How can I grab a website and place it on my server for a faculty member who wants to archive student web projects?

Several times I've been asked by faculty to copy student web projects over to our server, or in one case to burn a CD. The utility I use is wget and among its advantages, it will rewrite the links in each page to local links so they'll work on your server.

- http://www.gnu.org/software/wget/

Notes – Mike Franks, Social Sciences Computing – 3 May 2006

The hard part was getting all the parameters right, so that it didn't dig too deeply, but did go far enough to get everything.

If you're interested, here are my student programmer's notes from the last time we did this.

The command I finally ended up with was:

wget -Pccsa -Ducla.edu -nH —cut-dirs=3 -r -l inf -k url_to_copy_goes_here

-P for the directory to save to, -D to limit what hosts of URLs should be followed, -nH to not make a directory based on the host URL, —cut-dirs=3 to have it not recreate the site's hierarchy for directories, -r to make it process pages recursively, -l inf to follow the recursion infinitely, and finally, -k to convert links to local, relative links. Also, I had to create the directory that I was to save the data to, and I entered the command while in the directory above where I saved them to.

Revision #4
Created 2006-05-03 13:28:25 UTC by Franks, Mike
Updated 2006-05-03 13:35:59 UTC by Franks, Mike