Aesthetics Program Web Archive Proposal

Danielle Janney

School of Information, College of Arts and Sciences, University of South Florida

LIS 6515: Web Archiving

Professor Jinfang Niu

July 5, 2024

Author Note

This document is for learning purposes only. Storm Coast University and the proposed web archive are fictional.

Aesthetics Program Web Archive Proposal

Vision and objectives

Storm Coast University (SCU) is organizing an innovative Aesthetics program within its College of Arts in response to emerging trends in popular and workplace cultures. The Aesthetics program coordinators and associated professors have reached a solution to create a web archive that will enhance the discoverability of knowledge and criticism of related fields to produce a dynamic structure of scholarly development.

Appraisal, selection, and scope

Websites to be selected, in whole or in part, for the archive will be related to the following subjects: aesthetics, art, art history, art theory, design principles, and philosophy—of art, history, language, mathematics, mind, politics, religion, and science. Selected websites, so far, include:

- Art F City - http://artfcity.com/
- Art Newspaper - https://www.theartnewspaper.com/
- ARTnews - https://www.artnews.com/
- Artsy - https://www.artsy.net/
- Frieze - https://www.frieze.com/
- Hyperallergic - https://hyperallergic.com/
- Krisis - https://krisis.eu/
- MoMA: Magazine - https://www.moma.org/magazine/
- PhilArchive - https://philarchive.org
- Project Gutenberg: Books in Philosophy - https://www.gutenberg.org/ebooks/bookshelf/57
- Sacred Texts - https://sacred-texts.com/

- Tanner Lectures - https://tannerlectures.utah.edu/

- Whitehot Magazine - https://whitehotmagazine.com/

It is anticipated that websites will be added to this list as SCU personnel see fit, whether from personal research or student learning products.

Data capture

HTTrack will be used to download the selected websites on a dedicated PC. HTTrack has been chosen over other web crawling options because it is free, easy to learn, allows HTML editing, and has the option to update existing downloads. Each website will be given its own name, and the category of Art or Philosophy, to begin the project of copying web content. The preference and mirror options in the HTTrack window can set a proxy, scan rules, limits, and flow control; and offer options for links, building local structure, spider crawls, file types and MIME identity, browser ID, and log, index, and cache. Expert options can set a primary scan rule, travel mode, global travel mode, and rewrite links.

Basic guidelines

The following guidelines have been agreed upon to responsibly represent relevant websites and reduce potential problems in downloading:

- Websites that feature regularly uploaded content will be crawled weekly, whereas more static websites will be crawled every six months for any changes.

- Most file types will be accepted for archiving, though efforts will be made to reduce the inclusion of advertisements.

- Select 'Get HTML files first' and 'Test validity of all links' boxes in Links tab of crawl options

- Select "No error pages,' 'No external pages,' and 'Do not purge old files' boxes in Build tab of crawl options

- Select 'Do not re-download locally erased files' box in Log, Index, Cache of crawl options

Storage and preservation

Data captured from HTTrack will be transferred from local PC storage to a Backblaze cloud storage account. The data upload schedule from the local storage to Backblaze will be continuous as long as the web archive project is pursued and will omit local files not related to the web archive. In addition to the original downloaded websites, the final edited archive website will also be uploaded to Backblaze.

Resources and workflow

At SCU, hands-on learning is valued and the motivation to provide opportunities to computer and information science students is comprehensive. As the head of the Aesthetics program development approves websites for archiving, a professor of computer science has agreed to collaborate and lead a high-level group of qualified students to help with streamlining and maintaining the web archiving project; this will alleviate the Aesthetics department staff who may not be familiar with technical processes and reduce overall operation costs. The web archive will be hosted separately from SCU's website on a dedicated PC to reduce server overload.

The computer science professor and students will be tasked with organizing the downloaded websites--as outlined by the Aesthetics head--and bulk modifying the HTML code with SCU's license of Microsoft's Visual Studio to create the web archive site. Examples of tasks are to batch find-and-replace the archived website names with the web archive's scuaesthetics.com domain name; insert navigational HTML code; set up a login form on the scuaesthetics.com index page; and employ RDFa metadata tags.

Downloaded website files from the dedicated PC will be selected, compressed into a ZIP

file, and made available to students to modify with HTML codes and RDFa tags. Student modifications will similarly be compressed into a ZIP file and submitted to the professor for review and submission.

Budget

Contractual Services	Expenses	Comments
GoDaddy	$44	Domain name subscription; price for first three years
Backblaze	$297	Cloud storage; price for three years of single dedicated PC storage
Total	$341	Approx. $113.67 per year for the first three years

Quality assurance

As part of the students' responsibilities in the web archiving project, they will be asked to click through the copied websites to ensure the operability of the archive after the HTML modifications have been made. Any broken links or error messages will be brought to the computer science professor's attention to resolve.

Access and use

The Aesthetics program web archive will be made available to SCU students only and require login credentials through the library's website or online learning management system.

Risk management

SCU recognizes it is not the owner of the content selected to digitally archive for the

Aesthetics program and therefore will include a disclaimer on its web archive site that states:

> Copyright in all forms of the publications, both within the SCU Aesthetics Archive and
> on the publisher's sites, are held by the authors and/or publishers of the
> publications. We do not own the rights to materials in our collections. You should
> determine if an item is protected by copyright or in the public domain and then satisfy
> any copyright or use restrictions when publishing or distributing materials from our
> collections. Transmission or reproduction of protected items beyond what is allowed by
> fair use or other exemptions requires written permission from the copyright holder.

References

Backblaze. (n.d.). *Computer cloud backup*. Retrieved July 5, 2024, from

      https://www.backblaze.com/cloud-backup/pricing

Bragg, M., & Hanna, K. (2013, March). *The web archiving life cycle model*. The Internet

      Archive. https://archive-it.org/static/files/archiveit_life_cycle_model.pdf

GoDaddy. (n.d.). *Web hosting*. Retrieved July 5, 2024, from

      https://www.godaddy.com/hosting/web-hosting

HTTrack Website Copier. (2007). *Option panel*. https://www.httrack.com/html/step9.html

Library of Congress. (2020, December 17). *Legal*. https://www.loc.gov/legal/#copyright