

CalSAWS | CalSAWS Imaging Committee – OCR Statistics Update



July 15, 2021

Agenda

- ① Terms
- ② OCR Training Process
- ③ Overlapping Terms
- ④ OCR Index Rates
- ⑤ Next Steps

Terms

OCR

Optical Character Recognition

Brainware

Hyland's proprietary OCR software

Confidence level

Also referenced as Confidence Threshold

Exception rate

How often a document is sent to the Exception Queue

Splitting

The automatic separation done by OCR to a set of documents imaged at once

Index

The naming of a document

A vertical grey line is positioned on the left side of the slide, extending from the top to the middle. At the bottom of this line is a solid green circle.

OCR Training Process

OCR Training Process

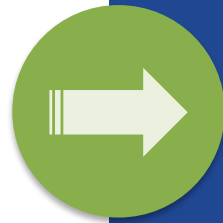
The What

- We are currently only training Person level documents provided by customers
- The initial training and tuning have been completed
- We are working to improve the current index rates by identifying overlap and gathering more sample training documents

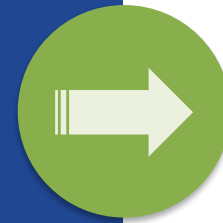
OCR Training Process

The How

For each Imaging Form name that will be OCR'd, a sample set of documents were pulled from C-IV loosely based on the C-IV document types.



Brainware reviews each document. It records words and phrases found on the submitted set of documents along with a location, page number and frequency.



Our Hyland partners review the results to determine which words or phrases should be used in the final configuration.



Overlapping Forms

Overlapping Forms

Things to monitor

Court Records - Divorce Decree - Child Support - Child Custody

- Domestic Violence case that references child custody
- Divorce Decree that determines child and spousal support
- Protective order that temporarily assign child custody

Laboratory Work – Hospital Reports – Immunization Records – Pregnancy Verification – Medical Expense

- Statement with labs
- Doctor bill with immunization charges
- Lab test for pregnancy

Shelter – Utility

- Rental receipt that includes sewer, water, gas and electric

Overlapping Forms

Informational

Receipts – Medical Expense – Educational Expense

- Walmart receipts can be either “Receipt” or “Medical Expense”
- Amazon purchases can be either “Receipt” or “Educational Expense” for a textbook
- Home Depot receipt can be either “Receipt” or something else

Miscellaneous pages that are related but have little information related to the document type

- Fax coversheets
- Pages advertising other products and services (e.g., Bank, Utility, etc.)
- Envelopes
- Receipt stamps and notes on the back of a page

Workers Comp – Third Party Liability

- A large number of Third Party Liability documents are centered around a Workers Comp claim



OCR Index Rates

OCR Index Rates

Form Name Statistics

- **Indexed** is the percent of sample training documents were accurately indexed and archived with the correct form name.
- **Exception** is the percent of sample training documents that could not be recognized and were sent to the Exception Queue.
- **Misindexed** is the percent of sample training documents that were archived with an incorrect form name.
- **Monthly Volume** is the percent of C-IV documents that were indexed to the C-IV equivalent document type. The percentage only represents person level documents and does not account for barcoded documents.

FORM NAME	OCR TRAINED	INDEXED	EXCEPTION	MISINDEXED	MONTHLY	Notes
					VOLUME as % of Total C-IV Docs	
Adoption Documents	N				0.42%	Too generic to train
Assessment	N				0.26%	Too generic to train. Conflicts with OCAT and WEX. In C-IV, this included both C-IV and county created forms.
Attendance Records	Y	72.22%	27.78%	0.00%	0.88%	
Authorized Rep Designation	N				0.00%	Too generic to train. This will most likely be the ID card for an AR.
Bank Verif	Y	90.75%	9.25%	0.00%	5.16%	
Bills	Y	83.33%	16.67%	0.00%	0.15%	
Birth Certificate	Y	73.08%	26.92%	0.00%	4.20%	
Burial Fund/Funeral Statements	Y	83.33%	16.67%	0.00%	0.07%	
Certificate of Completion/Diploma	Y	93.33%	6.67%	0.00%	0.04%	
Child Custody	Y	80.95%	19.05%	0.00%	0.00%	
Child Support	Y	73.68%	26.32%	0.00%	2.49%	
Citizenship Verification	Y	85.00%	10.00%	5.00%	3.56%	
Contracts	N				0.00%	Too generic to train
Court/Hearings Documents	Y	73.68%	21.05%	5.26%	1.38%	
Death Certificate	Y	50.00%	50.00%	0.00%	0.13%	Work on improving indexed rate
Disability (SDI)/Unemployment (UIB)	Y	90.00%	10.00%	0.00%	13.01%	
Disability Verification	N				0.00%	Too generic to train
Divorce Decree	Y	81.25%	18.75%	0.00%	0.02%	
Drivers License	Y	80.00%	20.00%	0.00%	6.92%	
Earnings	Y	79.75%	20.25%	0.00%	24.71%	
Educational Expense	Y	47.37%	52.63%	0.00%	0.06%	Work on improving indexed rate
Employer Verif	Y	75.00%	25.00%	0.00%	3.52%	
FC Application	N				0.00%	Will most likely all be county created forms.
Financial Aid	Y	66.67%	33.33%	0.00%	0.94%	
Foster Childs Data Record AFDC-FC Certif	N				0.00%	
Gifts/Loans	N				0.00%	Too generic to train
Healthcare Insurance	Y	81.82%	18.18%	0.00%	0.00%	
Hospital Reports	Y	76.47%	23.53%	0.00%	0.00%	
Hunt v. Kizer Record of Bills	N				0.00%	Training this would conflict with Medical Expenses, Receipts, etc. It was agreed when this was added that this would not be OCR'd.
Immunization Records	Y	100.00%	0.00%	0.00%	1.75%	
Income Tax/Tax Return Forms	Y	64.29%	35.71%	0.00%	1.18%	Work on improving indexed rate
Institutional Verification	Y	95.24%	0.00%	4.76%	0.00%	
IPV	N				0.00%	Too generic to train
Laboratory Work	Y	61.54%	38.46%	0.00%	0.00%	
Learning Disability	Y	57.14%	28.57%	14.29%	0.33%	
Life Insurance	Y	65.00%	35.00%	0.00%	0.00%	
Long Term Care Facility Verification	N				0.00%	
Lost warrant replacement Affidavit	N				0.00%	Mostly county forms
Marriage/RDP Certificate	Y	68.75%	25.00%	6.25%	0.25%	
Medical Expense	Y	52.63%	47.37%	0.00%	0.81%	
Mortgage Deed	Y	65.00%	35.00%	0.00%	0.05%	
OCAT Appraisal Summary & Recommendation	Y	100.00%	0.00%	0.00%	0.00%	
OP/OI Info	N				0.00%	Too generic to train
Other ID Cards	Y	40.00%	60.00%	0.00%	8.79%	
Passport	Y	63.16%	36.84%	0.00%	1.08%	Work on improving indexed rate
Police Report	Y	65.00%	25.00%	10.00%	0.01%	
Pregnancy Verification	Y	70.59%	29.41%	0.00%	0.55%	
Promissory Note	Y	86.67%	13.33%	0.00%	0.01%	



Next Steps

Project Next Steps and On-Going Work

- OCR training will be on-going and improving for the foreseeable future (at minimum through all county go-lives)
- BA to manually search C-IV for new sample documents
- Any changes or updates will be made in consideration to other form names. Does improving the recognition of one form name negatively impact another?
- [CA-228869](#) was created to request overrides for OCR and OCR splitting