

Saving the spirit: Digitizing the inventory of the Summer School of Alcohol Studies

Daniel Geary, MA, MLIS '14

*Center of Alcohol Studies
Rutgers, The State University of New Jersey*

For the past few months, CAS Library has been working with a document management software company to digitize pieces of its collection.

One sub-section is the collection of Summer School materials. These include: lectures, Alumni News, photos, rosters, brochures, flyers, and registration cards, with some items dating back to Jellinek's first Summer School session in 1943. We chose its software due to its zonal OCR and access permissions abilities.

Zonal OCR allows the scanner to choose a particular area to scan and then automatically translate to text and index. While the default is using a rectangular selector, the program can even use the surrounding pixels, such as a title, as points of reference to make the scan more accurate. Our first attempt at zonal OCR used registration cards, which we have for all of the educational opportunities the Center has ever offered dating back to 1943. These have the names of alumni in the top left corner. If we can use zonal OCR accurately, then we will be able to top-load a stack of registration cards in the copy machine and let the computer do the rest. We will need to adjust the program a bit for each year, but if zonal OCR works well with the registration cards, then we can try using it on any type of

form, thus saving time, energy, and money in the digitization effort.

After attempting the process, we found that it did work well. We were afraid that cards in which the year was printed on a line would disrupt the zonal OCR, but there was no issue. Although the text was in a slightly different location for each card, even within the same year range (because the print was added with a typewriter rather than printed from a computer), the system had no trouble locating the font, translating it to computer-readable text, and inserting it in the correct index field. We could not have multiple zonal OCRs happening simultaneously. For instance, we could not run a card through the procedure and have the zonal OCR pick up and index both the Name and Address. We had to scan each card twice, once for the Name and once for the Address. But, this was no hindrance compared to inputting all the information manually.

The second plus is access permissions. Due to the Center's sensitive subject matter, we need to make sure that the library's digital collection can differentiate between in-house and public files. The in-house files include personal records, like the registration cards. Their use is limited to official access only; any use by the public would be a violation of

privacy. Additionally, there are copyrighted pieces, such as the *Quarterly Journal of Studies on Alcohol*. The range between public and private material led the library to establish various levels of access, with the software allowing rights to be assigned at the point of digitization.

In our access permissions workflow, undergraduate assistants scan documents and deposit them in the central folder. They check for readability and approve the document for review. Then, graduate students or librarians approve OCR processing. And lastly, grad students or librarians double-check that the item is correctly placed in the in-house or public folder. All in all, the staff checks the document four times. Plus, the program does a regular sweep to make sure all files have made it to the correct directory. Therefore, by the end of

a document's journey it will be checked, double-checked, triple-checked, and so on.

At the start, we did have some trouble with documents going to the correct folders, as shown in the workflows section of the poster. However, these problems would be a document skipping processing or moving back to the initial folder after being approved. One file was accidentally copied millions of times. But, there were never any problems in which a file that was supposed to be in-house went public. And, working with the vendor, we were able to figure out how to adjust the system and have the documents sent to the correct folders at the correct time.

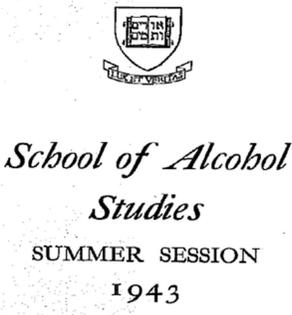
With these tools at our disposal, this software has the potential to be a valuable asset for digitizing our Summer School collection and other pieces of CAS Library's growing digital footprint.



RUTGERS

Saving the Spirit: Digitizing the Inventory of the Summer School

By Daniel Geary



Our Goals

Started at Yale in 1943 by E.M. Jellinek, the Summer Session at the School of Alcohol Studies became a large part of the organization. Now, as we are working towards digitizing a large portion of our collection, we are trying to bring together information about the Summer School. We would like to use document management software to index and OCR various items including: lectures, Alumni News, photos, rosters, brochures, flyers, and registration cards and to create an online database with sub-sections for public and in-house use and field and full-text advanced search.

Zonal OCR

One advantage of our document management software is zonal OCR. While OCR can translate a document into searchable text, zonal OCR can translate specific areas. A prime example is registration records. With names in the top left, Zonal OCR should allow us to draw a box around the desired area (shown in red), OCR only that, and then automatically move it to a corresponding index field. The scan can use surrounding pixels, like a title, to be more accurate. We will need to change it for each year's form, but if this method works it will save time and energy.

Document Management

- [-] Folders
 - [+] MFP Scans
 - [+] MFP Summer School
 - [+] Processing
 - [+] Rescan Queue
 - [+] Review Queue
 - [+] CARRF
 - [+] In house - CAS Archives
 - [+] Public - CAS Archives
 - [+] Summer School

Workflow

Undergrad interns scan and send documents to a central MFP Scans folder. The interns check for readability and send the document to the review queue. Grad students or librarians approve OCR processing. And lastly, grad students or librarians assign the item to the appropriate, public or in-house folder. As a result, the document is checked by staff four times, while an additional automated check verifies the document went to the appropriate folder.

Sensitive Materials

Due to the nature of our specialization, we need to be careful with our sensitive records. Our registration cards are a prime example. While we wanted everything digitized, we needed to make sure only certain items would be publicly available. Our document management software offers a solution in the form of access permissions folders with items being assigned public or in-house. However, it is not that simple. Each item must be looked at carefully. So, as to avoid any problems with accidental publication, we have a strict workflow.

Based on information from the Rutgers' University Center of Alcohol Studies, 2014.

