An Approach to Increase Availability of Cultural Archives: Pine Mountain Settlement School. Tradd Schmidt, Bethanie Williams, and Mario Nakazawa. Computer Science Department, Berea College, Berea, Kentucky, 40404.

Abstract

The Pine Mountain Settlement School has an extensive collection of interesting and culturally important material that is slowly being digitized, but the images and metadata were incorporated into a WordPress blog about the school. This platform is not usable for research because archived items are often stored separately from its metadata in the backend, and they are associated in a blog post, often interspersed with interpretative information. To make connections or search items, the user often travels from post to post. Even the platform's search function often does not work as expected. We developed a web crawler and web scraper using a Python library called BeautifulSoup that extracts from the entire blog site all the filenames of visual and document photographs, any relevant metadata that can be associated with these images, pairs the information together, and finally outputs the combined data into a comma separated value spreadsheet. All original scanned items are currently being uploaded into an archiving platform called CONTENTdm along with the corresponding extracted metadata, making searching and research more effective. The ultimate goal is to provide an effective tool for the community members and other researchers of the Pine Mountain Settlement School to access the digitized archival material.

19th Annual Berea College Undergraduate Research and Internship Symposium, October 18, 2019, Berea College, Berea, Kentucky (Poster Presentation)