

How Historians use Historical Newspapers

Robert B. Allen

College of Information Science and Technology
Drexel University
Philadelphia, PA 19104 USA
+1-215-895-0460
rba@drexel.edu

Robert Sieczkiewicz

University Archives
Drexel University
Philadelphia, PA 19104 USA
+1-215-895-1757
robs@drexel.edu

ABSTRACT

Newspapers have long been rich resources for historians. In the past several years many historical newspapers have been digitized, offering the promise of improved access and powerful searching. In this research, we focus on historians' needs for searching collections of newspapers and managing the information they find. This is a deeper and more targeted investigation than much previous work that was based on surveys rather than personal interviews. We interviewed eight academic historians who largely embraced digitized newspapers but suggest the current systems still have many limitations. We also discuss the implications for the design of interfaces and services that would serve as a historians' workbench.

Keywords

Newspapers, Historians, Humanities, Interviews, Interface, Search, Workbench

INTRODUCTION

Humanities Scholars' Use of Sources

There is a large body of work on use of literature by humanities scholars (e.g., Toms & O'Brien, 2008) that demonstrates a need for different information and communication technology tools for humanists than for other scholars. Humanists, Toms and O'Brien argue, have differing search needs that require databases with non-traditional metadata (p 123). Their proposed workbench would support downloading, storing, organizing and markup of information. While their conclusions point to the need for a humanists' workbench, historians made up only a fraction of the sample: fewer than 20 in a pool of 169. Of respondents describing their research, only 6% reported doing "historical analyses." Other work has focused on historians. Duff et al. (2004) surveyed 600 historians about their use of primary sources, identifying barriers and preferences relevant to this study. The most commonly cited barrier to accessing information was geographic location, cited by nearly two-third of respondents. While the largest group used original material in their primary source research, the second greatest preference was for microfilm material, used by 82% of respondents; this was far higher than the number who used electronic, or digital, reproductions (21%). We investigate whether these barriers and preferences apply to historians' use of

newspapers as well. The extensive study by Tibbo (2002) showed that newspapers remain vital to historians, identified as both historians' "most important" and "most often used" type of material. There are also informal sites which offer advice to researchers. Rules of thumb for effective newspaper searching have been proposed: "Rachel"¹, who describes herself as a doctoral student in history, suggests using multiple newspapers, reading beyond articles yielded by keyword searching and consulting secondary literature. Our research confirms that most of these techniques are broadly used, yet also demonstrates great variety in the variety of uses historians have for newspapers.

Goals and Outline of this Research

Unlike most of the previous work, we focused on interacting with collections of newspapers because there is now an especially large amount of digitized newspapers available for research, with much more being digitized each year. Moreover, we are exploring features which might be included in an interface. We have three major goals for this research. The first is to identify how historians use newspapers, regardless of the format. A second goal is to discover the problems researchers encounter in their research using newspapers. Finally, we seek to identify historians' needs for new interfaces and workbenches for using digitized newspapers in research. A more targeted and deeper analysis of usage than most previous studies of historians and humanities scholars, this work should be useful for the development of systems to facilitate historians' use of digitized newspapers.

PROCEDURE

There is a value to direct interviews rather than surveys. The interviewer was a historian. The interviews lasted about 15-30 minutes each. For the first three interviews, we relied on the interviewer's notes while the final five interviews were recorded and transcribed later.

We interviewed a total of eight academic historians from two universities who are likely users of collections of U.S. newspapers. There were 2 women and 6 men. Their dates of PhD ranged from the early 1970s to the early 2000s.

RESULTS

Here, we highlight some of the main findings from the interviews. Quotations from the interviews are indented followed by a code number that identifies the historian.

Microfilm versus Digital Access

All of these historians had experience with both microfilm and digital. Most reported satisfaction with both formats, but all had found obstacles to using microfilm effectively.

I haven't thought about it, but I am pretty satisfied [with digital newspapers]. I'd just like to see more things that would be the biggest thing: more years, more newspapers. I'm not a fan of browsing microfilm, if I have to I can do it, but I'd be much more prone to look at the NY Times or the Philly Inquirer or any other digitized newspaper rather than sit down with the Philadelphia Record or the Public Ledger [on microfilm]. [285]

I try to avoid microfilm as much as possible. Not only because they're a little bit harder to access [in that] you actually have to go to the library and use a microfilm reader, but the microfilm readers are just so uneven. I mean sometimes you get to the microfilm reader and the light is so dim that you can't even see it. So I try not to use microfilm [unless I have no choice].... I would much prefer the high contrast between black and white that you get on a computer screen and of course the ease in which you can do it at home at your convenience. [164]

Several of the historians noted limitations in availability of digital resources. One noted a regional bias in what has been digitized, with a preponderance of materials from the northeastern coastal cities. Another researcher mentioned a lack of digitized material from small towns. This echoes the conclusion of Dalton and Charnigo, who report that, unlike social scientists who suffered from information overload, historians want more information. They want access to more databases, to more texts in those databases, and to more years of those texts. Is it the case, Dalton and Charnigo ask, that "the more they have, the more they want?" (p. 415). This is a key theme of our respondents: lack of digitized years and issues of the newspapers that interest them. Beyond the amount and representativeness of digitized newspapers available, concerns about quality were noted:

While problems were most common for small collections, even for leading tools such as Readex, users reported metadata errors, mislabeled articles and incomplete articles.

Research Strategy

The historians reported a variety of strategies for using newspapers. One of them described using newspapers to "fill in gaps in research and corroborate information from other sources". Another historian's exploratory searching included looking at larger issues and events such as elections and campaigns. One of the historians used newspapers to find public opinion about changes in liquor license laws – to get a sense of "the texture of the city... how the city was thinking". Another would typically find a date in a court case of interest then search newspaper articles around that date. Newspaper stories provided more details to research, such as names of witnesses or investigators, as well as location information and images. For the historian who researched the internment of

Japanese-Americans in the 1940s, newspapers from different towns are used to compare and contrast:

We separate them and then go through in sort of a constantly comparative method I guess you know we read them repeatedly looking for themes and then we'll go through the columns and go through the editorials because they serve different functions in the newspaper so columns are obviously the personal opinions and then editorials reflect the judgments or the view points of the editorial board of the newspaper, this is what we believe as a newspaper as an institution. [706]

All interviewees used other sources in conjunction with the newspapers. Several described their method as going from archival source to newspaper. The historian of engineering consulted "every single institutional document" created in the first few decades of the Institute he studied, while also using the archives at a nearby college and state government records. The historian of criminals relied upon court records such as trial transcripts, bills of indictment, court rosters, and intake documents, as well as personal narratives. Another of the historians relied on microfilmed Mexican parish records. The baseball historian used free online resources compiled from primary sources, particularly Retrosheet for box scores and www.baseball-reference.com for player and team statistics.

Searching

The humanist is typically a browser rather than searcher (Toms & O'Brien, 2008; Dalton & Charnigo, 2004). This tendency to browse came up often in our interviews: historians who read papers sequentially across months or years; the respondent who described his peers as browsing though an entire reel of microfilm rather than searching across years. This may be a question of the limitations of print and microfilm. Are historians browsers by nature or have they just acted that way because their sources have required it? Further, how well does past research about humanists using humanities literature apply to historians using primary sources?

Scope of Searches

The historian of racial equality used a combination of big city and small town newspapers: the former for urban classified and display ads to see division of jobs along racial lines and the latter for research on Maryland race riots in the 1940s. The most common type of search was for specific keywords. Keyword searches could be used not merely to find information, but to rule out possibilities and circumscribe dates of study. For example, the historian of Mexico used keyword searching to "be sure there had been no earlier mention of cholera." Full-text keyword searching was mentioned as most important for one historian – both for names of individuals and for subjects (e.g., lynching, mob justice). The ability to limit keyword search results by location (such as county or town) and by date or time period was also described as very important.

Searching on Specific Attributes By Date and Event

Most interviewees indicated that searching and narrowing by date was essential. Event searching was felt to be useful in some cases (such as for the baseball historian, who

would use it to find articles about Opening Day or the World Series, or a game between particular teams). Events such as holiday celebrations and military events (e.g., “the attack on Pearl Harbor or the Battle of Gettysburg”) were also mentioned, as was the technique of searching on anniversaries of important events when there would often be summary or reflection. Another interviewee described using date searching of newspapers to verify or question primary source accounts of a murder:

When I looked at it, I said "Wait a minute I know there is a guy writing a diary here on this day," so I could just look up that day and there it was, in the evening he describes the murder.... I would want to see the newspaper of that day to see what he says. [683]

Searching by Name and by Author

One of the historians used quarterlies; the articles there had specific authors which the historian found it useful to search. The baseball historian found it useful to follow specific players as well as the work of a particular columnist. Names searched were likely to be subjects rather than creators of articles, and were satisfactorily found with keyword searches. Several others indicated that by-lines were rare or non-existent for their period of research.

I do know of colleagues who will analyze the collective writings of... a particular columnist or reporter and look for all of [her] columns or stories.... For me personally not right now [706]

Limitations of Searches

One of the historians noted the possibility that keyword search would be incomplete. Another, the historian of racial equality, found keyword searching overwhelming, noting that while the *Chicago Tribune* database is good for searching names, searching on broader topics – e.g., “race relations” -- brings back too many results [730]. Another expressed a reservation that keyword searching was a shortcut or deviation:

When [historians] use that kind of feature they should note it in their work, in the bibliography or in their notes. In other words, I wouldn't want someone to think necessarily that I had read every issue of a newspaper between 1850 and 1855 because that would be a misrepresentation of what actually happened. [164]

Searching non-English papers presented complexities:

The databases that we have that don't have page views, which I don't like, partly because the Spanish transcriptions are often wrong. It makes a big difference if the diacritics aren't there and you can't tell what a word is. And it's also going to be a problem with anything in a foreign language; when searching for it, what are the transliterations going to be. [683]

Browsing

Browsing is distinct from searching and is widely used to explore the context of events. One historian suggested that it would be great for broad topics, but not so much for narrow topics. In the same vein, another suggested:

Historians are always looking for both [context and facts]. You can't have the specific facts without the context and that's part of the reason why I at least skim the whole daily newspaper, just to see what kinds of things are there. Where [an article] is in the paper, and what surrounds it, matters. There isn't a lot of advertising, but what there is often can be interesting too. It's

the things that you find when you're not looking for them that are really fun. [683]

Confirming the research of Toms and O'Brien, among others, one historian characterized his peers as people “who would rather browse” an entire year of newspapers on film than search for specific articles. [285] However, chaining of browsed articles is uncommon [164].

Specific Content Types

There is also perceived value to searching specific content types. We consider a sampling of content categories. These are generally well recognized categories but they are not always clearly differentiated (Allen & Hall, 2010).

Advertisements

Ads were treated as distinct from images, which we consider separately in this paper. Several of the historians we interviewed treated advertising as key pieces of the cultural context:

If again I was writing [economic or cultural history] and I wanted to give the flavor of the times, you know, and say on such and such a date Macy's was advertising this kind of thing and this play was in its 32nd week on Broadway but certain backrooms somewhere these people were planning this that and the other thing. [535]

Usually they're at the back. It seems like there may [have been] some space left at the end, they put in ads... With the epidemic I am particularly interested in things people are selling to prevent or cure cholera. [683]

Editorials

Opinion pieces were felt to be useful, when available. One interviewee noted that editorials were not always distinct from articles, such as in the “partisan papers in the nineteenth century.” [683]

There are some regional differences and the press had different priorities and cultural differences in a way certain things got expressed. Especially since the middle of the twentieth century there is a certain homogenization, but there is a certain variety. [535]

Obituaries and Death Notices

Obituaries were judged to be “occasionally useful” [164] and one historian found them particularly helpful:

If they make a big thing about [Babe Ruth] being The Sultan of Swat then that would be one note on a note card and if they talk about overcoming disadvantages of growing up in an orphanage... [I] use that for dealing with his earlier life. [535]

[In teaching] I have the students try to find the obituaries of people depicted in films. And say ok that's just how they were presented, how their deaths were noticed and maybe they were icons even before Hollywood got their hooks on them. The Hollywood treatment affects the way the press talked about them. [535]

Images

Not everyone finds images useful [706]. The most common application was for illustrating publications. For instance, the cover for one of the historian's books features an illustration from a historical newspaper. Others found them helpful in rounding out a topic otherwise described purely by text.

Managing Search Results

Noting that searches lead to other searches, one historian wished to see how searches are nested within each other, in order to get back to earlier search results.

A visual map telling you where you are in your search would be especially helpful [as would] a system that lets you easily use multiple windows. [730]

Printing results was a common. How do they manage these printouts? One admitted that the “worst case” is to photocopy articles and “mash them into a ‘to sort’ folder.” A better system is to read articles and take notes on a laptop. The “best case would be to enter all the information into Endnote” [730]. Chronological arrangement also remains popular: However, methods of managing results varied with the nature of the research:

It depends on what I'm working on. If it's a book I [organize notes] by chapter. I don't really use note cards and I don't actually think a lot of historians do now use note cards. Although that was very common earlier I don't think people do it that way, at least I don't. [164]

Another [164] indicated that in her first book, which focused on multiple people, she organized her search results (printouts) by person. In her current research, which focuses on one criminal case, she sorts results by newspaper. The ability to tag articles with terms/keywords to help find them later “would be incredibly useful” [730]. One of the historians looks at articles as threads in a tapestry and is not always sure where in the tapestry they will go when she finds them. The ability to tag articles with terms/keywords to help find them later would be great. That same historian said that she uses ancestry.com for census data. She doesn't print this data out, but keeps it in a “shoebox” in ancestry.com where it is easy to find later. This might be useful in a digital newspaper system.

Reaction to Possible New Services

There was not full agreement among the interviewees, but the majority thought that new services could be useful, particularly those that offered coordination among multiple newspapers. One historian said that the ability to tag articles with terms/keywords to help find them later would be “moderately useful” while another said she would not use tags but might use annotations. Boolean and proximity searches were seen as the best alternative for supporting broad searches Web2.0 features were seen by some as not useful; others believed they would be useful provided that the researcher's privacy could be respected. One historian also questioned the usefulness of mass digitization of archival materials:

Once you get into this level of detail about stuff... not many people are looking at it. I think the people who claim that everything will be online eventually just have no idea of how much stuff there is that nobody looked at since it was put in the archive. So I'm not optimistic about everything being there and being connected I just don't think it will ever happen. [683]

CONCLUSION

These interviews reveal a wide range of needs for historians working with collections of digitized newspapers. Supporting many of these needs is a matter of

broadly deploying known technologies but other needs will require development of new services. Historians will clearly benefit from improved search tools and interfaces which would support richer ways of interacting with collections. Two specific additional search services would appear to be particularly useful. The first is searches restricted by a variety of attributes. While many existing services support restricting by newspaper title and by date, there are many other attributes which could be applied. In part, this is a matter of coding the sections but as we noted earlier, there is little agreement about what the sections should be and the content itself is highly variable. The second big challenge that became apparent was searching on broad topics such as “race relations”. Visualization tools can also be useful for supporting search (Allen, 2005). As we add more features – especially features that require more complex indexing and coordinating of multiple newspapers (see Allen 2010) – the interface may be thought of as a workbench. Such a workbench might support all aspects of a single research project or even several research projects (cf. Toms & O'Brien, 2008). While we have mostly focused on digitized newspapers, we are entering an era of a wide range of historical resources. Therefore, for instance, the workbench could support access to census data, records, maps, photographs, and books and manuscripts. It could also support access to personal notes (Unsworth, 2000). We have focused on academic historians but there are other significant groups. These might include what Herbert and Estlund (2008) have called “citizen historians”.

REFERENCES

- Allen, R.B., (2005) Using Information Visualization to Support Access of Archival Records. *Journal of Archival Organization*, 3(1), 37-49.
- Allen, R.B., (2010) Toward a Platform for Working with Sets of Digitized Historical Newspapers. *IFLA International Newspaper Conference: Digital Preservation and Access to News and Views*, New Delhi, 54-59.
- Allen, R.B., and Hall, C., (2010) Automated Processing of Digitized Historical Newspapers beyond the Article Level: Finding Sections and Regular Features, *ICADL*, June 2010.
- Dalton, M., and Charnigo, L., (2004) Historians and their use of information sources, *College & Research Libraries*, 65(5), 400-25.
- Duff, W., Craig, B. and Cherry, J., (2004) Finding and Using Archival Resources: A Cross-Canada Survey of Historians Studying Canadian History, *Archivaria*, 58, 51-80.
- Herbert, J and Estlund, K., (2008) Creating Citizen Historians, *Western Historical Quarterly*, Autumn 39 (3).
- Tibbo, H., (2002) Primarily History: Historians and the Search for Primary Source Materials. *JCDL*, 2002, 1-10.
- Toms, E.G. and O'Brien, H.L., (2008) Understanding the Information and Communication Technology Needs of the e-Humanist. *Journal of Documentation*, 64(1), 120-130.
- Unsworth, J., (2000) Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? <http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html>