

Indexing a local newspaper using dBASE IV

Maureen Kilcullen and Melissa Spohn

Describes the design and creation of an index database, including software selection criteria, record and database design and useful program features. Discusses the value to the community of comprehensive newspaper indexing, and the need for a customized thesaurus.

In the summer of 1993, Maureen Kilcullen, Reference Librarian at the Stark Campus of Kent State University, and Melissa Spohn, Media Librarian at Kent State University, Kent Campus, decided to investigate indexing the local newspaper, the *Canton Repository*. The decision to index the newspaper was in response to an observed need. Although Canton, Ohio and the surrounding area are rich in local history, the libraries and historical societies lacked easy access to the basic historical information about local events, people, government and politics. Information about historical events and characters could sometimes be located in a six volume set entitled Stark County History. However, the information contained in this work would merely hint at the answer. There was no convenient access to the area's unique history.

Value of hidden information

In the past, attempts had been made to create indexes to this type of information, but there was no real effort to develop a subject authority thesaurus or a complete index. At the newspaper clipping library located at the *Canton Repository*, one was apt to find the same information under several different headings, which were not cross-referenced. And at the Stark County District Library, attempts have been made to index parts of the newspaper, (e.g. vital statistics), but no one had yet provided the community with a completed index. Clearly what was needed was a comprehensive index to the information found in the local newspapers.

Spohn and Kilcullen chose the newspaper the *Canton Repository* as the initial project because it had been consistently published since 1815. The *Canton Repository* was originally published by John Saxton under the title Ohio Repository, beginning as a weekly in 1815. Due to the historical importance of John Saxton's family, (his daughter, Ida, married William McKinley, the 25th President of the United States), combined with the industrial importance of the county (which includes the Timken Company, Republic Steel, and the Hoover Company), an index to these newspapers would be of considerable historical significance. The index would provide community access to information that has been virtually impossible to retrieve. Additionally, it would serve as an educational and community interest tool for the residents of Stark County, Ohio.

The extent of other indexing projects was confined to personal names and select subjects of interest. Attempts to limit the scope and effort involved in indexing the local news would reduce the comprehensiveness of the index.

Following the advice of Hans H. Wellisch¹, it was decided, that all local news and events would be indexed since national news is indexed elsewhere. Local interest items include all persons associated with Stark County; artistic and cultural

activities; first and last events (e.g., the open and closing of a store); local and state politics; court cases; business news; disasters (natural and accidental); local sporting events; crime; and social trends such as the rise or fall of population or changes in the ethnic composition of the city or the county.²

Local product advertisements, vital statistics and editorials which pertain to Stark County would also be included in the index.

Software selection

Kilcullen and Spohn knew from the outset that the index would not only interest the Kent State University Stark Campus but also Stark County area public, academic, and special libraries, as well as the local historical societies. Eventually, interested libraries would either be sent a printed index, or have electronic access to the index. The librarians agreed with Barbara Semonche in that computers facilitate production and distribution and enable the establishment of essential subject authority files³. With this in mind, it was decided to index the newspaper with the aid of a computer. The next step was a search for software that would meet the indexing needs of the project.

Several factors were considered while evaluating indexing programs:

- 1 ease of use from the aspects of data entry and database end use;
- 2 program flexibility;
- 3 printing capabilities and versatility;
- 4 search features;
- 5 product support;
- 6 hardware requirements;
- 7 cost and suitability to this project.

Three software programs were evaluated: Procite, Cindex and dBASE IV. The librarians had also investigated the possibility of testing the program Macrex, but were unable to obtain a demonstration copy to examine. They were fortunate to discover that the Kent State University School of Library and Information Science had copies of Cindex and ProCite and were able to evaluate them.

Final decision criteria

Although both ProCite⁴ and Cindex⁵ were powerful information managers, the librarians were concerned not only with ease of use but also how the final index would be distributed. They had thought the index could be distributed to the area libraries in disk format, but discovered that the libraries did not own either program. Paper copies of the index would be difficult to supply without continuing grant support. Since the necessary funds were slow in coming Kilcullen and Spohn decided not to use either specialized program and began

to consider using dBASE IV. dBASE IV proved to be quite acceptable for the following reasons:

- 1 Accessibility — the Stark Campus already owned the program and had enough site licenses for us to use it.
- 2 Support — the Stark Campus provides computer support and in-house workshops on the use of dBASE IV.
- 3 Versatility — the program provides a number of ways to organize data within database files including sorting and indexing.
- 4 Ease of use — once the database is set up, data entry is very simple.
- 5 Construction of a thesaurus — dBASE IV features the ability to construct a subject authority thesaurus.
- 6 Report generating capabilities — reports can be generated for the following:
subject bibliographies — (local elections, crime, vital statistics)
bylines
personal names
- 7 Familiarity — Kilcullen had used dBASE III+ in another newspaper indexing project and had designed a periodical database which she frequently updated using dBASE IV.
- 8 Other libraries in the area used dBASE.

Database record structure

One of the first steps in creating a newspaper database index using dBASE IV is to determine the basic structure of the database record. dBASE IV requires that information be stored as either character fields, numeric fields, date fields, or memo fields.⁶ To determine the types and lengths of fields, samples of the index were created by actually indexing the newspaper. The following character fields were chosen:

- 1 Subject [subject of the article]
- 2 Heading [title of the article]
- 3 Byline [author]
- 4 Name [qualifiers: occupation or location]
- 5 Newspaper title (abbreviated); page and column
- 6 Date

Terminology control

A major concern was choosing the terms for topics and events. The *Canton Repository's* newspaper library consists of drawers of clippings filed under subject headings. The subjects chosen were solely dependent on the decisions of the person clipping the newspaper. There was no thesaurus. Common and general headings could be taken from the *New York Times* and *Newsbank*, but the concern was that such predetermined lists would not be able to handle topics that would relate to local events. Many stories in the *Canton Repository* could not usefully be indexed under standardized headings. The decision was made to supplement the combination of the *New York Times* and *Newsbank* headings with the thesaurus found in the *Napa City—County Local History Indexing Project Indexer's Manual*⁷ as well as to look for usable subject headings from the headline, or first or second paragraph.⁸ Headings would be revised to reflect events in Stark County. These three sources along with subject headings found in the vertical files containing local history collections in the area libraries would provide the basis for a thesaurus to the two newspapers.

Useful program features

The next step was to be able to pull out these terms from the database to provide a separate listing. A useful feature provided by dBASE IV is the ability to create a thesaurus. The *set unique on* feature allows one to extract the subject headings and/or names without duplication. This provided a basic authority list that could then be checked for accuracy and revised as needed.

Set commands change the dBASE IV defaults and allow data to be copied from a previous record into a new record. A data entry person can issue the command *set carry on* after opening the database file to be updated. If the data entry person is inputting records from one issue of a newspaper, he or she can use the command to make the system copy information entered in the first entry's date and newspaper fields to the next new entry. After determining which field(s) are to be carried over from record to record, the data entry person simply issues the command *append* which allows records to be added to the end of the file.

Another feature is the use of function key **f8** which allows the data entry person to choose the field(s) to be duplicated while inputting records. Data entry using **f8** is quick and easy and relieves tedious repetition. Often **f8** and *set carry to* are used together.

For example, in 1815, the Ohio Repository included the following particularly frustrating area: *List of Letters*, a weekly listing of the names of people who had received mail but had not yet collected it. Each entry had to be indexed by the name of the person no matter how many times the name was listed in ensuing weeks. To eliminate the constant repetition, **f8** was used to duplicate information from the previous record when the name, subject and heading were the same, while *set carry to* was used for date, newspaper title, page and column. Combining these features virtually eliminated the frustrating repetition in the *List of Letters*.

A powerful feature of dBase IV is the ability to create a database management system consisting of two or more database files. If there is a need to combine information from several database files, it is possible to *link* fields from the databases. To do so, *query* is used, enabling one to combine elements of two or more databases and create a totally different database without re-entering the data. Creating a *query* design also allows one or more files to be viewed at the same time.⁹

dBASE IV provides the ability to specify the order in which the data can be displayed. For example, one can organize the database file by any field (such as subject or name) and the results can be in ascending or descending order. In ascending order, uppercase letters are before lowercase, e.g., *Zeppelin* appears before *apple*. In descending order, *apple* appears before *Zeppelin*. One may also sort the database in an ascending or descending dictionary format. This latter type of sorting is not case sensitive and therefore protects against capitalization inconsistency.¹⁰ An added feature is the ability for dBASE IV automatically to update all the indexes whenever one modifies or adds records to the database.

dBASE IV's ability to generate reports as well as its print capabilities is an added attraction. The *custom report* feature enables one to add an introduction and a summary, change the layout of the fields and alter the fonts. It is also possible to save the report to a DOS file.

continued page 22

tionary makers and encyclopedia editors alike in connecting related headwords.

Many reference works are arranged in a classified sequence, and a detailed alphabetic index provides access to specific pieces of information. If indexers were brought in at the design stage, the decision might in many cases be made not to classify the body of the work, but rather to have an alphabetic arrangement with classified indexes. For example, an association that I belong to publishes a membership directory that features a primary geographic arrangement and an index of personal names. The most frequent use of the directory, however, is to find the address and phone number of a single individual—not to identify members in a given region; the index and the body of the directory should be reversed to permit direct access to personal names, with secondary access to geographic groupings.

Besides thinking about user needs and envisioning the information that will be sought in a work, indexers pay attention to typographic details such as variation of fonts to create entries of maximum clarity, as well as to the use of running heads, guidewords, and continuation headings to orient readers to the section of a book that they are consulting. Indexers often create special locators for illustrative material, a component of *Wordsmanship* for which no access was provided. Mulvany notes the advisability of bringing in indexers to plan the locators of multipart and/or frequently revised technical and policy manuals;⁷ indexers should also be consulted on the locators of bibliographies and other types of reference works that have numbered entries.

There exist many poorly designed reference works, and even excellent indexes cannot compensate for their design flaws. A classified body with an alphabetic index may seem like a sophisticated structure for a reference work, but it often frustrates the user who would prefer a single lookup in a self-indexing tool. That is why the alphabetic encyclopedia is a more popular form of reference work than the systematically arranged German *Handbuch*. In a recent article in *The Indexer*, David Crystal describes a time-consuming search for a specific piece of information in a classified reference work without a detailed index.⁸

Indexers who suggest the rearrangement of a classified work into a self-indexing tool—when warranted—are not limiting the demand for their services, as there is so much that they can contribute to an alphabetically arranged reference work in the areas of control of internal cross-references, typographic design, and overall user-friendliness. It is the latter quality which should give indexers a feeling of superiority, not the use of esoteric words that are accessible only to a few OROTUND (headword 'pompous') individuals.

Acknowledgements

The author expresses her appreciation to Dr James A. Benson, who read a draft of this paper and contributed valuable observations.

Our thanks are due to Dr Laurence Urdang for permission to quote passages and to reproduce figures from *Wordsmanship*.

References

1. National Information Standards Organization. *Guidelines for the construction, format, and management of monolingual thesauri*. Oxon Hill, MD: NISO Press, 1994 (ANSI Z39.19-1993). Glossary, pp 35-8.
2. Cerf, Christopher, Navasky, Victor. *The experts speak: The definitive compendium of authoritative misinformation*. New York: Pantheon Books, 1984. Index by Sydney W. Cohen.
3. Benson, James A. Review of: *Reference and information services: An introduction*, Ed. by Richard E. Bopp and Linda C. Smith. *Journal of Education for Library and Information Science* 33(2) Spring 1992, 166.
4. Wellisch, Hans H. More on indexes in encyclopedias (and on one encyclopedia index). *The Indexer* 13 (1) April 1982, 3-5.
5. Fienberg, Anna. Curious Ferdinand and his magic spectacles. In *The magnificent nose and other marvels*. Boston: Little Brown, 1991, 28.
6. Cross, Ruth Canedy. *Indexing books*. Cambridge, MA: The Word Guild, 1980, 4.
7. Mulvany, Nancy C. *Indexing books*. Chicago: University of Chicago Press, 1994, 237.
8. Crystal, David. Some indexing decisions in the Cambridge encyclopedia family. *The Indexer* 19 (3) April 1995, 178-9.

Dr Bella Hass Weinberg is a Professor in the Division of Library and Information Science, St John's University, Jamaica, New York, and a Past President of the American Society of Indexers.

Indexing a local newspaper using dBASE IV — continued

The dBASE IV software has proved useful due to its efficiency, availability and ease of use. These features, in combination with guidelines for indexing and revisions of the thesaurus and the index lead the librarians to believe that this will be at least a seven-year project — more than 180 years of local information will be indexed. The index will lead the patron to the appropriate date, page and column of the full text available on microfilm from the newspaper and the District Library.

References

1. Wellisch, Hans H. *Indexing from a to z*. New York: H. W. Wilson, 1991. 249-250.
2. *Ibid.*
3. Semonche, Barbara P. 'Newspaper indexing: in search of a solution'. *Collection Building* 7 (4) 1986. 25.
4. Cibbarelli, Pamela. 'Cibbarelli's surveys: user ratings of bibliographic

citation management software'. *Computers in Libraries* 15 (4) April 1995. 36.

5. Holmes, Olive. 'Cards to keyboard: indexing by computer'. *Scholarly Publishing* 24 (1) January 1993. 114.
6. Chou, George Tsu-der and W. Edward Tiley. *dBASE IV handbook*. 3rd edition. Carmel (Indiana): Que Corporation. 1989. 82.
7. Vierra, Bobbie and Maureen McCarthy. *Napa city-county local history indexing project indexer's manual*. Napa (California): Napa City-County Library, 1979. 21-63.
8. Wellisch, Hans H. *op cit* 253.
9. *Using dBASE IV*. Scotts Valley (California): Borland International Inc, 1988.
10. *Ibid.* 5-17

Maureen Kilcullen is a reference librarian for the Stark Campus of Kent State University. Melissa Spohn is the media librarian at Kent State University.