

Blake/An Illustrated Quarterly Issue Archive

Solr Search Indexing

Note: Because of (our) HTML's inconsistent structure, we can't directly import it into Solr, but instead have to preprocess it using PHP. XML files can, however, be directly imported. Consequently, we import HTML and XML data into Solr in different ways.

Sourcetree (Git): Pull latest BQ code.

MAMP/browser: Update **HTML** Solr file.

Run local server in MAMP.

Run script in browser: <http://localhost:8888/bq-tools/bq-htmlforsolr>

Locate file in folder: bq-tools / bq-htmlforsolr / new / solrFile.xml [file's "modified" date/time should be the time you ran the script just now]

Cyberduck/Fetch: Upload **XML** files and **HTML** Solr file to *both* Solr servers.

Servers:

Webapp-dev (Solr server for BQ-dev): webapp-dev.libint.unc.edu

Webapp (Solr server for BQ): webapp.lib.unc.edu

For *each* server:

Open the folder data/solr/bq on the server.

Delete the "docs" folder on the server and upload the "docs" folder from your local BQ folder.

Delete the "solrFile.xml" file in the "html" folder on the server, and upload the "solrFile.xml" file from your local BQ-tools folder (see above).

Terminal: Import the files from the servers into the matching Solr instances. For each server:

Log in with SSH (secure shell). Paste the following command (selecting based on the server, and substituting your Onyen for "afmccune") into Terminal, and hit Enter.

[for Webapp-dev — i.e., Solr for BQ-dev]

```
ssh -L 8200:127.0.0.1:8200 afmccune@webapp-dev.lib.unc.edu
```

[for Webapp — i.e., Solr for BQ]

```
ssh -L 8200:127.0.0.1:8200 afmccune@webapp.lib.unc.edu
```

Enter your password when prompted, and hit Enter.

Enter the following commands, hitting Enter after each. These will move to the folder which "solrFile.xml" is in, and then import that file into Solr (indexing all HTML issues).

```
cd ../../data/solr/bq/html
```

```
curl http://127.0.0.1:8200/solr/bq/update?commit=true -H "Content-Type: text/xml" --data-binary @solrFile.xml
```

Browser: While still logged in to the Solr server in Terminal (again, you have to do this for each server), import the XML files into Solr by going to the following URL:

```
http://127.0.0.1:8200/solr/bq/dataimport?command=full-import
```

Terminal: Exit one server so that you can log in to another with SSH (and repeat the above steps for the other server). Enter the following command and hit Enter.

```
exit
```

At this point, the index(es) should be updated. You'll want to test them by doing test searches on BQ-dev and BQ to make sure the latest text is indexed. (Remember than on BQ, only published articles will show up in searches.)

For example, to see if the latest HTML issue has been added, search something ubiquitous (like "Blake") on BQ-dev, then sort by "newest" to see if the newest HTML issue shows up in the search.

Bear in mind: These instructions are for adding/updating articles in Solr. To delete old entries (say, if a filename/idno is changed and you don't want search results to show up for the old, nonexistent URL) you can wipe the whole index and then add/update using the following instructions. To wipe the whole index, use this command after logging in with SSH:

```
curl http://127.0.0.1:8200/solr/bq/update?commit=true -H "Content-Type: text/xml" --  
data-binary '<delete><query>*:*/query</query></delete>'
```