time spent in clipping and/or filming more than one copy of a story, and then cross-filing these multiple copies manually in different locations. The filming of a single clipping in the MIRACODE system will provide cross references for retrieval purposes and a multiple indexing of the desired number of references to both major topics and major persons mentioned in the story. The process will also maintain the integrity of the morgue files, since the original material would never be taken out to a reporter's desk for use—and perhaps for eventual return to the files.

In summary, on the basis of this pilot project, we believe that the MIRACODE system holds considerable promise for use in newspaper morgues. It appears to provide economies of both space and manpower, a means for continually updating the index to the morgue holdings, and the advantages of easy conversion from manual operations. We submit that this may well be a logical step on the road from manual operations to full-scale computerization, and that its possibilities should be tested further by those newspapers which are not yet ready to take on the costs, lead time requirements, and sophisticated operating procedures of fully computerized operations.

Notes and Literature Cited

1. See *Editor and Publisher*, p.946 (Apr 5, 1969). Also, *ICRH Newsletter*, p.1-2 (Apr 1969), published by the Institute for Computer Research in the Humanities, New York University.
2. At present, the *Globe and Mail* is developing plans for indexing its news stories through the use of the tape used in typesetting. That paper's current intention is to combine the program it now uses for computerized indexing of photographic negatives with a modification of the program researched at the Massachusetts Institute of Technology

for the American Newspaper Publishers Association. The Toronto paper now uses Kalvar microfiche to store its clippings, but has not yet gone to computerized indexing or mechanical retrieval of anything but the photographic negatives. Letters to David Gordon from D. A. Rhydwen, Chief Librarian, Toronto *Globe and Mail* (Mar 7 and Apr 18, 1969).

3. The International Comparative Political Parties Project is supported by the National Science Foundation, Grants GS-1418 and GS-2533. A general description of the project's objectives and methodology can be found in Kenneth Janda, "Retrieving Information for a Comparative Study of Political Parties," in William J. Crotty, ed./ *Approaches to the Study of Party Organization*. Boston, Allyn and Bacon, 1968. p.159-215. This article is reprinted along with more technical material on the MIRACODE system and also on computer systems for information retrieval in Kenneth Janda / *Information Retrieval: Applications to Political Science*. Indianapolis, Bobbs-Merrill, 1968.
4. "Free-format" coding is used in the MIRACODE application to political parties research. In that project, any given page may be tagged with varying numbers of codes assigned to the page according to its content. At the retrieval stage, we want to find pages that discuss the topic or topics of interest, regardless of the location of the discussion on the page—hence the use of "free-format" codes and searching.
5. Anne Lechtenberg, assisted by Dana Whalen, patiently and efficiently keypunched and microfilmed the clippings. James and Jill O'Donnell helped considerably in editing the microfilms after processing.
6. For examples of a six-digit code applied to literature in another field, see John E. Rickert / *Urban Thesaurus: A Vocabulary for Indexing and Retrieving Urban Literature*. Kent, Ohio, Center for Urban Regionalism, Kent State University, 1968. This

Table 3 (contd.)

| | | Prim. | Sec. | Tert. |
|---|---------------------------------------|-------|------|-------|
| HEALTH (910-924) | | | | |
| 910 | Evanston-Northshore Health Department | 1 | | |
| 912 | Community Hospital | 2 | 2 | |
| 913 | Evanston Hospital | 14 | 4 | 1 |
| 914 | St. Francis Hospital | 7 | 2 | 1 |
| 915 | Mental Health Association | 13 | 2 | 1 |
| 917 | Hospitals/Deaths 1969 | 3 | 2 | 1 |
| 918 | Student Health 1969 | 10 | | 1 |
| 919 | Skokie Valley Community Hospital 1969 | | | 1 |
| EVANSTON TOWNSHIP HIGH SCHOOL SPORTS (950-959) | | | | |
| 950 | ETHS Sports—General | | 2 | 1 |
| 951 | ETHS Basketball | 19 | 1 | |
| 952 | ETHS Swimming | 9 | | |
| 953 | ETHS Wrestling | 7 | | |
| 954 | ETHS Gymnastics | 7 | | |
| 955 | ETHS Track | 5 | | |
| NORTHWESTERN UNIVERSITY SPORTS (960-979) | | | | |
| 960 | NU Athletics—General | 16 | 4 | 2 |
| 961 | NU Athletic Department | 23 | 2 | |
| 962 | NU Basketball (Big 10 Conference) | 61 | 23 | |
| 963 | NU Basketball (Non-conference) | 13 | 3 | |
| 968 | Big Ten Conference—Basketball | 5 | 1 | |
| 970 | NU Wrestling (Big 10 Conference) | 22 | 2 | |
| 971 | NU Wrestling (Non-conference) | 17 | 4 | 1 |
| 975 | NU Hockey | 3 | | |
| MISCELLANEOUS (980-999) | | | | |
| 999 | Obituaries | 3 | 1 | |



Kenneth Janda is professor of political science at Northwestern University. David Gordon is assistant professor of journalism at Northwestern's Medill School of Journalism. The paper was presented before the Newspaper Division on Jun 2, 1969 during SLA's 60th Annual Conference in Montreal.



thesaurus was designed for a MIRACODE system.

7. By entering the first seven characters of the last names of individuals in the clippings, we were able to accommodate the complete names in most cases. No actual count of the completion percentage was obtained, but another study involving the names of authors publishing in *The American Political Science Review* disclosed that 57% of the authors' names could be recorded with six characters, although only 10% of the complete names could be recorded with four characters. See Kenneth Janda, ed. / *Cumulative Index to the American Political Science Review, Volumes 1-57: 1906-1963*. Evanston, Northwestern University Press, 1964. p.xx. Of course, limiting the number of letters for personal names to seven characters (plus two for initials) was an arbitrary decision based on using only three columns of code on the film for each name. If greater precision is desired, more characters can be encoded—at the cost of one extra column of code for each set of three letters.
8. Procedures were set up in this project to allow clippings to be transferred from a general folder to a new and more specific one, and to be refilmed in the process. This procedure never was required during the quarter. However, it would result in the most recent recoding being placed most recently on the film (and out of chronological order with other clippings). This would mean that the recoded film would be searched first in any "browsing" operations using its newest topical designation. In any indexing procedure, the recoded clippings would show up under both their old and new designations, which would provide a handy cross reference. This procedure might, however, make it necessary to include a code for the year of publication (probably using the last three digits of the year in question), to allow for full and clear information to be presented in an index of the morgue holdings. During the particular 10-week period covered by this project, no clippings were coded for Category No. 111, "City Data Processing."
9. Note that, if a separate file had been set up for the James Brothers controversy, all this searching would be unnecessary, because all earlier clips would have been recoded for this single topic category and refilmed; and all clippings filed subsequent to the establishment of the new file would have been coded properly the first time. In this case, however, there were too few clippings on this topic to warrant a separate file.
10. This was actually determined by electronic scanning of both reels of microfilm, with the retrieval station set to search for any references to James. Because of time limitations, we did not prepare a program to provide an index based on the punched cards, but it would be fairly routine for users of this system to maintain an updated list of all names and other information contained in the coded material once such a program were prepared for the first time.
11. See, for example, Chapter I, "Indexing by Alphabetizing Keywords," in Kenneth Janda / *Information Retrieval: Applications to Political Science*, Ref. 3.

Received for review Jun 9, 1969. Accepted for publication Nov 26, 1969.

• • •