CSU San Bernardino

"People are very happy with the speed of the scanner, the way it seems to auto-crop, and that it converts things quickly into the format they want and then easily uploads the file to Google Drive and Gmail."

– Jill Vassilakos-Long, Government Information Specialist and
Coordinator for Special Collections

Scannx Helps Tackles OCR Challenges In Special Collections

When Jill Vassilakos-Long walked past the Scannx booth at the midwinter ALA convention, she wasn't looking for a new scanner. But when she overheard a discussion about the superiority of its integrated ABBYY OCR engine – the most accurate on the market – she changed her mind.

"We'd had a lot of problems with OCR," she recalls. As coordinator for CSU San Bernardino's Special Collections in the John M. Pfau Library, Vassilakos-Long oversees a number of projects that digitize collections into searchable PDF documents, but found that many older materials were incompatible with the OCR engine that came with their existing flatbed scanner software.

Particularly problematic, were typescripts – oral histories transcribed by typewriter in the 1980s. Typewriters tend to create documents with some letters lighter or darker than the rest, or slightly out of alignment. That can prevent some OCR engines from recognizing the words and make the resulting document less searchable.

Scannx uses ABBYY FineReader OCR software, with unmatched text recognition accuracy and conversion capabilities, to turn scans into searchable and editable formats, including PDF and Word, virtually eliminating the need to retype or

reformat the materials. It supports text recognition and conversion for English, French, Italian, German and Spanish text, with an optional upgrade to support up to 189 languages. And Scannx uses the ABBYY OCR engine to deliver the fastest OCR image processing. Other systems wait until scanning is completed to start the OCR process, but the Scannx Book ScanCenter OCRs each page as the user continues scanning.

Book ScanCenter Makes Scanning Easier and Faster

Vassilakos-Long installed the Scannx Book ScanCenter in the Special Collections Reading Room, and was immediately pleased with the results. "We've been using the Scannx for manuscripts, newspapers, and texts," she said, "and have been happy with the OCR." In addition, she said, the students "really like it. It's much faster and easier than our other scanner. The touch screen is very clear."

She sees the Book ScanCenter as an important driver in the ongoing push to digitize the Special Collections materials. "We have special collections that have not really been curated until recently. Now we're working our way through it."

Most special collection materials are available only in the reading room, and today's patrons – and donors – expect to be able to access them online. An area publisher, for example, donated a complete run of the Inland Empire Business Journal, the newspaper he founded. At the time, the newspaper was too large to fit on the old flatbed scanner, so the library was using a digital camera held over the page to capture images of each issue, then cropping the images with Photoshop. The BookScan Center, with its oversized scanning bed and auto-crop feature made the job easier and faster. "I wish we'd had it earlier," she said.

The Book ScanCenter also helps Vassilakos-Long build two new local history collections. "In San Bernardino, we have a large Latin American community with a rich history and a large African American community with a rich history," she said. "Many people are happy to contribute to the collection, but want to retain their

original photographs and other materials. Through scanning we can take a copy, make it available to researchers, and return the originals to the people who are helping us build these collections."

More pressing, however, was a project to get as much of CSU San Bernardino's own history online in searchable formats in time for the university's 50th anniversary in 2015. Those materials included things like old course catalogs, and bound copies of the student newspaper, The Coyote Chronicle, which were much easier to scan thanks to the patented Book-Edge scanner design of the Book ScanCenter. The beveled edge of the scanning bed allows each page of a book to lay flat on the glass, protecting the spine and eliminating the blackened and distorted "gutter" for clearly readable text and better OCR.

University Eyes Taking Scannx Beyond Special Collections

While the Book ScanCenter was first installed within Special Collections, Vassilakos-Long says the library staff in other areas have taken notice. "The people in reference have been talking about the possibility of getting one to try out for public service," she said, recommending they get one. "I told them that the menuing system was clear. It is kind of like a camera set on point and shoot: it does everything for you at a pretty good level. With students, that would be good, because they won't go home and find out they have something they can't use."