Notes for Archiving

Background

At the end of each academic year, we need to make a static copy of the courses that have just completed. The static copy is then moved to the archive server - (archive.st-andrews.ac.uk).

This was previously achieved by using a tool called HTTrack to get the static pages.

Before these copied course pages were moved to the archive server, they had various updates applied such as: Adding a green banner to the top of the page and various link and title changes to fit the form of their new home and to flag that they were archived.

These edits where completed using the find and replace function in the code editor.

Issue

There is an issue with this process as the HTTrack tool can no longer be used to get the static pages due to system changes that block it’s use.

This solution uses PHP and hopes to replicate the HTTrack tool by doing the following.

Note this code is run locally.

1. Get a list of the course URL’s for UG and PG:
   1. Note – run the script for each UG & PG separately.
   2. Use a test sample initially - 2-3 files.
2. Loop over these files
3. Create a local folder structure matching that of each URL form the list.
4. Add an index.html file to each folder.
5. Get the content from each URL.
6. Update the contents making changes; Add the green banner and breadcrumb links etc.
7. Write the updated files to the index.html files within the new folder structure.
8. There was an issue with the removal of the footer, so this was done once the above was completed, using Find and Replace in the code editor.
9. Check the files locally and make sure they look fine.
10. Move these files to the archive server – courses – add a folder for the appropriate year.
    1. Note: The URL structure doesn't match the structure of the current archived content, so manual creation of the folder structure on the server is necessary. Afterward, the appropriate files should be copied to each of these folders.