***Blake/An Illustrated Quarterly***
**Issue Archive**

Processing HTML Articles

These instructions assume that you have the BQ and BQ-tools repositories synced (using Git/Sourcetree) to a local computer which is equipped with MAMP, and that they are synced to "bq" and "bq-tools" directories, respectively. It also assumes you have access to subscriber-only BIQ issues, on the BIQ subscription site. Because some of the steps involve using our custom-made PHP scripts, you may also need to identify and troubleshoot errors using the "php_error.log" file—these instructions assume that nothing unexpected comes up and everything works smoothly.

1. **Go to the HTML issue** you want, on the BIQ subscription site. (Use Chrome; Safari saves webpages slightly differently.)

    LOGIN:
    http://blakequarterly.org/index.php/blake/login
    username: ******
    password: ******

2. **Download PDFs** for all the articles in the issue (linked from the table of contents).
    2a. For each PDF: in the PDF URL, look for the pattern "keywordvolumeissue" (e.g., essick484), and rename the downloaded PDF using the pattern "volume.issue.keyword.pdf" (e.g., 48.4.essick.pdf). (Use the same keyword as the original, but change it to lowercase if there are any uppercase letters.)
    2b. Once you have all the issue's PDFs, upload them to the "pdfs" folder on both the public BQ server and the BQ-dev server (using an FTP client such as Cyberduck or Fetch).
    2c. Move the PDFs to the "pdfs" folder in your local computer's BQ folder.

3. **Download the HTML and images** for the issue (table of contents* and all articles)
    3a. Use command-S or (in Chrome menu) File > Save Page As
    3b. Set "Format" to "Webpage, Complete" (this saves both HTML and images)
    3c. Use the filename format "volume.issue.keyword.html."
        - *table of contents:* Use the keyword "toc" (e.g., 48.4.toc.html).
        - *article:* Look for the pattern "keywordvolumeissue" (e.g., essick484) in the URL, and use the pattern "volume.issue.keyword.html" (e.g., 48.4.essick.html). (Use the same keyword as the original, but change to lowercase if there are any uppercase letters. This should match the filename of the corresponding PDF.)
    3d. Click "Save." (Notice that, in addition to the HTML file, you get a folder named "[filename]_files", full of images and other supporting files.)
    (*NOTE:* check for links to images not embedded in the page—e.g., "see enlargement"—and save those images separately. These will need to be uploaded with the other images in step 5.)

4. **Fix the HTML image paths and image filenames**.
    4a. Open “bq-tools / bq-htmltransform” and (if it doesn’t contain the folders “old” and “new” already) create folders called “old” and “new”.
    4b. Move the HTML files and associated folders (“[filename]_files”) into the “old” folder.
    4c. Start servers in MAMP. This should open your localhost in your browser.
    4d. In the URL after the localhost, replace “MAMP/?language=English” with “bq-tools/bq-htmltransform”. (This will run a PHP script that copies edited versions of the files to the “new” folder: [1.] copies an edited version of the HTML files—the new version lacks the “[filename]_files/” in the image filepaths, and [2.] copies and renames the relevant images into a folder called “volume.issue”—e.g., “48.4”.)

5. **Move the HTML and image files** to the correct places.
    5a. Move all HTML files from “bq-tools / bq-htmltransform / new” to “bq / html”.
    5b. Upload image folder labeled “volume.issue” (e.g. “48.4”) from “bq-tools / bq-htmltransform / new” to the “img / illustrations” folder on both the public BQ and the BQ-dev servers.
    5c. Move image folder labeled “volume.issue” (e.g. “48.4”) from “bq-tools / bq-htmltransform / new” to the “bq / img / illustrations” folder on your local computer.
    5d. Before syncing BQ-tools in Sourcetree, you will want to delete all the files in the “old” directory. You may, however, want to wait until you are sure the processed files work correctly.

6. **Test the HTML issue locally** to make sure nothing is broken.
    6a. Change the URL in your browser from localhost “bq-tools/bq-htmltransform” to localhost “bq”. This will load BQ, hopefully including the latest HTML issue you added at the top, complete with a miniature cover image.
    6b. Open the issue to check the table of contents and its images, and open all the articles and PDF links to make sure all the HTML and PDFs open correctly and all the images load correctly.**
    6c. Make sure symbols are displaying correctly.***
    6d. If everything looks right, sync to the BQ repository in Sourcetree.

7. **Test the HTML issue on BQ-dev** to make sure nothing is broken.
    7a. Sync the repository to BQ-dev. (Mike Fox has instructions for this.)
    7b. Open bq-dev.blakearchive.org. The latest HTML issue you added should be at the top, complete with a miniature cover image.
    7c. Open the issue to check the table of contents and its images, and open all the articles and PDF links to make sure all the HTML and PDFs open correctly and all the images load correctly.

8. **Index for searching**.
    8a. Change URL to localhost “bq-tools/bq-htmlforsolr/” and load the page. (This runs a script which—if “bq-tools” and “bq” are in the same directory—should create an index file for the html files in “bq / html”. This index file is “solrFile.xml” and should appear in the directory “bq-tools / bq-htmlforsolr / new”.)

8b. The file "solrFile.xml" in the directory "bq-tools / bq-htmlforsolr / new" should be imported into Solr for both the public BQ site and BQ-dev. (Mike Fox can do this or provide instructions.)

8c. Make sure the search can now find the new issue. Go to bq-dev.blakearchive.org and search for a keyword in one of the new articles, and sort by date so you can find the newest HTML article if it has been indexed.

9. **Create RDF data** for NINES.

8a. Change URL to localhost "bq-tools/bq-html-to-rdf/" and load the page. (This runs a script which—if "bq-tools" and "bq" are in the same directory—should create RDF files for the NINES index in the directory "bq-tools / bq-html-to-rdf / new".)

8b. Locate the RDF files with names matching the new issue. Five years from now, when we publish this issue, these RDF files should be sent to NINES for indexing. (Each time we publish an issue or issues, the matching RDF files should be sent to NINES.)

_____

* TABLE OF CONTENTS (TOC) USING FRAMES

Some tables of contents use frames, and therefore can't be saved with the usual "Save As" in the browser—the relevant HTML code will not be saved. If you encounter problems, follow the following steps for the table of contents.

I) **HTML code**

right click — View Frame Source
select all (of frame source code) and copy (NOT "Save As," which will include the tags used to display the code)
paste into a blank file in TextWrangler
save file as volume.issue.toc.html (e.g., 40.4.toc.html)
edit image path, removing everything but the final file name (after the last /)
move file to BQ "html" folder

II) **Image**

manually save cover image (right-click, "Save Image As")
move to volume.issue (e.g., 40.4) folder in img/illustrations/

III) **Test as above**

** CHECK ASSOCIATED FILES (IMAGES AND PDFs)

Note: when manually saving images, name them using the last section of the URL and (if there is no file extension) add ".jpg"

e.g., for: http://blake.lib.rochester.edu.libproxy.lib.unc.edu/blakeojs/index.php/blake/article/viewFile/148/essick484html/1046
save as "1046.jpg"

Also note: I'm assuming that your local host root is http://localhost:8888/ — but substitute whatever MAMP loads up.

I) **Check for missing PDFs** (http://localhost:8888/bq-tools/bq-htmlcheck/toc-pdf-links.php)

look for "not downloaded to archive"

go back to the relevant TOC on the subscription site and download the missing files which are linked from that article
put the downloaded files in the relevant folder (pdfs) on both servers (BQ and BQ-dev) and locally

II) **Check for missing embedded images** (http://localhost:8888/bq-tools/bq-htmlcheck/img-src.php)

look for "not downloaded to archive"
go back to the relevant article on the subscription site and download the missing files which are embedded in that article
put the downloaded files in the relevant folder (that would be volume.issue—e.g., 48.4—in img/illustrations) on both servers (BQ and BQ-dev) and locally
you can ignore the following—images from the header and footer, which we don't display:
fulltext_open_medium.gif
fulltext_restricted_medium.gif
UofR.gif

III) **Check for missing linked images** (http://localhost:8888/bq-tools/bq-htmlcheck/file-links.php)

look for "not downloaded to archive"
go back to the relevant article on the subscription site and download the missing files which are linked from that article
put the downloaded files in the relevant folder (if they are images, that would be volume.issue—e.g., 48.4—in img/illustrations) on both servers (BQ and BQ-dev) and locally

*** <u>CHECK SYMBOL DISPLAY</u>

*Note:* this requires a little bit of PHP editing, so it wouldn't hurt to have a developer look over your shoulder the first time you do it.

*Also note:* I'm assuming that your local host root is http://localhost:8888/ — but substitute whatever MAMP loads up.

I) **Load** (with MAMP on): http://localhost:8888/bq-tools/bq-htmlcheck/characters.php
Search for any entries with a non-zero count in square brackets, and with symbols listed. (These symbols are not being properly processed and will not display correctly in the listed article.)

II) **For each listed symbol**:

ii.A) **Open the corresponding article** as rendered in BQ, and open its raw HTML file.
(For example:
Rendered: http://localhost:8888/bq/47.1.bentley
Raw: http://localhost:8888/bq/html/47.1.bentley.html )
Search the symbol in the raw HTML, where it should be rendered correctly.
(*Note:* if the symbol is a variant of a Roman letter, the browser on-page search will—unfortunately, for our purposes—find every version of that Roman letter. You can open the HTML file in TextWrangler instead to get more precise searching.)

Find the same place in the rendered HTML (by searching for a nearby word), and you will (almost certainly) see the symbol incorrectly displayed.

ii.B) **Search the symbol** on graphemica.com to get the html code for it: “HTML Entity (Decimal)”

ii.C) **In BQ, open include/functions.php** in TextWrangler.

*Note:* be very careful in this file, as a mistake in it can disable the whole local instance of the BQ website until fixed.

Find the bottom of the HTML character list.

The entries look like this: $list['Ā'] = 'Ā'; )

Add a new entry with the new special character and its html code.

*Note:* you must follow the format exactly, semicolons and single quotes and all.

Save the file.

ii.D) **Reload the rendered HMTL** of the relevant article and find the same place again. You should see the symbol correctly rendered now, matching the raw HTML.

ii.E) **Clear the symbol from the errors** by adding the symbol to the list in characters.php

In BQ-tools, open bq-htmlcheck/characters.php in TextWrangler.

*Note:* again, be very careful editing this file, so as not to disable your error-finding tool.

Find the bottom of the HMTL character list.

Again, the entries look like this: $list['Ā'] = 'Ā'; )

Copy and paste your new character’s entry.

Save the file.

Load http://localhost:8888/bq-tools/bq-htmlcheck/characters.php — the character you’ve fixed should no longer be listed.