RSC19
OCLC RESOURCE SHARING CONFERENCE

Steps Towards Accessibility: Improving the User Experience through OCR

Melissa Eighmy Brown & Carol Nelson
University of Minnesota Libraries

Austin Smith & Hilary Thompson
University of Maryland Libraries

#OCLCRSC19

OCLC

Melissa Eighmy Brown

Manager, Interlibrary Loan Borrowing &
Digital Delivery Services
University of Minnesota Libraries

Hilary Thompson

Acting Director, User Services & Resource Sharing
University of Maryland Libraries

Carol Nelson

Minitex Resource Sharing Manager
University of Minnesota Libraries

Austin Smith

Resource Sharing Supervisor
University of Maryland Libraries

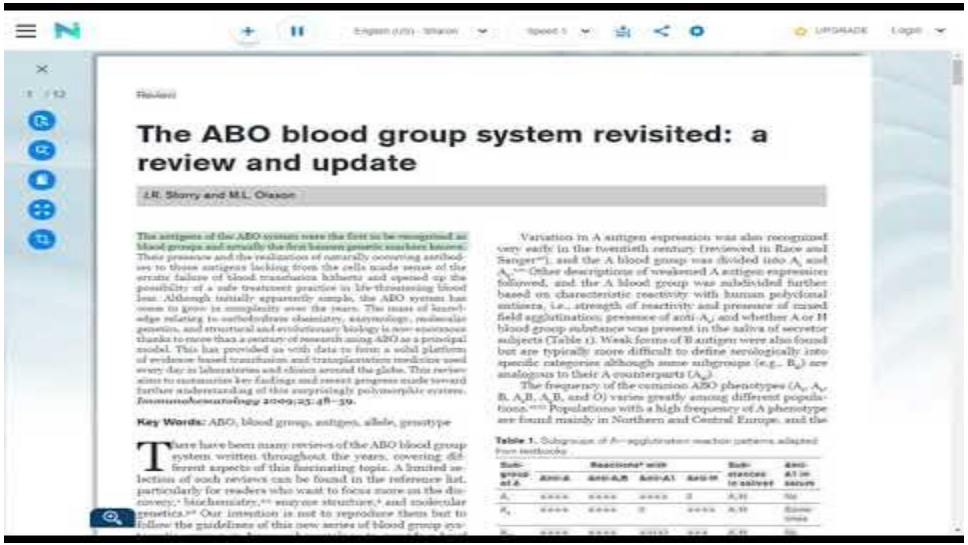
What is OCR?

Optical Character Recognition (OCR) is the recognition of printed or written text characters by a computer.

Benefits:

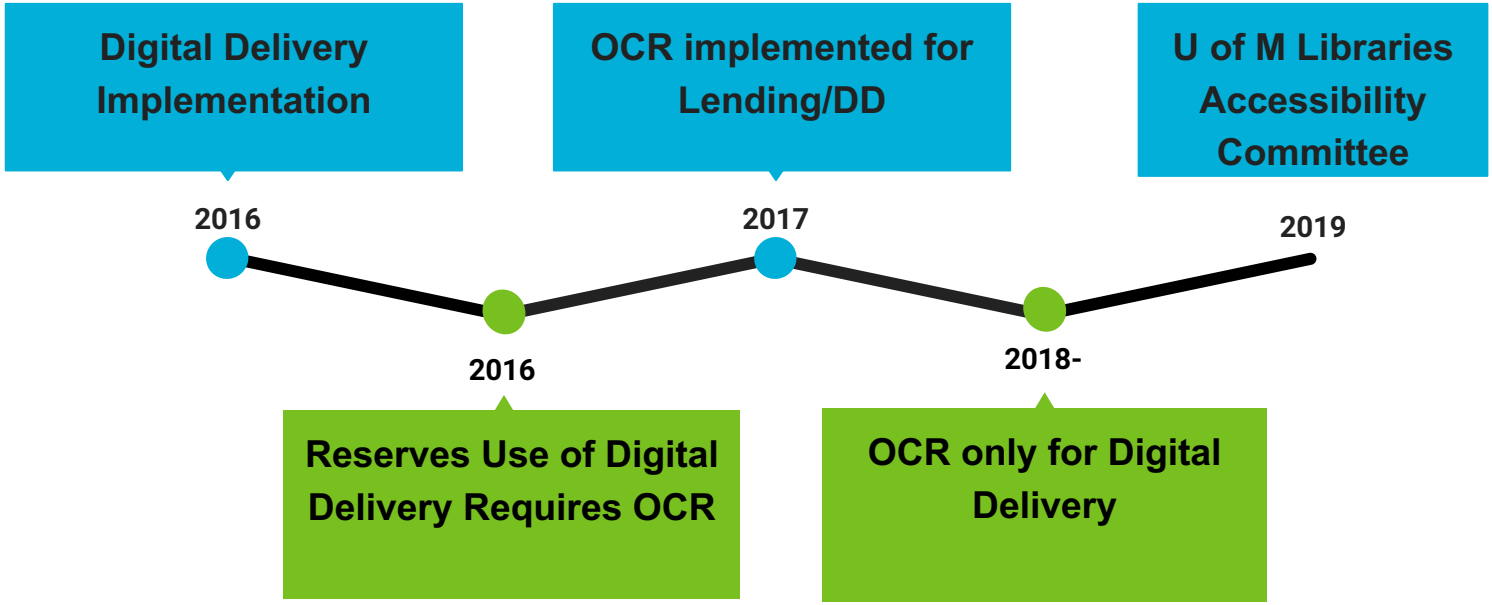
- Searchability
- Text to speech capabilities
- Screen reader compatibility
- Copy and paste
- Supports text mining

OCR + Text to Speech Demo

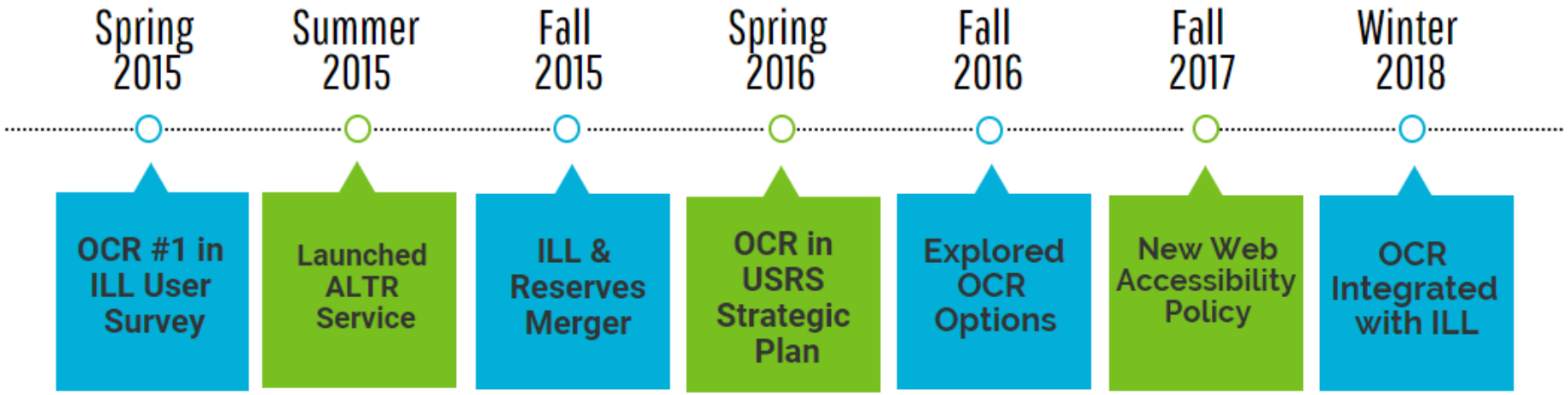


https://youtu.be/MGE9VTmBtDU

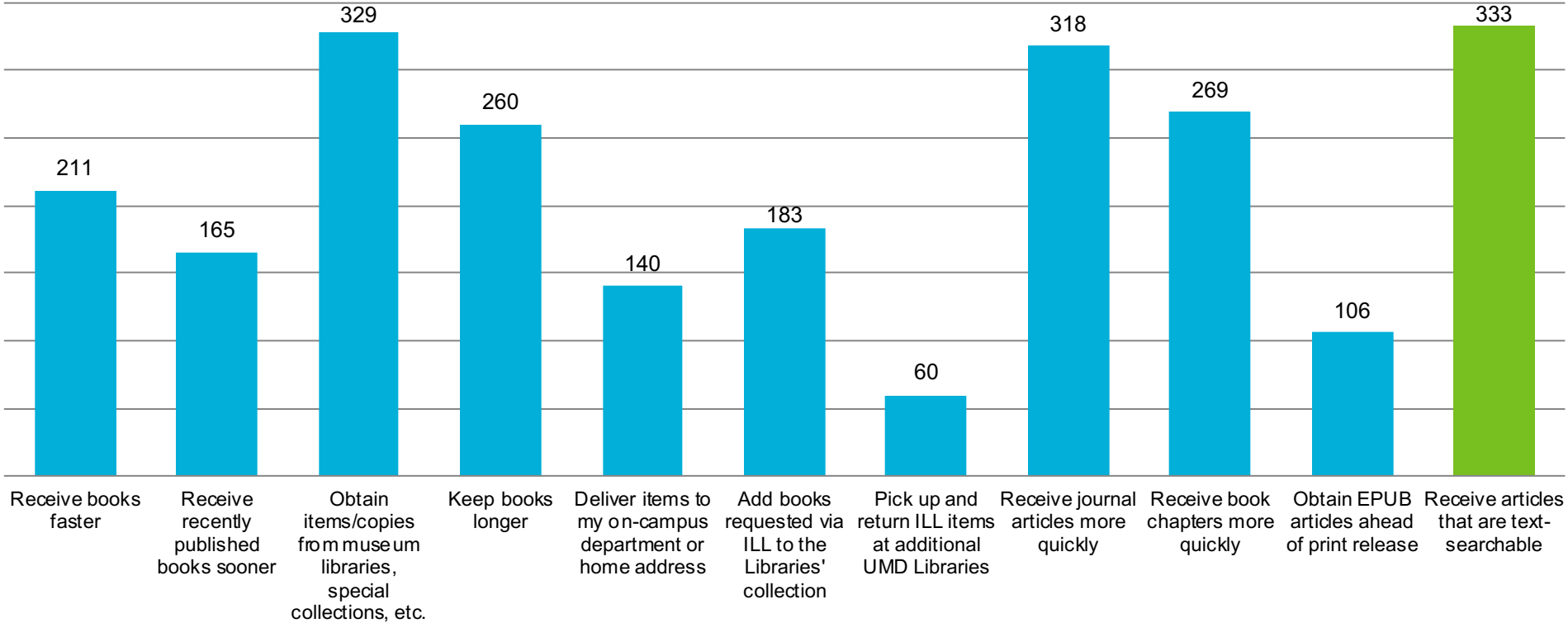
Why OCR at Minnesota?



Why OCR at Maryland?



2015 ILL UMD User Survey Results: Desired Service Enhancements



Fact or Fiction?

"When ILLiad receives PDFs, the OCR from the lender is not retained."

Fiction (provided your cover sheet has OCR).
See ILLiad documentation.

"It's not really important to users. I've never had a complaint."

(Most Likely) Fiction. Have you asked your users?
→ 333 of 1,251 UMD respondents (27%) said it was.

Fact or Fiction?

“It would take too much staff time/effort.”

“It would slow delivery to our users.”

“It would cost too much.”

Whether fact or fiction depends on implementation.

University of Maryland: Borrowing & Document Delivery

Choosing a System

Automated delivery 24/7

- Server-based, not workstation-based
- No staff mediation required
- Minimal IT maintenance

Language & Script Coverage

In 2018 ~15K articles delivered in 32 languages & 10 scripts

Normal Workflow

Articles processed automatically when they arrive.

OCR complete within 10 minutes of delivery

(no staff intervention)

However:

Processing only handles languages written in Latin script.

In 2018: 14,600 articles (97.5% of total) in 16 languages

Exception Workflow

In some cases, default workflow cannot provide OCR

In 2018 375 articles (2.5% of total) in 7 scripts

Patron can request OCR through their ILLiad account:

Electronically Received Articles

Transaction	Title	View	Size	Author	Expires	Delete	PDF not searchable?
1232799	College & Research Libraries: Borrowing Latin American Materials in the Big Ten Academic	View	5.62 MB	Thompson, Smith, Ostos, Gardinier	12/13/2018	Delete	Request OCR

ILL staff process & re-deliver within 1-2 business days

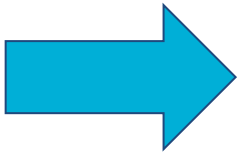
Challenges & Considerations

- Page count must be estimated in advance
- Imperfect language differentiation
- Decreased need due to increased lender-side OCR?

University of Minnesota: Lending & Document Delivery

Software choices at MNU

GdPicture
+ tesseract open source OCR



pypdfocr
+ tesseract open source OCR

Search Minitex Requests | Open Request Window | Advanced Search

Mira Client - 3.9.1.0 - vm3020

Workflows

Combine Wiscat Backup Files (Downloads)
Start Verification (Verification)
Start Referrals (Referrals)
Start Typed Queue Review, Start JRET Review, 100 Year Review, Start Message Review (Review)
Start Reports, Soon to Expire (Reporting)
RPA, Campus Loans, Campus Reporting, Batch HCO Fills (Updating)
Como Printing (Printing)

Navigation

Not on Shelf Report

System

- Library Closings
- Locations
- Lender Exceptions
- Campus Branches
- Campus Branch Name Lookups
- Download Holds
- DTU Campus Review Locations
- In Constant Demand
- Minitex Requests
- Minitex Request Locks
- Minitex Request Holds
- Referrals
- New OCLC Updates
- Request Systems
- Users
- VDX Referral Groups
- Soon To Expire Requests
- VDX Format Lookup
- Borrower Note Filters
- Alexis Requests

Scanning

- Scanning
- Documents
- Scanning Users
- Scanning Log
- Scanning Alerts

How do you decide what OCR software works for your library?

ABBYY

- Full featured
- Excellent quality OCR
- Used at U of MD
- Included with Scannx
- Cost

Tesseract OCR

- Works with scanning software
- Needs IT support to configure
- Open source

Adobe Acrobat Pro

- Your institution may already provide access
- High quality OCR
- Works best for low volume OCR needs

Audience Poll:

Is your ILL unit providing OCR?

If yes, how?

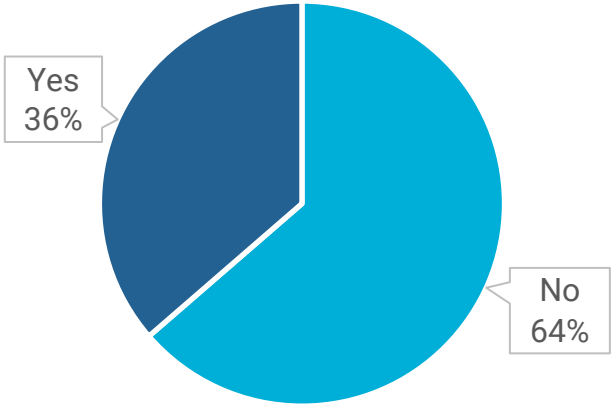
If not, do you plan to do so in 1-2 years?

Image: Raised Hands by Karen Arnold / CC0 Public Domain

BTAA ILL Coordinators Survey

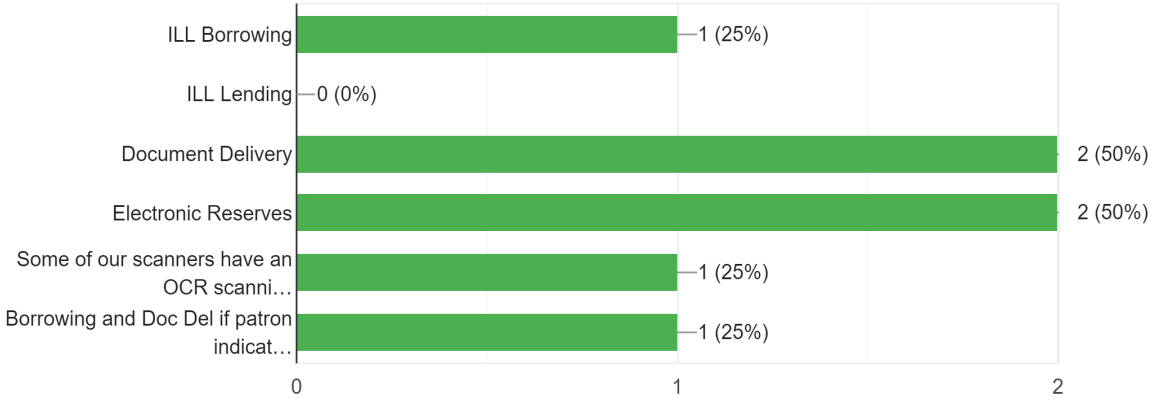
Does your unit or library apply optical character recognition (OCR) to PDF files delivered electronically?

11 responses



For which service(s) does your library apply OCR to PDF files? Please check all that apply.

4 responses



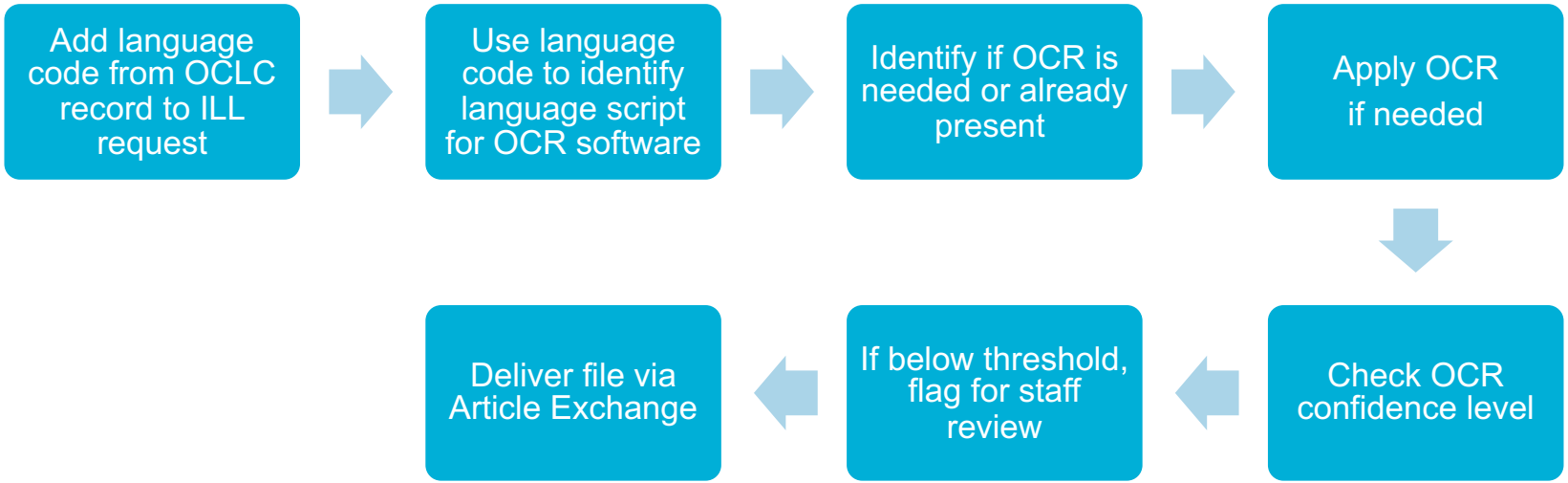
How can we work together to provide OCR more programmatically?

Options include:

1. "OCR required" option + OCR provider group
2. ALA Interlibrary Loan code or consortial agreements
3. Feature of next generation resource sharing systems

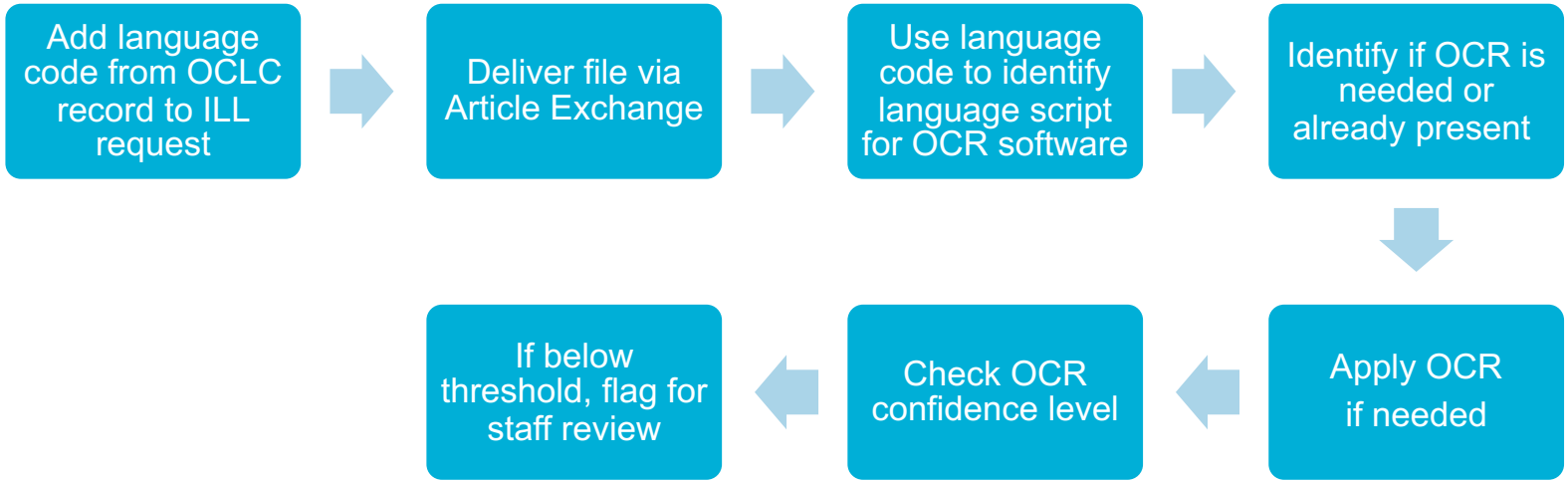
OCR and Next Generation ILL Systems

Workflow A: Supplying Library Participates in Systematic OCR



OCR and Next Generation ILL Systems

Workflow B: Vendor Provides Systematic OCR



Learn More

Atlas Systems: OCR Functionality in ILLiad

UIU: An Introduction to OCR LibGuide

Natural Reader: Text to Speech Application -- try it out!

Convert scanned PDF to OCR using pypdfocr

Check out GdPicture SDK with add-on OCR

RSC19
OCLC RESOURCE SHARING CONFERENCE

thank you

Melissa Eighmy Brown
University of Minnesota Libraries
eighm002@umn.edu

Austin Smith
University of Maryland Libraries
afsmith@umd.edu

Carol Nelson
University of Minnesota Libraries
c-nels@umn.edu

Hilary Thompson
University of Maryland Libraries
hthomps1@umd.edu