A New Trend in Newspaper Indexing

There seems to be a new trend in newspaper indexing. Consider the difficulty vendors face with the continual demand for more online newspaper content, combined with the fact that metadata needs to be created for each article within a daily paper. To assign professional indexers to create that much metadata is extremely costly.

In an effort to add nearly 100 years of historic content from the *Denver Post* (1894–1989) in a quick manner, *NewsBank* added fully scanned issues. Retrieved results do not point to discrete newspaper articles, but to hits on the scanned pages. This difference in document processing is not evident to the user, especially when historic issues without article metadata are combined within the interface with newer issues with metadata (see figure 1).

![Digital newspapers under NewsBank showing image format and text format search form.](image)

Notice the format field in figure 1. Image files are being cross-searched with text files. When testing out these integrated files with a default search for “Ku Klux Klan,” several observations can be made (see figure 2).
Articles from image searching and metadata searching are interfiled. Notice that there are two entries for April 13, 2018; two from February 22, 2018; and two from February 10, 2018. For each of these dates one of the indexed articles contains metadata for the article title, while the other says only “News Article.”

A second observation is that some of the dates only contain an “image” article. We see this for April 30 and for April 1. Both of these are editorial perspective columns, and it is possible that these will not be indexed in the same manner as regular news articles.

A third observation is not illustrated in figure 2. When doing a “mouse over” on the titles, “News Article” image titles show a snippet of the page image, whereas articles with full metadata show only text.

Another example of this trend is seen in the ProQuest Digitized Newspapers (formerly known as ProQuest Digital Microform) product. In this product, article metadata does not include title or author, just administrative metadata like date, title of publication, and pagination (see figure 3).
Figure 3. ProQuest search results page for ProQuest Digitized Newspapers.

The “browse this issue” feature mimics the microfilm experience, with thumbnail images across the bottom of the screen. This is useful when trying to find the financial sections within the Wall Street Journal, one of the publications available through this ProQuest platform.

Figure 4. Thumbnail browsing with ProQuest Digitized Newspapers.
It remains to be seen whether this type of indexing (or lack of indexing) will work for researchers, or whether it is just an advantage to publishers.