Introduction

For family historians, historical newspapers are a wonderful source of information. Two freely available newspaper archives, Chronicling America at the Library of Congress and The Portal to Texas History at the University of North Texas, will be used to illustrate techniques for improving research effectiveness and sharing results.

Chronicling America [http://chroniclingamerica.loc.gov/]

Chronicling America is a Website providing access to information about historic newspapers and select digitized newspaper pages, and is produced by the National Digital Newspaper Program (NDNP). NDNP, a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC), is a long-term effort to develop an Internet-based, searchable database of U.S. newspapers with descriptive information and select digitization of historic pages. Supported by NEH, this rich digital resource will be developed and permanently maintained at the Library of Congress.

The Portal to Texas History℠ [http://texashistory.unt.edu/explore/collections/TDNP/]

The Portal is a gateway to Texas history materials. From prehistory to the present day, its unique collections from Texas libraries, museums, archives, historical societies, genealogical societies, and private family collections, provide researchers with primary source materials. The University of North Texas (UNT) Libraries maintains the Portal, which includes the Texas Digital Newspaper Program (TDNP) collection. The UNT Libraries is the lead institution in Texas selected for the National Digital Newspaper Program, Chronicling America. TDNP is a partnership providing broad geographic access to digitized Texas newspapers as far back as 1829. It currently includes 80,486 newspaper editions (618,619 files).

Types of Information

- Births and deaths
- Marriage announcements
- Military Service
- Land purchases
- Promotions
- Advertisements: Family businesses
- Travel announcements
System Components

Before demonstrating search techniques, it helps to understand a few key components of a newspaper archive’s system architecture. Put another way, here is an under-the-hood view of Chronicling America and The Portal to Texas History℠. Both systems use metadata, linked data, OCR software, and extensive application programming to optimize viewing, searching and browsing items.

1. Metadata

Descriptive information about a newspaper is contained in structured metadata, which often include elements such as collection, title, publisher, date, language, description, subject heading, and location. Structured metadata is used by a system to parse items and create relationships.

2. Linked Data

The representation of data from different collections in a common structure enables a system to cross reference or link data. This allows connections to be made among disparate collections and enhances browsing capability.

3. OCR Software

The digitization of newspapers involves scanning the source documents either from paper or microfilm. The result is a digital master image of the source document, typically in tiff format, as well as other formats generally used for viewing and downloading, such as jpeg or PDF. Optical Character Recognition (OCR) software converts the images into text format, thus enabling the full text of newspapers to be searched.

4. Application Programming Interface (API)

The system includes several different views of the data about the newspapers. The API allows programmatic access to these views and enables the data to be manipulated, viewed, and searched in multiple ways:

- Thumbnail view of newspapers
- All issues of one newspaper
- All pages of one issue
- Different sizes of one page
Mining Newspaper Archives

Kathleen Murray & Tara Carlisle

- Collections, titles, dates

Searching Historic Newspapers

The presentation will demonstrate search techniques for specific types of information of interest to family history researchers.

1. No hits? Don’t give up yet!
   - Boolean searching
   - Phrase searching
   - Searching names
   - Language of an era (runaway; ranaway)

2. Lots of hits? Narrow the field!
   - Faceted searching
   - Linked data
   - Advanced Search

Using Search Results

1. Printing
   - Benefits and pitfalls
- PDFs – zoom feature

2. Saving
- Saving pages
- Saving searches through persistent Links

3. Sharing
- Email
- Social Media
- Persistent Links

Bibliography


National Digital Newspaper Program

More information on program guidelines, participation, and technical information can be found at [http://www.neh.gov/projects/ndnp.html](http://www.neh.gov/projects/ndnp.html) or [http://www.loc.gov/ndnp/](http://www.loc.gov/ndnp/).