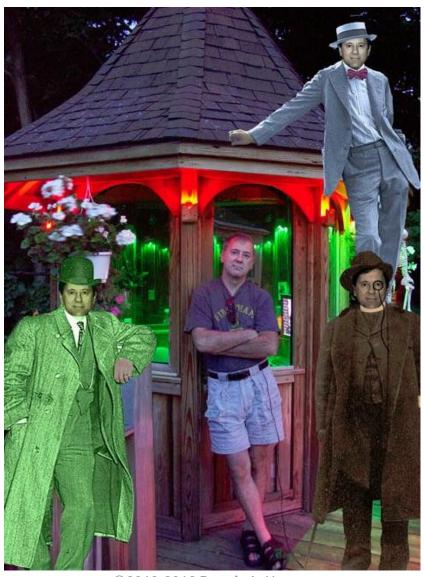


#### Tom & His Staff



©2012-2019 Dennis A. Hogan, www.dennisAhogan.com

...but really it's a 1 person operation!

#### Notes

 A copy of all slides of this presentation (with more detail than the handout) will be available on on my site, www.dennisAhogan.com

#### Why use Newspapers

- · Birth/Marriage/Death info
  - Substitute for Vital Records & Church Records
- · Info about relatives & associates
- · Learn tidbits about personality
- Get Lucky!

#### Newspaper Sites

- New York State Free Sites
  - www.fultonhistory.com 44+ million pages
  - http://nyshistoricnewspapers.org/8+million
  - http://chroniclingamerica.loc.gov/ 14+ million (only downstate papers)
- Other newspaper resources
  - www.theancestorhunt.com
  - https://www.cyndislist.com/us/

#### Some Website Background

- Tom was Corporate Director of Systems for a NYC firm
- In 1998 a friend's mother had some boxes of postcards from Fulton & upstate
- Some organizations that have used his site include: US Executive Office, Secretary of State, FBI, and CIA

#### In the News

- Genealogy/Tech blogger, Dick Eastman, highlighted Tom Tryniski and his site.
- Eastman Article link
- YouTube interview

#### Presentation Background

- The Church Records Preservation
   Committee (CRPC) of the Rochester
   Genealogical Society (http://nyrgs.org/)
   digitalizes records of the oldest
   churches in Rochester
- In August 2011, the CRPC visited Tom Tryniski in Fulton, NY.

#### OCR Intro

- OCR stands for Optical Character Recognition
- OCR software translates words in image files into text files
- Image file >> OCR >> text file

#### OCR for mere mortals...

- We would use an image of a newspaper clipping of so-so quality
- We would use free or inexpensive OCR software. See <u>www.freeocr.net</u> for a list of free OCR software.
- We would edit the resulting text file to make corrections.
- Note OCR is not perfect...a word that is human-readable may be interpreted differently by OCR.

### Tom's Approach

- High-end Hardware
- · High-end Software
- Automation

#### Scan Microfilm to digital images



- Microfilm is his input of choice
- •Needs 35MB image for successful OCR
- •This Wicks and Wilson 8850 scanner lists for \$80,000 he paid \$35,000

©2012-2019 Dennis A. Hogan, www.dennisAhogan.com

The "slides" on the screen are individual images as they are being digitalized from microfilm. The images scroll from right to left.

Once he has calibrated, the images fly by.



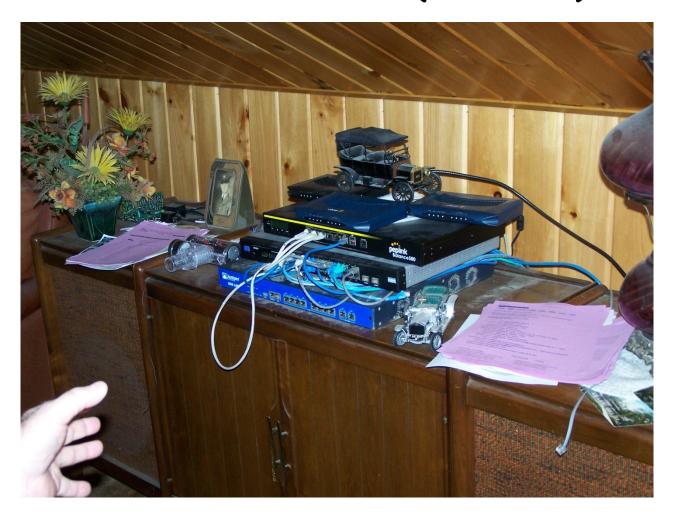
## Convert digital images to text OCR (ABBYY FineReader)

He builds his computers.



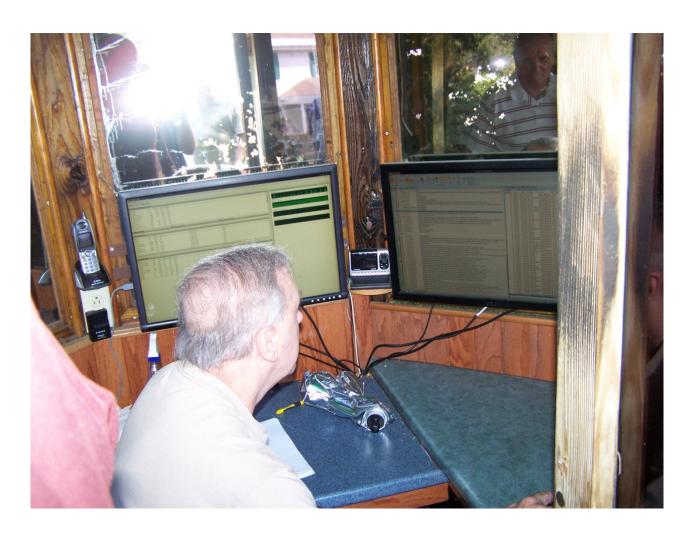
©2012-2019 Dennis A. Hogan, www.dennisAhogan.com

## 3 Business Class Road Runner Modems with a balancer (in 2011)



©2012-2019 Dennis A. Hogan, www.dennisAhogan.com

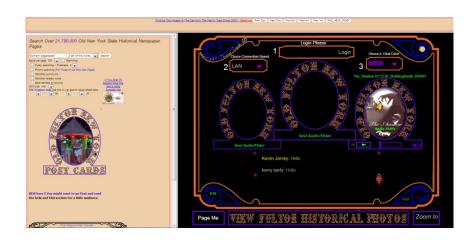
#### Tom Monitoring Usage in Gazebo



©2012-2019 Dennis A. Hogan, www.dennisAhogan.com

#### Website Usage

- www.fultonhistory.com
  - Bookmark the search page instead of the home page, <a href="http://www.fultonhistory.com/Fulton.html">http://www.fultonhistory.com/Fulton.html</a>

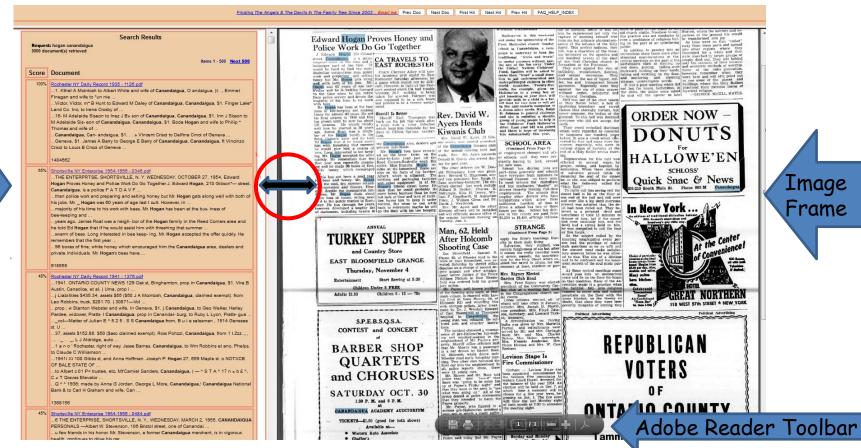


It's a free site, consider a donation



#### Website Usage

 Hovering cursor near bottom of image frame will display the Adobe Reader toolbar (unless a different reader is being used.



Results Frame

©2012-2019 Dennis A. Hogan, www.dennisAhogan.com

### "Results May Vary"

- · How this site displays, depends on your:
  - Computer pc, mac, tablet, ...
  - Operating System Windows 7, 8, 10, Mac, ...
  - Browser & version chrome, firefox, ie, ...
  - -PDF Reader & ver Adobe Reader, foxit, ...
    - chrome & firefox browsers have their own readers...
  - Version of the Hit Highlighter plug-in

## Search Hit Highlighting

- · Do a simple search, select a result
- You want your search hits to be highlighted, like in this example:



# What if your search hits are NOT highlighted?

## [Note - If you are seeing highlighted search terms, you do NOT need to follow the steps below.]

Follow these 4 steps to implement Adobe Reader.

- Install or Upgrade Adobe Reader (free) to the latest version
  - http://get.adobe.com/reader/
- Install the highlighter plug-in for Adobe Reader
  - Click on FAQ\_HELP\_INDEX button (upper right corner of fultonhistory.com)
  - Ctrl-F to search for highlighter
  - Click on Hit Highlighter for Acrobat 10 for instructions on installing the plug-in.
- Upgrade your browser to the latest version
  - Chrome: www.google.com/chrome
  - Firefox: www.mozilla.org/en-US/firefox/new/
  - IE: windows.microsoft.com/en-us/internet-explorer/download-ie
- Modify your browser to use Adobe Reader as the default PDF reader
  - Chrome: http://www.thechromesource.com/how-to-enable-adobes-pdf-reader-in-chrome/
  - Firefox: http://howto.cnet.com/8301-11310\_39-57570666-285/how-to-change-the-built-in-pdf-viewer-in-firefox/
  - IE: http://www.ehow.com/how 5806814 enable-adobe-reader-helper-ie.html

#### Search Hit Workaround

 When image displays in the Image frame,

- Do a CTRL-F
- Select Highlight All
- Enter search word
- Click Next



## Searching

Search Over 21,790,000 Old New York State Historical Newspaper Pages

Hogan Canandaigua all of the words ▼ Search

- Enter your string
- Choose from dropdown (click on arrow):
  - All of the words
  - Any of the words
  - The exact phrase
  - Boolean
- Press Enter or click Search

## Searching Examples

- john hogan canandaigua with all of the words chosen, yields pages containing all 3 words which may be scattered. (case does NOT matter)
- "john hogan" canandaigua with all of the words chosen, yields pages containing all 3 words with john & hogan next to each other.
- "john hogan" -hulk with all of the words chosen, yields pages containing john & hogan next to each other and NO occurrence of hulk.

#### <u>Searching - Potential Problems</u>

- Middle Names/Initials "John Hogan" vs
   "John F. Hogan" vs "Hogan, John"
- · Abbreviations "William" vs "Wm" vs "W"
- Married names "Mary Hogan" may be in the newspaper as "Mrs. John Hogan"
- Hyphenation "Hog-an" vs "Hogan"
- Long words (Canandaigua, O'Shaughnessy) increase the odds of OCR "mistakes"

#### Searching with Wildcards

- Wildcard characters:
  - Partial-word wildcards:
    - ? matches 1 character, ex: H?gan would match Hogan, Hagan, Hegan, etc.
    - \* matches 0 to many characters, ex: Wil\* would match Wil, Wilson, Willson, Willoughby, etc.
  - Full-word wildcard:
    - \* matches 1 word, ex: "John \* Hogan"
  - ~~ numeric range, 1912~~1921 matches any page containing a number from 1912 to 1921

#### Other Wildcards

- ~ invokes stemming, ex: hero~ matches hero and heroes.
  - % fuzzy wildcard (not as effective as fuzzy checkbox)
  - & synonym wildcard, headstone& matches tombstone Must select a synonym box!

#### Boolean Searching

- MUST select the boolean option in the dropdown menu, for the following operators to function correctly:
  - A and B, both must be present
    - A B gives different results than A and B
  - A or B, at least 1 must be present
  - -A w/5 B, A must be within 5 words of B
    - A w/1 B, A must next to B (also finds B A)
    - A w/2 B, this is useful to allow for a middle name or initial (also finds B A)
  - A not B, A must be present but not B

- With the boolean option selected in the dropdown menu:
  - -AB not =A and B
    - John Hogan = "John Hogan" (with the "all words" option)
    - John and Hogan = John Hogan (with the "all words" option)

- With the boolean option selected in the dropdown menu:
  - -AB not =A and B
    - John w/1 Hogan Canandaigua
       Hogan must be next to Canandaigua
    - John w/1 Hogan and Canandaigua Hogan does NOT have to be next to Canandaigua

- With the boolean option selected in the dropdown menu:
  - A w/2 B, this is useful to allow for a middle name or initial (also finds B A)
    - John w/2 Hogan finds
      - -John Hogan
      - -John Francis Hogan (or John F. Hogan)
      - -Hogan, John

- mary w/2 hogan and (john w/2 hogan or francis w/2 hogan)
- Use parentheses to organize complex boolean search strings.
  - Contents of parentheses are evaluated first.
     Logic is evaluated from left to right.
  - This string will give completely different results if the parentheses are left out.

#### Searching & OCR Results

- So far we've seen many ways to tailor our search string.
- These work fine assuming the OCR process has provided "correct" results (Hogan in the source doc is OCR'd as Hogan).
- What can we do when the OCR process is not successful? Examples:
  - Hogan OCR'd and indexed as Hagan
  - The source has a typo, Hogen

## Fuzzy Searching

Search Over 21,790,000 Old New York State Historical Newspaper Pages				
Hogan Canandaigua	all of the words	•	Search	
Items per page 200 Stemming				
Fuzzy searching Fuzziness 0				

- OCR is not perfect. OCR may produce something like Hogam Camandugw instead of Hogan Canandaigua.
- As long as I search for Hogan, I will NEVER find instances where OCR gave incorrect results for Hogan...like Hogam or Hojan or Hagan or ...

## Fuzzy Searching - Cont'd

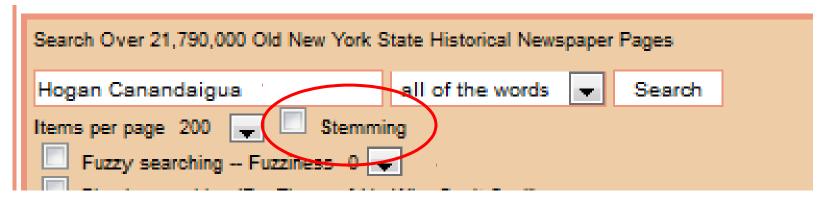
Search Over 21,790,000 Old New York State Historical Newspaper Pages					
Hogan Canandaigua	all of the words	-	Search		
Items per page 200 Stemming Fuzzy searching Fuzziness 0					

- Fuzziness is a way to do non-exact searches. A
  fuzziness of 1 could be thought of as allowing 1
  character to be different than what you had
  searched for.
  - \_ogan or h\_gan or ho\_an or hog\_n or hoga\_ or hogany
- Strategy: First search with NO fuzziness, then try fuzziness of 1, then 2, etc.

### Fuzzy Searching Alternatives

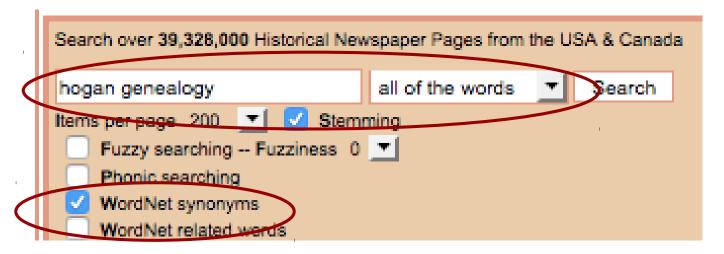
- The fuzzy search in FultonHistory is a great tool, but as an alternative try substituting letters in your search string. This is also a useful technique to use for search engines that do not offer fuzzy searches.
- For example consider searching for a word containing rn. The OCR process could easily yield m instead of rn. So substitute a m and repeat the search.
- A list of potential substitution pairs can be found at <u>http://www.theancestorhunt.com/blog/the-one-absolute-best-way-to-find-more-ancestor-articles-in-historic-newspapers-online#.</u>

## Stemming Searches



- Stemming, when selected, adds extra search words for you. For example, if you search for run, the search engine might also search for runs, running, ran, runner, etc.
- Stemming is off by default.
- Example: Searching for hero does not find heroes.
   Same search with the Stemming box checked, will find heroes.

## Synonym Searches



- Selecting "WordNet synonyms" will cause searches for synonyms of genealogy, like family tree.
- Selecting "Stemming" will cause searches for singular & plural forms.

©2012-2019 Dennis A. Hogan www.dennisAhogan.com

## What Newspapers are included?

- Click on FAQ\_HELP\_INDEX button (upper rt)
- Click <u>Here</u>, which is in the 1st sentence.
  - But results are NOT in alphabetical order
  - So click Down Load This Index as an Microsoft
     Excel file in the upper right corner
    - If you don't have MS Excel, choose the *Save* option to save to your computer. Then use FREE spreadsheet software to open: <u>Google Drive</u>, <u>Zoho</u>, <u>LibreOffice</u>.

## Onondaga County Content

											Time	Period	t									
Newspaper Title	County	1800	1810	1820		1840		1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Skaneatles NY Columbian 1835 - 1853	Onondaga				1835		1853															
Skaneatles NY Democrat 1843 - 1921	Onondaga					1843	1859							1921								
Skaneatles NY Free Press 1892 -1989	Onondaga											1892								1989		
Syracuse NY Daily Courier 1856 - 1898	Onondaga						1856				1898											
Syracuse NY Evening Telegram 1898-1922	Onondaga										1898			1922								
Syracuse NY Standard 1851-1899	Onondaga						1851				1899											
Syracuse NY Daily Journal 1847-1921	Onondaga					1847								1921								
Syracuse NY Evening Herald 1877-1897	Onondaga								1877		1897											
Syracuse NY Post Standard 1900-1921	Onondaga											1900		1921								
Tully NY Independent 1967 - 1978	Onondaga																	1967	1978			
Tully NY Times 1855 - 1945	Onondaga						1855									1945						
Fayetteville NY Bulletin 1901-1989	Onondaga											1901								1989		
Fayetteville NY Luminary 1840 - 1842	Onondaga					1840 1842																
Fayetteville NY Onondaga Independent 1899	Onondaga										1899											
Fayetteville NY Weekly Recorder 1866-1897	Onondaga							1866			1897											
Fayetteville NY Thrift News 1935 - 1941	Onondaga														1935	1941						
Fayetteville NY National Citizen & Ballot Box 1878-1881	Onondaga								1878	1881												
Baldwinsville NY Republican 1843 - 1846	Onondaga					1843 1846																
Syracuse NY Daily Star 1847- 1852	Onondaga					1847	1852															
Syracuse NY Empire State Democrat 1843 - 1844	Onondaga					1843 1844																
Syracuse NY Evening Chronicle 1853 - 1855	Onondaga						1853 1844															
Syracuse NY Gazette & General Advertiser 1828 - 1829	Onondaga			1828 1829																		
Syracuse NY L'UNIONE (Italian Newspaper) 1908 - 1909	Onondaga											1908 1909										
Syracuse NY Onondaga Standard 1841 - 1849	Onondaga					1841 1849																
Syracuse NY Republican 1853 - 1854	Onondaga						1853 1854															
Syracuse NY Sunday Times 1876-1880	Onondaga							1876	1880													
Syracuse NY The Carson League 1851- 1854	Onondaga						1851 1854															
Syracuse NY Weekly Express 1887 - 1892	Onondaga									1887	1892											
Syracuse NY The State league 1851 - 1869	Onondaga						1851	1869														
Syracuse NY Wesleyan 1853 - 1856	Onondaga						1853 1856															
Syracuse NY Religious Recorder 1851 - 1853	Onondaga						1851 1853															
Baldwindsville NY Onondaga Gazette 1846 - 1857	Onondaga					1846	1857															
Baldwinsville NY Messenger 1941 - 1988	Onondaga															1941				1988		
Baldwinsville NY Gazette Farmers Journal 1879 - 1965	Onondaga								1879									1965				
Marcellus NY Observer 1879 - 2007	Onondaga								1879													2007
Baldwinsville NY Gazzette Anniversary 1846-1896	Onondaga					1846					1896											
Syracuse NY Herald 1904-1929	Onondaga											1904		1929								
Syracuse NY Journal 1922-1945	Onondaga													1922		1945						
Cicero NY Patriot 2000-2013	Onondaga																					2000-

## How to Search in a Specific Newspaper

- Assume I want to search the Holley NY
   Standard for time frame of 1931 or
   1932 for surnames Molison and Totter
   within 20 words of each other.
- Select Boolean from the drop down

## How to Search in a Specific Newspaper

- Then enter the following:

  Holley NY Standard and 1931~~1932 and Molison w/20 Totter
  - Name of paper must be in same format as in the fulton history list. (same as used in the results frame)
  - $-\sim$  = 2 tildas (left of the 1 key) indicates year range (year range is optional)
  - -w/20 indicates within 20 words

## Search in a Specific Newspaper II

- Select Boolean from the drop down
- Then enter the following: (Filename contains (Holley NY Standard)) and 1931~~1932 and Molison w/10 Totter

OR

(Filename contains (Holley))...

-"Filename" refers to the PDF filename like what is listed in the results frame on the left

## How to Display the Next Page of the Paper (or the Previous)

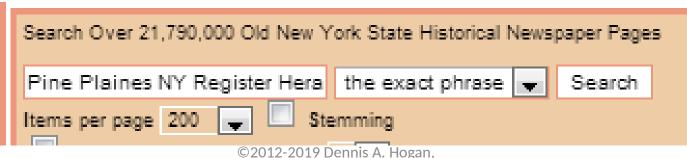
100%

Pine Plaines NY Register Herald 1997 Grayscale - 0520.pdf ...NY Register Herald\Pine Plaines NY Register Herald 1997 Gr

- In the Results frame, the name of the newspaper is listed followed by a 4 digit number & .pdf
- For the Next Page, you want the 4 digit number + 1 & for the Previous Page, you want the 4 digit number - 1
- For the Front page of the issue, you need to work your way back. If the page number is displayed in the current image frame, you may be able to calculate the number for the Front page.

## How to Display the Next Page of the Paper (or the Previous)

- From the listing in the Results frame:
  - select the underlined listing & copy it.
  - Paste into the search box
  - Adjust the 4 digit number as desired
  - Select the exact phrase from the drop down menu
  - Click Search



## Another way to Display the Next Page of <u>the Page</u> (or the Previous)

Click on FAQ\_HELP\_INDEX button (upper right)

Click <u>Here</u>, which is in the 1st sentence.

Click on the desired newspaper (or click Down Load This
 Index as an Microsoft Excel file in the upper rt corner &
 click on the desired newspaper)

Click on the desired folder, which yields
 a list of PDF files (1 file for each page)



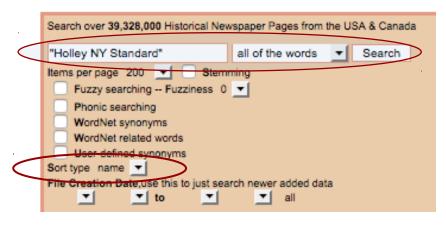
Click on the back arrow (left) to take you back to the PDF listing where you can choose the next or previous (or any) page.

## YET Another way to Display a Sequential <u>list</u> of pages of a paper

Enter the standard

Name in quotes

Select name from



## the "Sort type" pull-down menu

· Click Search

```
Holley NY Standard 1897-1903 - 0621.pdf
...Block, Holley. CROCKERY, FURNITURE AND UNDERTAKING. Filename:
I:\Newspaper 15\Holley NY Standard\Holley NY Standard 1897-1903\Holley NY
Standard 1897-1903 - 0621.pdf ...

956160

Holley NY Standard 1897-1903 - 0622.pdf
...Imf S. C» DMsfltt *□*ha******* Filename: I:\Newspaper 15\Holley NY
Standard\Holley NY Standard 1897-1903\Holley NY Standard 1897-1903 - 0622.pdf ...
```

# Helpful Tools for working with FultonHistory.com

- Graphics Tools to edit an image file
  - Paint: free & easy to use. [(Windows) Start Button > All Programs > Accessories] Note can Not edit PDF files.
  - Photoshop Elements: relatively inexpensive & sort of easy
  - Gimp: free & has learning curve
  - Photoshop: expensive & deadly learning curve
- Snipping Tools to capture a portion of a screen
  - Snipping Tool: free & easy to use. [(Windows) Start Button > All Programs > Accessories]
  - To find others: search on free snipping tool
- Image Organization Tools to organize image files & do basic editing
  - Picasa: free & easy to use not currently supported
  - Vivid-Pic Restore: An new program you might like

## Saving an Article

### Some approaches:

- 1)Screen Shot
  - 1) Open your graphics program & minimize
  - 2)Press *Print Screen* button on keyboard (this copies image of screen to the clipboard)
  - 3)Paste (CTRL-V) to the graphics program

#### 2) Whole newspaper page

- 1) Right click on link in Results frame & Select Save Link as... to save PDF file.
- 2)If editing is necessary, open the PDF file in a graphics program (Paint will not open a PDF.)

### Saving an Article - Cont'd

- 3) Snip relevant portion of the page
  - 1)Open your graphics program & minimize
  - 2) Snip the banner (paper name, date, page) and article. (I also copy the blue link from the results frame.)
  - 3)Paste (CTRL-V) to the graphics program

## Saving an Article - Cont'd

Example (text file saved as pdf):

From <a href="http://nyshistoricnewspapers.org/">http://nyshistoricnewspapers.org/</a>

<u>Livingston Republican. (Geneseo, N.Y.), March 23, 1933, Page 7, Image 7</u>



## Saving an Article - Cont'd

- Suggestion: Create a master source in your genealogy database and attach the resulting image.
- If the image is hard to read, create a transcription in a text file and attach that to the master source also.
- · An Easy Way to create a Transcription:
  - Open your word processing program (Word, notepad, etc.)
  - In the image frame of fultonhistory.com, select & copy the desired text [Don't snip which results in an image file]
  - Paste into the word processing program (editing will be necessary)

### Printing an Article

- First, save the article (see previous slides)
- Then, open the file for the saved article (double click). The program that opens the file should be able to print the file.

## Cite what you Sight

- When you use information from the Fulton
  History site, you are looking at a website, not
  at a newspaper. So, you must cite the website
  as your source (as well as the newspaper).
  - If you view a microfilm, cite the microfilm. If you hold a newspaper in your hands, cite the newspaper.
- Citation Content
  - author, article title, newspaper name, date, page, column, indicate digital image, website name, home page URL, date accessed, optional citation detail

## Cite what you Sight - Cont'd

#### Example Citation

— "Pair to Mark 51st Wedding Date," Rochester (New York) Democrat and Chronicle, 9 Jan 1939, p. 12, col. 5; digital image, Fulton History (http://www.fultonhistory.com : accessed 9 Mar 2013), Rochester NY Democrat Chronicle 1939 - 0333.PDF.

#### Citation Authority

- Elizabeth Shown Mills is the recognized authority on citing historical records. For more information about this topic, see her website at <a href="https://www.historicpathways.com">www.historicpathways.com</a>.
- See Evidence! Explained, section 14.22 for citing online images of newspapers.

# site: Another Way to Access FultonHistory.com

- Google can be used to search fultonhistory.com by using the site: command in the google search box
  - ·site:fultonhistory.com hogan
  - "john hogan" OR "john \* hogan"
     OR "hogan john"
     site:fultonhistory.com

## Another Way to Access FultonHistory.com - Cont'd

- · "Quick & Dirty" way to search Fulton History site
- Works for devices without "Flash" like iPads
- Must use google syntax. No way to simulate important Fulton History features like fuzzy search.
- site: command is useful for GenWeb-type sites!!
  - "john hogan" OR "john \* hogan" OR "hogan john" site:http://ontario.nygenweb.net/ [Ontario County (NY) genweb site]
  - site:http://ontario.nygenweb.net john AROUND(2) hogan

#### Want Your Microfilm OCR'd and Indexed?

"If you are a <u>non-profit organization</u> which has microfilm of any New York State Newspaper or other N.Y. Historical Documents which you would like to have converted to digital format and OCR'd for your non-profit use, and you would allow this site to share a copy of it so all could enjoy, I would consider doing it at no charge. Contact me at tryniski@fultonhistory.com"

# Happy Hunting