

# 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. The historians we interviewed largely embraced digitized newspapers but suggest the current systems still have many limitations. We also discuss the implications of the input from historians for the design of interfaces and services that would serve as a historians' workbench.

## Keywords

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

## INTRODUCTION

New resources for historians are rapidly becoming available in digital form; notably there are large collections of digitized newspapers. Historians use newspapers in conjunction with other types of primary and archival materials.

## 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). They demonstrate there is 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. They note that for some historians "period newspapers were the only source of information that existed on aspects of their research and that microfilm copies greatly facilitated their work and in some cases were the only remaining evidence to make the investigations possible."

Another thread of research has focused on developing tools to managing resources generally (e.g., Brockman, et al. 2001; Unsworth, 2000).

In addition to the formal research literature, there are also many informal sites which offer advice to researchers. Historians traditionally use newspapers for such tasks as establishing general outlines, facts and trends; searching for descriptions or details about people, places and events.<sup>1</sup> Rules of thumb for effective newspaper searching have been proposed: "Rachel"<sup>2</sup>, who describes herself as a

This is the space reserved for copyright notices.

*ASIST 2010*, October 22–27, 2010, Pittsburgh, PA, USA.  
Copyright notice continues right here.

---

<sup>1</sup>[http://chnm.gmu.edu/worldhistorysources/unpacking/news\\_how.html](http://chnm.gmu.edu/worldhistorysources/unpacking/news_how.html)

<sup>2</sup> Rachel, A Historian's Craft:  
<http://idlethink.wordpress.com/2009/06/16/on-newspapers-as-sources/>

doctoral student in history, has a set of techniques for searching newspapers:

- Have a list;
- Find a thread of some kind;
- Don't just use one newspaper;
- Don't fall into the trap of only reading articles that your keywords throw up;
- Use existing secondary literature;
- Keep really, really scrupulous notes; and
- Don't neglect the letters and the advertisements.

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.

### **Collections of Digital Newspapers**

In the past few years there has been a transition from microfilm to digital in providing access to historic newspapers, to the point where there are now quite a number of large digital collections available to most academic researchers.

In North America, these collections include commercial, public and academic digital newspaper collections, and general collections that include newspapers. Commercial collections include the backfiles of city papers such as the Chicago Tribune and the New York Times; databases that include many papers, such as the products of ProQuest and Readex<sup>3</sup>. Each of these has useful features but each also has substantial limits. For instance, Readex focuses on early American history. Public and academic resources include grant-funded projects like the Making of America collection and government-supported efforts like the National Digital Newspaper Project (NDNP) of LC and NEH (Murray, 2005).

NDNP has made many newspapers available in a digital format. These were microfilm reels created by the United States Newspaper Program (USNP) that were digitized and then subjected to optical character recognition. The materials include some of what would be called magazines. Public access is provided by the Chronicling America<sup>4</sup> interface. The limitations of the NDNP newspaper collections include the inaccuracy inherent in OCR vs. key-stroked text; efforts to overcome this problem include processing OCR text with reCAPTCHAs and other collaborative correction techniques.

### **Goals and Outline of this Research**

Unlike most of the previous work, our goals are more specifically focused on interacting with collections of primary sources and newspapers than on understanding the use of the broader scholarly literature in the humanities.

We focus on the use of newspapers because there is 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, from search to managing the results of search. While much work on the information needs of historians ends once the historians have found the information they seek (Toms, p 104), our research includes an investigation of post-search. For instance, how does the historian annotate, save and organize her notes on an article that has been identified as interesting?

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.

This more targeted and deeper analysis of usage than most previous studies of historians and humanities scholars should be useful for the development of systems to facilitate historians' use of digitized newspapers.

### **PROCEDURE**

#### **The Historians**

We interviewed a total of eight academic historians from two universities. There were 2 women and 6 men. Their dates of PhD ranged from the early 1970s to the early 2000s. We sought out historians who are likely users of newspaper collections. Their fields of research covered diverse aspects of primarily North American history from the mid 19<sup>th</sup> century to the mid 20<sup>th</sup> century. Their research topics included:

- 19<sup>th</sup> century engineering education
- Racial and occupational equality in the Civil Rights era
- Black women criminals in the early 20<sup>th</sup> century
- Baseball history
- The middle class in the antebellum U.S. South
- Popular perceptions of war crimes
- A cholera epidemic in Mexico
- Japanese-American internment camps

#### **Method**

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. A few minutes of the recording of one interview was lost.

#### **Interview Questions**

Those questions indicated with an asterisk \* were asked only of randomly selected participants.

---

<sup>3</sup> <http://www.newsbank.com/readex/press.cfm?press=33>

<sup>4</sup> <http://chroniclingamerica.loc.gov/>

### *Use of Newspapers for Research*

Briefly, describe your experience with newspapers for historical research. Describe a specific experience using newspapers for research. Did you search only one newspaper title or did you use several in the course of this search? What was that title(s)? How long did you spend searching? What were you seeking in that case: (a) general information (exploratory), (b) specific fact(s), (c) mixture? What was the primary format of the newspapers you used: Microfilm, Print, Digital? What parts of the newspaper are most relevant to your research (e.g., articles, advertisements, death notices)? \*Do you use newspaper images in your research? How are they useful to you as historical sources?

### *Searching*

If you have searched digital newspaper files, how well did that meet your needs? What sort of searching would you want to do in a digital newspaper search interface? For example, by words (e.g., full text), by date, by author, by event, by links to other articles.

[Interviewees shown screen capture of *Chicago Tribune* online archive.] Here is an example of a typical online search interface. Which of these search options would be particularly useful to you? Are there search options missing from this interface that you would like to see? \*Would you find it helpful to be able to access many different resources (e.g., several different newspapers or data sources from one interface)? If so, could you give an example?

### *Managing Ongoing Digital Searches and Search Results Sets*

What do you do typically with your search results? How valuable would the following be in a digital newspaper search interface: A log of all searches? The ability to save links? The ability to add and save annotations on articles? The ability to tag articles with terms/keywords or to put them in folders to help you find them later? \*Would you be interested in seeing user-contributed annotations (such as others' comments or tags)? Would you like to add your own annotations?

### *Wrap Up*

We're interested in developing an interface for digitized historical newspapers. Are there any features you would particularly like to see in such a system? Are there any parts of your comments which you would rather we didn't attribute directly to you in publications of this work?

## **RESULTS**

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

## **Perceived Value of Newspapers as Historical Sources**

Interview subjects were selected based on our understanding that they used newspapers in their academic research. From this point of departure, we found that the medium varied in value to our subjects, from the baseball historian who uses them almost exclusively, to others who use them as a tool to facilitate research in archival records. Feelings ranged from passionate to ambivalence.

I think most professional historians wouldn't work exclusively from newspapers because the newspaper is [suspect]...It's often taught as a primary source up to high school but in college you start developing a suspicion for newspapers because it's already removed a step or two, and having somebody else describe what happens.... So there isn't any perfect source but I think there is a general level of suspicion about newspapers but at the same time there is so much information there that you can't ignore it but we also don't treat it as it's got to be true because it's in the newspaper. [683]

## **How Much Time is Spent Searching the Newspapers?**

Most of the historians reported spending a great deal of time searching. For instance:

Maybe a few months, probably half a day (per day) for a total of a month. [683]

Days of searching [558]

My whole life![laughs]. Over 10 years on this project. [039]

40-50 hours of searching (by a research assistant) [706]

## **Microfilm versus Digital Access**

All of these historians had experience with both microfilm and digital. Our interviewees generally had more experience with the former than the latter. Most reported general 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 you actually have to go to the library and use a microfilm reader, but the microfilm reader 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 no 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 rather than physically having to go to a library. [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 then 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 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:

Well if they only had those that would be perfect, if you're not adding more by leaving something out or missing something. What I have seen with Google Books was kind of spotty, actually with illustrations the same way. Sometimes you can see that there's one folded up that they didn't bother to unfold. [683]

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

### **Collections Used**

Commonly cited collections included Readex (described as “very nice”), the New York Times, Los Angeles Times, and Chicago Tribune. Most of the historians also used specialized collections such as Mexican newspapers. The baseball historian reported finding the *Sporting News* particularly useful, with access through his membership in the Society for American Baseball Research (SABR).

### **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]

### **Coordination with Other Primary Sources**

All interviewees used other sources in conjunction with the newspapers. 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. [039]

Multiple researchers described their method as going from archival source to newspaper. [730]

The baseball historian used free online resources compiled from primary sources, particularly Retrosheet<sup>5</sup> for box scores and [www.baseball-reference.com](http://www.baseball-reference.com) for player and team statistics.

Another of the historians relied on microfilmed Mexican parish records.

### **Searching**

One set of issues concerns searching itself. Research has repeatedly shown that the historian (a class to which the historian belongs) 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

---

<sup>5</sup> [<http://www.retrosheet.org/>]

racial lines and the latter for research on Maryland race riots in the 1940s. [730]

Another research had a well-demarcated project at the start of research, “trying to reconstruct a discourse,” so the language people used was more important than finding historical details. [558]

### *Keyword Searching*

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.” [683]

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/time period was also mentioned as very important. [039]

### *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.

An interesting possibility mentioned by one of historian was 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:

[W]hen 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. [The murderer was] described as a soldier who stabbed someone to death with a sword; my diarist says that he was a journeyman shoemaker. And he uses a verb that could mean to run somebody through with a sword and could also mean to stab someone with a blunt object. And I'm thinking if he is a journeyman shoemaker he's not likely to have a sword, he's likely to have an awl that he's using to punch the leather and if he hits somebody with that it's going to be ... described the same way. So I would want to see the newspaper on that day just to see whether it says the individual was a soldier or whether this is the historian's conclusion that

I'm reading based on the verb that's used. So, yeah, searching by date would be very important, especially for checking up on other people. [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. For instance, the baseball historian found it useful to follow specific players as well as the work of a particular columnist. [285]

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 but I think it could be a valuable tool. [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 from norms of research that should be disclosed:

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. So what you would want to do is to be honest and up front and say what I really did was this kind of specific detail search for a phrase. [164]

Searching non-English papers presented complexities for the researcher:

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]

There were also challenges of metadata that worked perfectly well for paper-based works but fail the user of digital resources. For example:

The *American Historical Review* has online articles but they don't tell you, you can't see the pages and so the

paragraphs are numbered and they say "Ok it's acceptable now to cite things by paragraph number." All right fine but I am not, it should maybe easier to find things if you know what paragraph rather than what page, and it's a little more precise. But it doesn't help you with somebody's citations that are to pages. [683]

Metadata that suits the modern newspaper, which appears once a day, presents a problem for the researcher using newspapers that came out several times a day.

### *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. [730]. 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]

A clear example provided by one of the historians was illustrated by personal interest rather than academic research:

On Armistice Day in November 1918, there was a picture of my father in the Public Ledger, standing outside selling newspapers when he was about 8 years old. He always told me about it, but we could never find it. Well, I went to the Library and I found it [by browsing microfilm]. [285]

The same historian characterized his peers in a manner than confirms the research of Toms and O'Brien, among others:

There are people I know who'd rather browse. They shun digitization, they just take 1938 [microfilm reel] and go through everything. [285]

However, explicit chaining of articles is uncommon [164].

### **Specific Content Types**

There's also perceived to be a great 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:

Well, you know, they can be useful for different types of history, you know, doing economic history, cultural history. If again I was writing something 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 32<sup>nd</sup> week on Broadway but certain backrooms somewhere these people were planning this that and the other thing. But, you know, in and of itself depending on the kind of work you're doing you derive an awful lot of substantive information from advertisements, so sure. [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*

Editorials or opinion pieces were felt to be useful, when available and distinct from articles

Sometimes it's hard to tell the difference in this era. They don't make the distinction between one and the other. I guess there are editorials, but there's also a lot of opinions in the articles. They're very partisan papers in the nineteenth century. You probably can make the distinction. One of the other things is that they print excerpts from other newspapers. [683]

I think [editorials are] important. 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]. One historian of the 20<sup>th</sup> century U.S. found them particularly helpful in conveying contemporary perceptions of a historical figure:

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]

Another noted that obituaries were rare in his period of interest:

There are some, particularly for prominent people, but not for ordinary people. If the president's wife dies, there's notice and a description of the funeral, but that's really not an obituary; it's news. So, maybe not, maybe there aren't obituaries. [683]

### *Images*

Newspapers have substantial numbers of images, which have several possible applications. However, not everyone finds images useful [706]. The most common application was for illustrations. For instance, the cover for one of the historian's books features an illustration from a historical newspaper. Others found them valuable in rounding out a topic otherwise described purely by text.

### **Managing Search Results**

Another set of issues deals with managing results from searches. 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 practice:

I will tend to err on the side of printing it out rather than not printing it out because I'm never sure if I'm going to be able to go back or remember to go back to the specific article and I don't necessarily like reading on the computer. [For a long article] I will probably print it out and then use a highlighting pen and go through the results more specifically and take notes on the page that I printed out. I suppose somebody, I'm sure they have this now where you can probably highlight digitally on line and you can annotate yourself but I wouldn't do that. I would likely print it off and then sit with it and read it and then highlight it. [164]

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 remains popular:

I keep my [printed] pages in chronological order so that when I want to go back [and find them], I've found it's a good way to do it. Whenever I want to go back and write about a particular year I can do it. Then I transfer the information into [dividers] on heavy paper, one for each year, trying to file all activities in that year [of the subject of research] in some kind of chronological order. [285]

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 respondent suggests that this report of the demise of the index card is premature:

I use index cards at the beginning: I write down the date, the article title, any other information. If there's more information than I can put on the card I'll frequently do one of two things. First I'll go to a bigger card, 5x7, then write a little on there. If there's more than that even. I'll print the page out, then write on my index card that I've printed the page. [285]

Another 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. [164]

Editing a book manuscript recently, one of the historians found it "hugely taxing" to find items she hadn't cited. The ability to add and save annotations on articles does not match how she works. She takes notes about the body of work, not individual articles. The ability to tag articles with terms/keywords to help find them later "would be incredibly useful." [730]

Interviewees were asked about the usefulness of a log of all searches. For one respondent, this would be helpful in recalling results of research that has grown cold:

Well, you know, as you know you start a project and you're able to do a lot of research and then something comes up and you have to attend to committee work and graduate [students] and it's kind of cold. Knowing what you looked at could be useful, being able to bring it up electronically instead of going through a heap of Xeroxes. [535]

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 consistency among the interviewees, but the majority thought that new services could be useful, particularly those that offered coordination among multiple newspapers. One historian said he could imagine that the ability to tag articles with terms/keywords to help find them later would be "moderately useful".

Tags: No Annotations: Maybe. [164]

Not interested in a log of all searches. The ability to save links would be moderately useful, as a backup. Would not use the ability to add and save annotations on articles [558]

The problem with microfilm is that it requires painstaking care to record citations. [558]

Boolean and proximity searches were seen as the best alternative for supporting broad searches as would the ability to exclude words.

It actually had to do with the verb in the stabbing, the verb that was used was stabbing with a particular kind of sword, they're kind of pointy but you can't use it for slicing, you can use it for stabbing, they used it to finish off the bulls in the bull ring and somehow you would end up with a Lamborghini name for that kind of sword.... Once I could get rid of anything that referred to Lamborghinis, then I was much closer to finding what I was looking for. [683]

Web2.0 features were not seen by some as useful but others believed they would be useful provided that the researcher's privacy could be respected.

But one of the things that happens too once you get into this level of detail about stuff is that 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]

### **New Content Combinations**

The possibility of being able to compare and contrast content seems useful:

Well, you know, if I was looking at the nation's split opinion on Abraham Lincoln during the election of 1860 and I could have a newspaper from Richmond and a newspaper from Mobile, and one from Chicago and one from New York and one from Cleveland up at the same time that would bring that variety that much quicker. [535]

Of course there are also standalone tools:

... I use [Nota Bene] to keep my bibliography together. It will search the notes for me using various combinations and telling me how many, I can get a nice, how far down I've narrowed it. If the word cholera appears 800 times in different files then I need to add another word, to get a reasonable number, say 20 or so that I can look through for the idea that I'm looking for. So that kind of thing I would think would be useful. You need to know how much filtering needs to be done before it's a manageable number. [683]

I tend to stick things where I have books; I tend to put little tags on them... I think something like that would

be useful sort of like electronic book marks like you can put into some electronic documents.... [If] you're going to do that then the ability to write on them, say what it is you're doing... would be useful. [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 meeting some of these needs will require development of new services.

### **Search Tools and Interfaces**

Historians will clearly benefit from improved search tools and interfaces which would support richer ways of interacting with collections. While basic searches are well supported, or at least are available in some current systems, 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". One mitigating step would be supporting rich Boolean searches so that complex queries can be specified. In the same vein, clusters of queries could be specified. For instance, some aspects of race relationships could be found in the Help-Wanted classified, some in sports stories, and some in feature articles.

Visualization tools can also be useful for supporting search (Allen, 2005). For instance, search hits could be posted against a timeline (Kumar, et al. 1995; Swan and Allan, 2000).

### **Toward a Historians' Workbench**

As we add more features – especially features that require more complex indexing and coordinating of multiple newspapers (see Allen 2010) – the interface may be best 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). As a first step, multiple newspapers from one town or region may have synergies which improve the text processing of each (Allen, 2010).

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). To the extent it is



tailored to the individual researcher's work, it may be seen as supporting personal information management.

### **Rights, Licensing, Policies, and Standards**

While we have emphasized the development of tools to support better access to digitized newspapers and other digitized historical resources, rights policies and standards for access are also major issues. Simply being authorized to access content from multiple right-holders in one session can be difficult but perhaps that could be handled with an API and licensing agreements. However, negotiating those rights might be much more complex if derivatives were produced which incorporated fragments of content from multiple sources. Another set of issues deals with developing standards. This might include standard ways of referring to newspaper titles and section or, perhaps, standards for sharing annotations.

### **Beyond Academic Historians**

We have focused on academic historians but there are other significant groups. We seem to be entering a new era of historical research: These might include what Herbert and Estlund (2008) have called "citizen historians". These other groups may be more willing to embrace Web 2.0 technologies, even though we did not find much support in the interviews for Web 2.0 interactivity (Anderson & Allen, 2009). This is consistent with the findings of both Toms and Dalton, which characterize the humanist/historian as solitary. Just as Toms reports that only 17% of respondents "share hyper-textual notes or common annotation tools," our interviewees showed little enthusiasm for sharing article annotations.

### **ACKNOWLEDGMENTS**

We would like to thank the historians we interviewed. We also thank Ashley Brandon, a work-study student who transcribed the tapes.

### **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, in press.

- Allen, R.B. and Nalluru, S., (2009) Exploring History with Narrative Timelines. *HCI*, LNCS 5617, 333-338.
- Anderson, S., and Allen, R.B., Envisioning the Archival Commons, *American Archivist*, (2009) 72(2), 383-400.
- Anderson, I., (2004) Are You Being Served? Historians and the Search for Primary Sources, *Archivaria*, 58, 81-129.
- Brockman, W.S., Neumann, L., Palmer, C., and Tidline, T.J., (2001) *Scholarly Work in the Humanities and the Evolving Information Environment*, CLIR Report #104.
- 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).
- Kumar, V., Furuta, R. Allen, R.B., (1995) Interactive interfaces for knowledge-rich domains. *Electronic Publishing*, 8, 235-246.
- Murray, R., (2005) Toward a metadata standard for digitized historical newspapers. *JCDL*, 330-331.
- Swan, R. and Allan, J., (2000) Automatic generation of overview timelines. *ACM SIGIR*, 49-56.
- 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>