

1-12-2016

## OCR for Digital Collections

Rebecca Bakker

*Florida International University*, [rbakker@fiu.edu](mailto:rbakker@fiu.edu)

Follow this and additional works at: <https://digitalcommons.fiu.edu/glworks>

---

### Recommended Citation

Bakker, Rebecca, "OCR for Digital Collections" (2016). *Works of the FIU Libraries*. 36.  
<https://digitalcommons.fiu.edu/glworks/36>

This work is brought to you for free and open access by the FIU Libraries at FIU Digital Commons. It has been accepted for inclusion in Works of the FIU Libraries by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).

# OCR FOR DIGITAL COLLECTIONS



*Rebecca Bakker*  
*Digital Collections Center*  
*Florida International University*  
*rbakker@fiu.edu*

# OCR - What does it mean and why is it important?



- Optical Character Recognition
- a/k/a “Text Recognition”
- OCR accuracy affects information retrieval
- Important step in Digitization

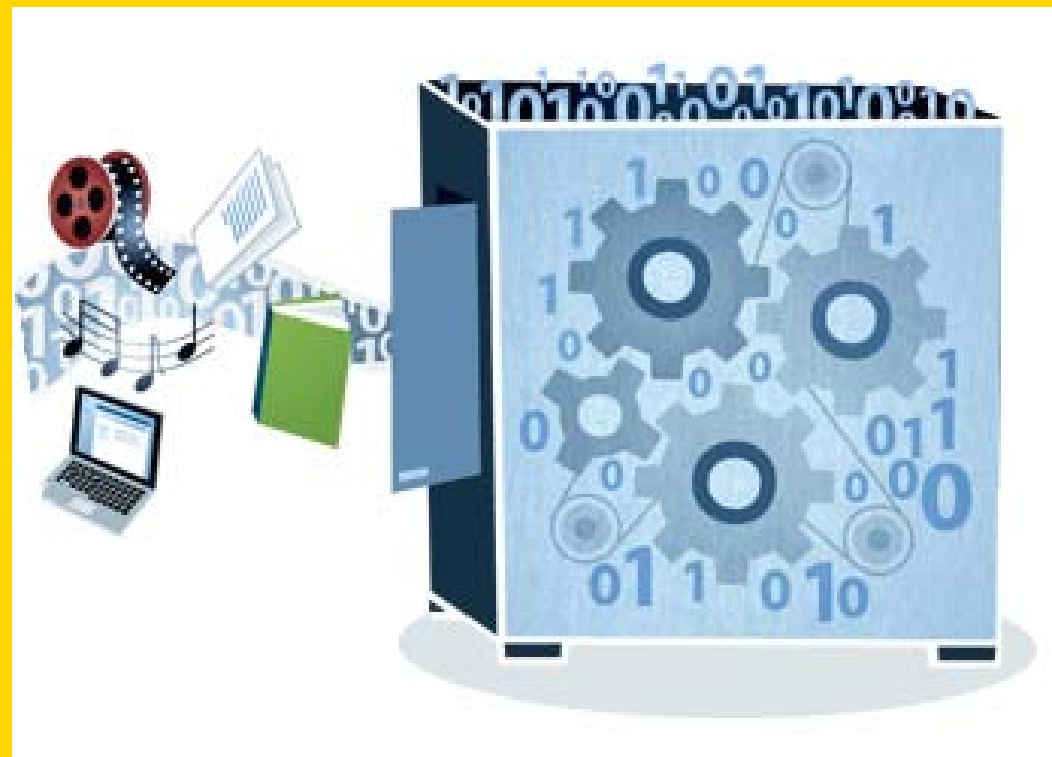
# Purpose of this Presentation

- Best practices
- Common problems and considerations
- Software available for your project

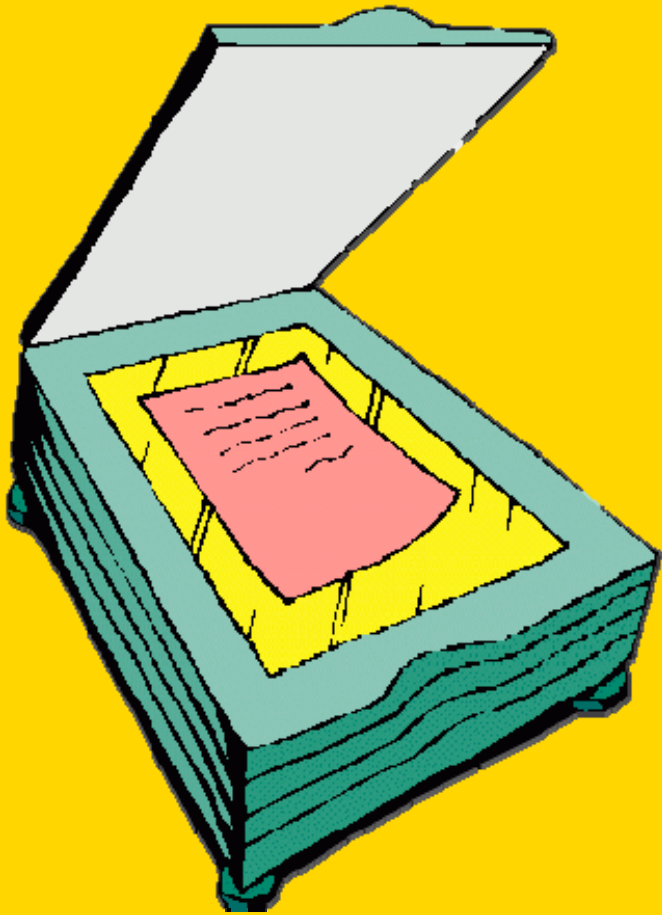


## Things to consider for your project:

- Quality of original
- Volume of files
- Preservation Intentions



# Best Practices for OCR



- Text:
  - *High contrast*
  - *Font size over 6 points*
  - *Published after 1950*
- Scanned:
  - *RGB Mode*
  - *300 dpi*
  - *Relatively straight!*
- Image format TIFF (Tagged Image File Format)
- OCR done as part of the image to PDF conversion
- Text-under-image in PDF

# Common difficulties

- Similar symbols (O's and 0's, 1's and l's, etc.)
- Punctuation and special symbols ☺
- Typography – *italics*, spacing, underlining
- Complex character scripts (non-Latin languages)
- Proper names
- Columns
- Bleed-through

340

**HERMAN J. ARRANT, Inc.**  
 "NOTHING BUT INSURANCE"  
 Huntington Bldg. Phone 3-5566

**POLK'S**  
 GREATER  
**MIAMI**  
 Directory of Householders, Occupants of Office Buildings and  
 Other Business Places, Including a Complete  
 Street and Avenue Guide  
**1939**  
 Including Coral Gables and Miami Beach  
 Copyright, 1939, by R. L. Polk & Co.

In the following list the named streets are arranged in alphabetical order, followed by the numbered streets in numerical order, the intersecting streets being shown. The number of each building on the street is given in numerical order, and opposite the number is placed the name of the occupants or houseowner or tenant. In the business blocks containing offices the occupants are shown by room numbers where practicable, in apartment houses the number denotes the number of the apartment. In all streets both sides of the street are included in one list. The profession or business of individuals or firms is given only at their respective place of business; to find occupation of others refer to Alphabetical List of Names.

The symbol Ⓢ following a householder's name indicates that we have received information during the canvass that the house is owned by some member of the family, but as the publisher cannot and does not guarantee the correctness of the information furnished, nor the complete absence of mistakes no responsibility for errors can be or is assumed, nor can the publisher furnish further information than that shown.

(c) denotes colored.

<p><b>ABACO AV (C Grove)</b>—From L. 14th southwest to Washington, 2 south of SW 28th</p> <p>2482 Rodgers East Ⓢ Ⓢ          2482 DeLuca Dentist          2497 Thompson Jas Ⓢ Ⓢ</p> <p style="text-align: center;">Intersecting streets</p> <p>2580 Love Doran E Ⓢ          2512 Smith Theo F Ⓢ          2519 Dowdell Tom J          2528 Torgolo Wm V Ⓢ          2541 Turk Wm J Ⓢ          2544 Mathews Wesley W Ⓢ</p> <p style="text-align: center;">Washington intersects</p> <p>2641 Beckett Linda Ⓢ Ⓢ</p> <p><b>ABBOTT AV (MB)</b>—From Atlantic dr north to 10th st, 2 west of Colby av</p> <p>6785 Caribbean Court Apts          Apartments:          1 White Jeanette Mrs          2 Sell Julia Ⓢ Mgr          Herc Aug Ⓢ          3 Dickinson Jessie Mrs          4-6-7-8 Vacant</p> <p>Street continued</p> <p>6785 Ray Breeze Apts          Apartments:          1 Vacant          2 Vacant</p>	<p>3 Bell Wm H Ⓢ          4-6 Vacant          6819-24-26-28-32-34 Biarritz Villas Apts          6816 Rank Arthur Ⓢ          6815 Abbott Apartments          Apartments:          1 Vacant          2 Young Chas Ⓢ          3-4-5 Vacant</p> <p>Street continued</p> <p>6815 Abbott Court Apartments          Apartments:          1-2-3 Vacant          4 Beasley Chas E          5-6-7-8 Vacant</p> <p>Street continued</p> <p>6814 Vacant          6814 Vacant          6822 Sharrigan Geo H          6823 Huntington Lewis          6824 Vacant</p> <p>6839 Cass Blanca Apartments          Apartments:          1 Rowe Emma Mrs Ⓢ          2 Vacant          3 Vacant          4 Columbia Ber L          5 Vacant</p>	<p>6 Ivan Harold A          7-8 Vacant          6850 Vacant          6854 south Geo E Ⓢ          6863 Wilking Laura Mrs Ⓢ</p> <p style="text-align: center;">68th st intersects</p> <p>6925 Ashley Harry L          6949 Douglas Carl T Ⓢ          6961 Brookwood Apartments          4706 Chas G mgr          6982 Hinkley Mason M Mrs Ⓢ          6985-86 Grasswood Apartments          Decuman Geo F Mrs mgr          7146 Whitmarsh Apartments The Whitmarsh Richmond D mgr          7147          7148          7149          7150          7151          7152          7153          7154          7155          7156          7157          7158          7159          7160          7161          7162          7163          7164          7165          7166          7167          7168          7169          7170          7171          7172          7173          7174          7175          7176          7177          7178          7179          7180          7181          7182          7183          7184          7185          7186          7187          7188          7189          7190          7191          7192          7193          7194          7195          7196          7197          7198          7199          7200</p> <p style="text-align: center;">71st st intersects</p> <p>7611 Nelson Harry Ⓢ          7629 Whites Apartments          7651 Goodwin Wm D Ⓢ          7628 Vacant</p> <p style="text-align: center;">77th st intersects</p> <p>7785 Vacant          7738 Landis Earl M Ⓢ          7800          7801          7802          7803          7804          7805          7806          7807          7808          7809          7810          7811          7812          7813          7814          7815          7816          7817          7818          7819          7820          7821          7822          7823          7824          7825          7826          7827          7828          7829          7830          7831          7832          7833          7834          7835          7836          7837          7838          7839          7840          7841          7842          7843          7844          7845          7846          7847          7848          7849          7850          7851          7852          7853          7854          7855          7856          7857          7858          7859          7860          7861          7862          7863          7864          7865          7866          7867          7868          7869          7870          7871          7872          7873          7874          7875          7876          7877          7878          7879          7880          7881          7882          7883          7884          7885          7886          7887          7888          7889          7890          7891          7892          7893          7894          7895          7896          7897          7898          7899          7900</p> <p style="text-align: center;">78th st intersects</p> <p>7886 Elgin Alf L          7838 Thor Apartments          7838          7839          7840          7841          7842          7843          7844          7845          7846          7847          7848          7849          7850          7851          7852          7853          7854          7855          7856          7857          7858          7859          7860          7861          7862          7863          7864          7865          7866          7867          7868          7869          7870          7871          7872          7873          7874          7875          7876          7877          7878          7879          7880          7881          7882          7883          7884          7885          7886          7887          7888          7889          7890          7891          7892          7893          7894          7895          7896          7897          7898          7899          7900</p> <p style="text-align: center;">78th st intersects</p>
---	---	--

**WASHINGTON STORAGE CO., Inc.**  
 Agents ALLIED VAN LINES Inc.  
 Moving — Packing — Shipping — Rug Cleaning — Baggage Transfer  
 Storage  
 1001 Washington Av. (M. B.) PHONE 5-3448

**PROTECT-O**  
 AWNING-SHUTTER CO.

**REED**  
 Construction  
**CORP.**  
 Engineers  
 and  
 Contractors

—

Bulkheading  
 Decks  
 Foundations  
 Wood Piles  
 Concrete Piles  
 Steel Sheet  
 Piling  
 Guniting  
 Bridges.

—

1345  
 20th St.  
 Miami Beach,  
 Fla.

Phones  
 5-5397 and  
 5-5398

P. O. Box  
 677

## Original scanned document:

### 2, 4A. TRANSACTIONS BY KEY

- A. This program will generate a list of all posting transactions by one key type.
- B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

- C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

(2,4A-2)

RESTAURANT TRANSACTIONS		1	31	1976
INTERNATIONAL INN TAMiami		TRAIL		
224	0 BARRY	15.00	69.08	1 39
303	0 JAMES	7.50	47.02	1 41
303	0 JAMES	29.00	84.78	1 43
201	0 DUCK	36.31	42.35	3 43
112	0 ROLLS	37.27	37.27	3 45



## Image processed document:

### 2, 4A. TRANSACTIONS BY KEY

- A. This program will generate a list of all posting transactions by one key type.
- B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

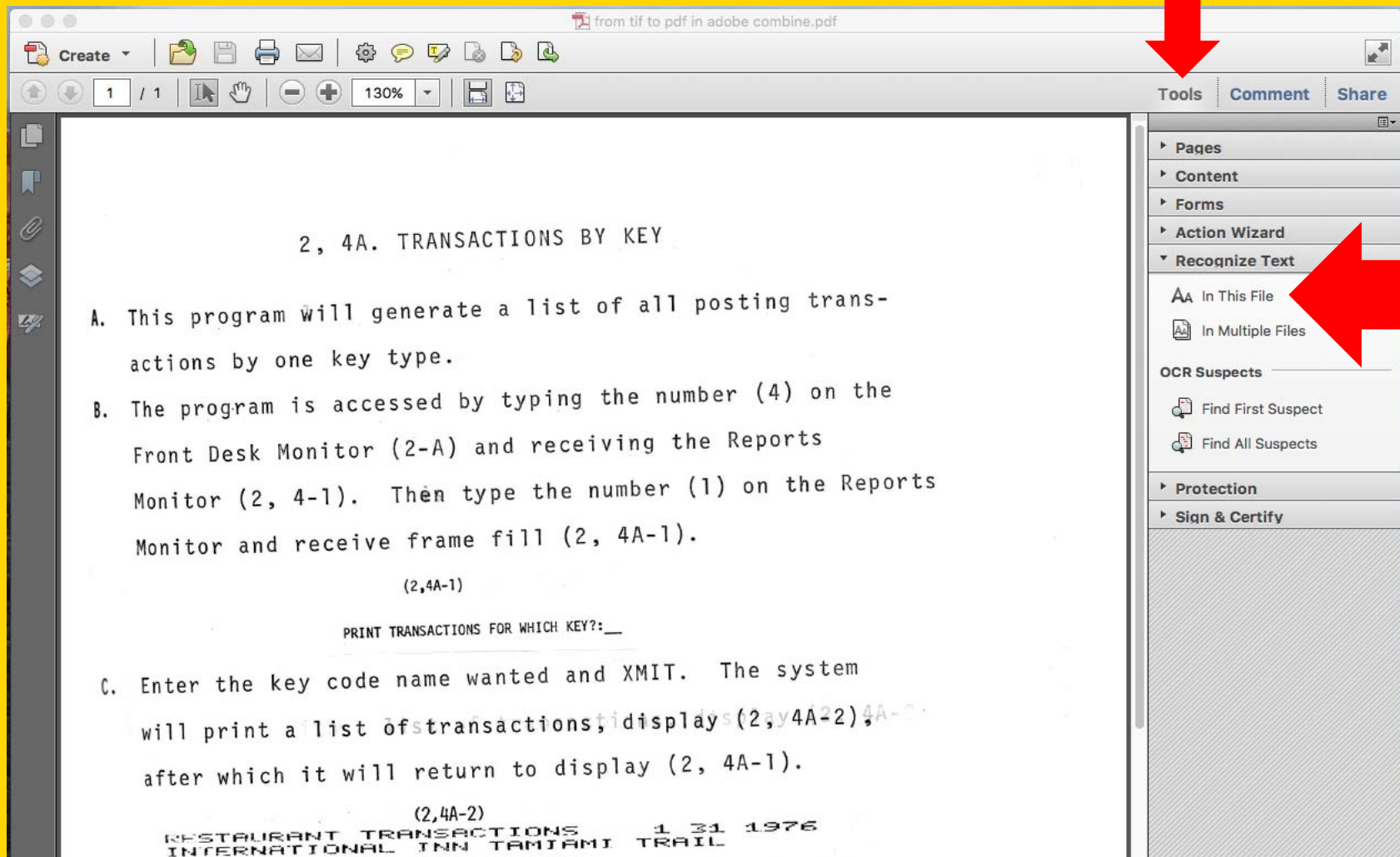
PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

- C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

(2,4A-2)

RESTAURANT TRANSACTIONS		1	31	1976
INTERNATIONAL INN TAMJAMI TRAIL				
224	Ø BARRY	15.00	69.08	1 39
303	Ø JAMES	7.50	47.02	1 41
		29.00	84.78	1 43

# OCR in Adobe Acrobat Pro



The screenshot shows the Adobe Acrobat Pro interface. The top toolbar includes icons for Create, Save, Print, Email, Settings, Comment, and Share. The main document area displays a page with the following text:

2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

(2,4A-2)

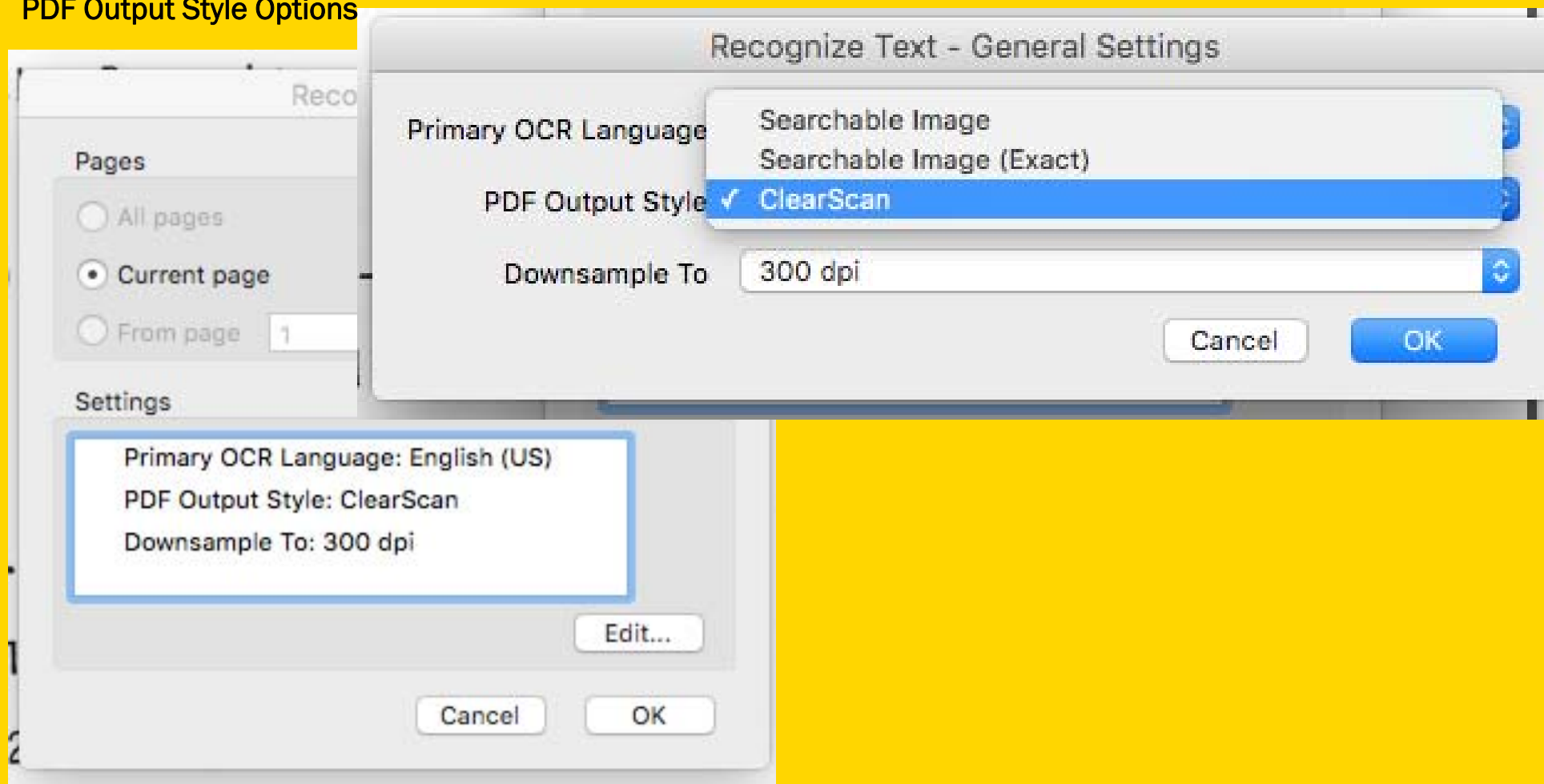
RESTAURANT TRANSACTIONS 1 31 1976  
INTERNATIONAL INN TAMiami TRAIL

The right-hand sidebar contains the Tools panel, which is expanded to show the 'Recognize Text' section. Two red arrows point to the 'Recognize Text' section and the 'In This File' option. The 'In This File' option is selected, and the 'OCR Suspects' section is visible below it, containing 'Find First Suspect' and 'Find All Suspects' options.

[Video link](#)

# OCR in Adobe Acrobat –

## PDF Output Style Options



Searchable Exact

This program will

Searchable




This program will

Clearscan

This program will

# Differences in Acrobat OCR Output Styles

	Deskews	File Size	Hidden Text Layer	Allows corrections	PDF/A Compliant
Searchable Image (Exact)	NO	Largest	YES	YES	YES
Searchable Image	YES	Middle	YES	YES	YES
Clearscan	YES	Smallest	NO	NO	NO

Name	Size ▾
 adobe searchable image exact ocr.pdf	287 KB
 adobe searchable image ocr.pdf	254 KB
 adobe clearscan ocr.pdf	135 KB

# Adobe Acrobat OCR Accuracy

2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

PRINT TRANSACTIONS FOR WHICH KEY?: \_\_

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

# Adobe Acrobat OCR

2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

C. Enter the key code name wanted and XMIT. The system will print a list of transactions; it displays (2, 4A) after which it will return to display (2, 4A-1).

(2,4A-2)

RESTAURANT TRANSACTIONS		1 31 1976
INTERNATIONAL INN TAMiami TRAIL		
224	0 BARRY	15.00 69.08
303	0 JAMES	7.50 47.02
304	0 JAMES	29.00 84.78
201	0 DUCK	36.31 42.35
117	0 ROLLS	37.27 37.27
		4302.00 4302.00
0	57 SALES	15.00 15.00

Find Element

Find: OCR Suspects

To correct an OCR suspect, click on the highlighted object in the document and type in the new text. After making the correction, click "Accept and Find" to identify the next OCR suspect.

TRANSACTIONS

Search Page  Search Document

Not Text

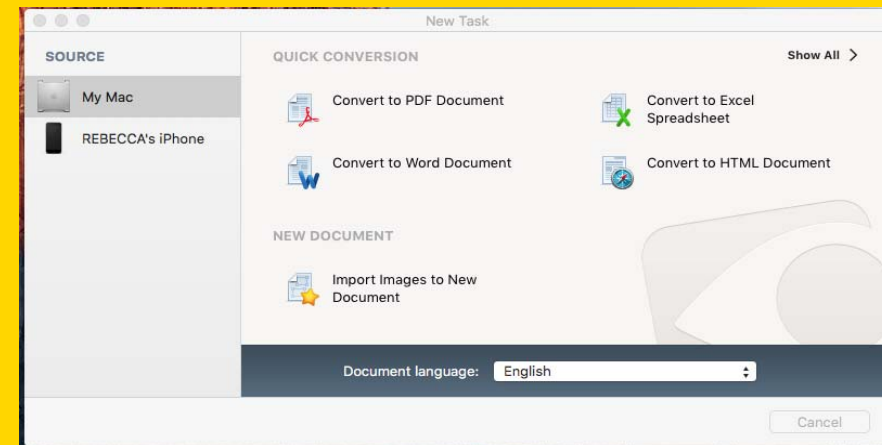
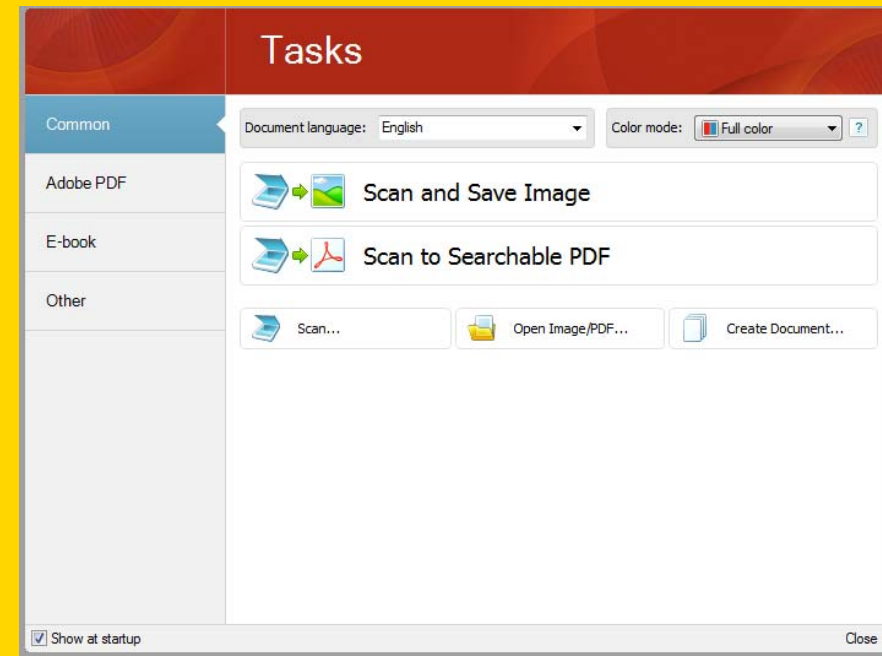
Find Next Accept and Find Close

[Video link](#)

# ABBYY FineReader

Recognizes 190 languages, in all combinations:

Abkhaz, Adyghe, Afrikaans, Agul, Albanian, Altaic, Arabic (Saudi Arabia) +, Armenian (Eastern) +, Armenian (Grabar) +, Armenian (Western) +, Avar, Aymara, Azeri (Cyrillic), Azeri (Latin) +, Bashkir +, Basic, Basque, Belarusian, Bemba, Blackfoot, Breton, Bugotu, Bulgarian +, Buryat, C/C++, Catalan +, Cebuano, Chamorro, Chechen, Chinese Simplified, Chinese Traditional, Chukchee, Chuvash, COBOL, Corsican, Crimean Tatar, Croatian +, Crow, Czech +, Dakota, Danish +, Dargwa, Dungan, Dutch (Belgian) +, Dutch +, English +, Eskimo (Cyrillic), Eskimo (Latin), Esperanto, Estonian +, Even, Evenki, Faroese, Fijian, Finnish +, Fortran, French +, Frisian, Friulian, Gagauz, Galician, Ganda, German (Luxembourg), German (new spelling) , German +, Greek +, Guarani, Hani, Hausa, Hawaiian, Hebrew +, Hungarian +, Icelandic, Ido, Indonesian +, Ingush, Interlingua, Irish, Italian +, Japanese +, Java, Jingpo, Kabardian, Kalmyk, Karachay-Balkar, Karakalpak, Kasub, Kawa, Kazakh, Khakass, Khanty, Kikuyu, Kirghiz, Kongo, Korean (Hangul) +, Korean +, Koryak, Kpelle, Kumyk, Kurdish, Lak, Latin +, Latvian +, Lezgi, Lithuanian +, Luba, Macedonian, Malagasy, Malay, Malinke, Maltese, Mansi, Maori, Mari, Maya, Miao, Minangkabau, Mohawk, Moldavian, Mongol, Mordvin, Nahuatl, Nenets, Nivkh, Nogay, Norwegian (Bokmal) +, Norwegian (Nynorsk) +, Nyanja, Occidental, Occitan, Ojibway, Ossetian, Papiamento, Pascal, Polish +, Portuguese (Brazil) +, Portuguese +, Quechua, Rhaeto-Romance, Romanian +, Romany, Rundi, Russian (old spelling) +, Russian +, Russian with accents <sup>new</sup> +, Rwanda, Sami (Lappish), Samoan, Scottish Gaelic, Selkup, Serbian (Cyrillic, Latin), Shona, Simple chemical formulas, Slovak +, Slovenian +, Somali, Sorbian, Sotho, Spanish +, Sunda, Swahili, Swazi, Swedish +, Tabasaran, Tagalog, Tahitian, Tajik, Tatar +, Thai +, Tok Pisin, Tongan, Tswana, Tun, Turkish +, Turkmen (Cyrillic), Turkmen (Latin), Tuvian, Udmurt, Uighur (Cyrillic, Latin), Ukrainian +, Uzbek (Cyrillic, Latin), Vietnamese +, Welsh, Wolof, Xhosa, Yakut, Yiddish, Zapotec, Zulu





# ABBYY FineReader OCR

Untitled document [1] - ABBYY FineReader 11 Professional Edition

File Edit View Document Page Areas Tools Help

Document language: English

Exact Copy

Redo Undo

Pages

Full color

Image

2, 4A. TRANSACTIONS BY KEY

This program will generate a list of all posting transactions by one key type.

The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

Enter the key code name wanted and XMIT. The system will print a list of transactions; it displays (2,y4A-2); 4A-2 after which it will return to display (2, 4A-1).

(2,4A-2)

RESTAURANT TRANSACTIONS 1 31 1976				
INTERNATIONAL INN TAMAMI TRAIL				
224	0	BARRY	15.00	69.08 1.39
303	0	JAMES	7.50	47.02 1.41
303	0	JAMES	29.00	84.78 1.43

Text

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2,y4A-2) after which it will return to display (2, 4A-1).

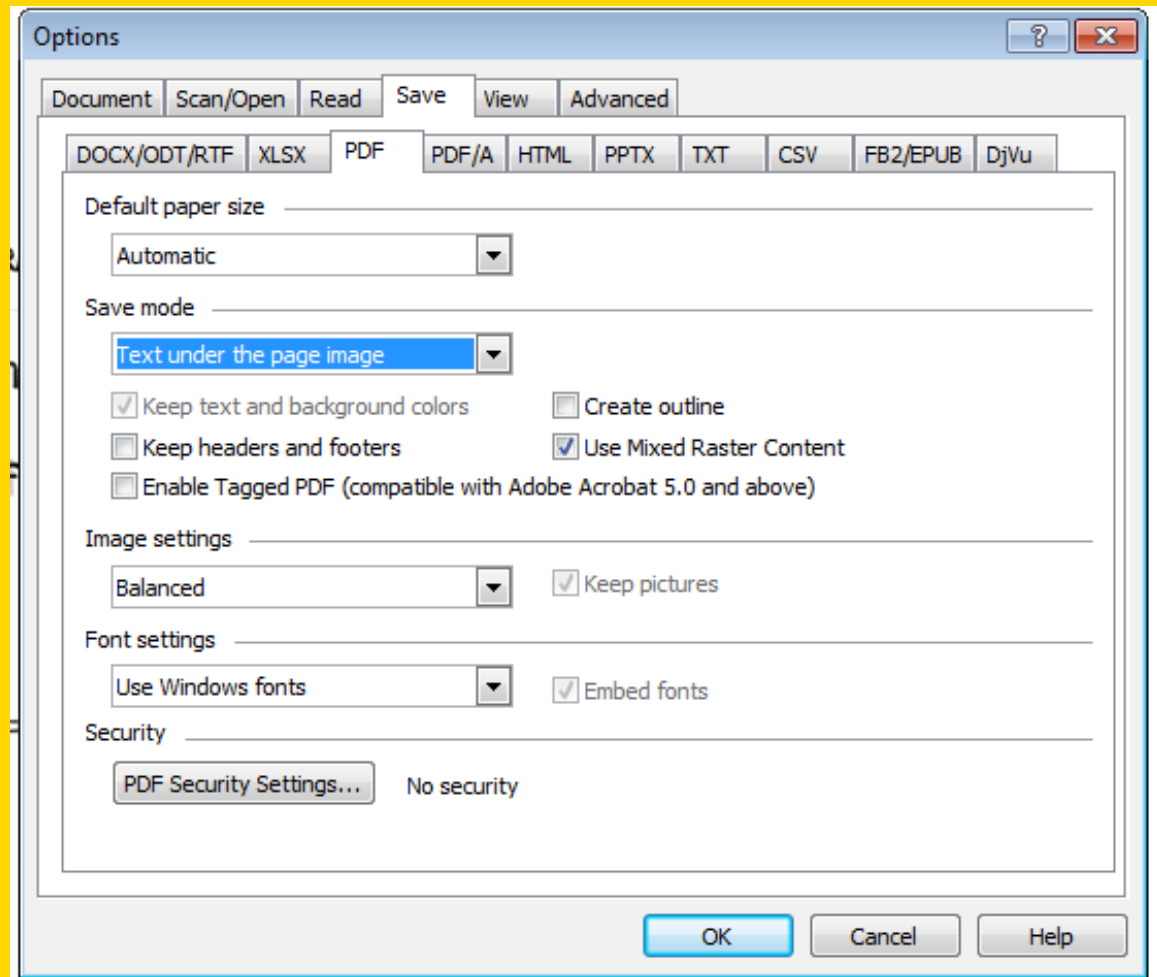
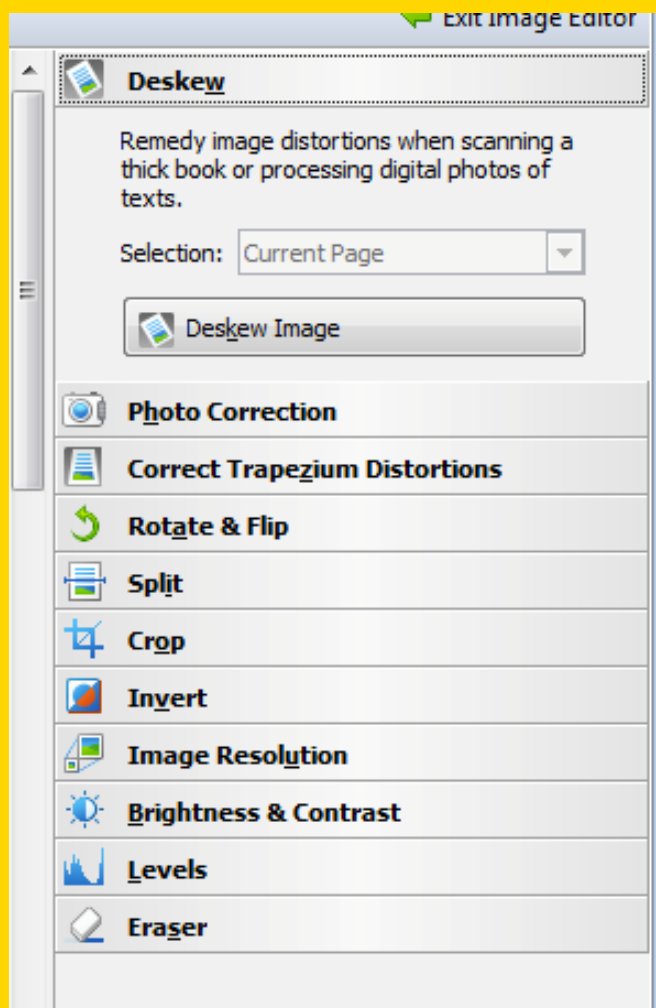
(UM-2)

RESTAL!!	ARWT IR	F-4 S	ent!	O T I.	O M S	DL	jSdl	1'37
X NTERNA	EDC T ONRL	T NE-4	TRMTRM.	IRR			I I	
224	0	BARRY				15.00	69.08	1.39
303	0	JAMES				7.50	47.02	1.41
303	0	JAMES				29.00	84.78	1.43
201	0	DUCK				36.31	42.35	3.43
111	0	ROLLS				37.27	37.27	3.45
0	57	SRILES				4302.00	4302.00	4.9

Click here to view zoomed image (Ctrl+F5).

Page 1: No warnings

# ABBYY FineReader OCR



# ABBYY FineReader OCR

## 2, 4 A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).  
(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

C. Enter the key code name wanted and XMIT. The system will print a list of stransacti o n s, display (2,y4A-2) after which it will return to display (2, 4A-1).

## 2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).  
(2,4A-1)

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:\_\_

C. Enter the key code name wanted and XMIT. The system will print a list of stransactions, display (2,y4A-2), after which it will return to display (2, 4A-1).

# Prime OCR

- 11 languages
- Uses up to 6 OCR Engines at a time
- Allows for high volume batch processing
- Vastly improved accuracy

The screenshot shows the PrimeOCR Job Server interface. The window title is "PrimeOCR Job Server". The menu bar includes "File", "Logs", "Configuration", and "Help". There are buttons for "Wizard...", "Setup...", and "Log". The status bar indicates "Status Processing Job".

On the right side, there is a text area showing job details: "PrimeOCR Reading C:\Prdev\Images\00001812.TIF Page 1 IMAGE SIZE=1632,2544 RES=300 Intermediate Results(% Complete)82".

The main area contains several summary tables:

# Images Processed - All Jobs			
	Total	Complete	Remaining
# Jobs	5	0	

Job C:\Prdev\Job\Copy (2) of Job1.job			
	Total	Complete	Remaining
# Files	1371	19	1352
Time(h:m)	0:50	0:00	0:49
Finish Time	Tue May 15 17:40:04		

Confidence	# of Chars	Time(sec.)	Average	
850	2924	2.19		
882	3412	1.86	00001807.TIF	C:\Prdev\Images\
853	3819	3.91	00001808.TIF	C:\Prdev\Images\
886	3192	1.41	00001811.TIF	C:\Prdev\Images\

At the bottom, there is a table titled "Jobs Processed" with columns: Job#, Completed, Files, Errors, #Pages, Time(h:m), and Job Name. The table is currently empty.

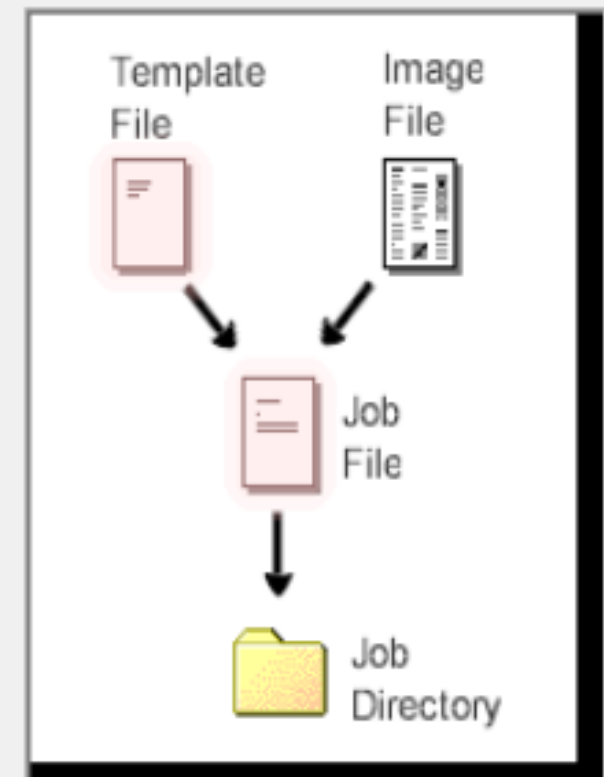
# Prime OCR

**Template File:** OCR engine configuration data, i.e. image enhancement steps, automatic zoning settings, output format.

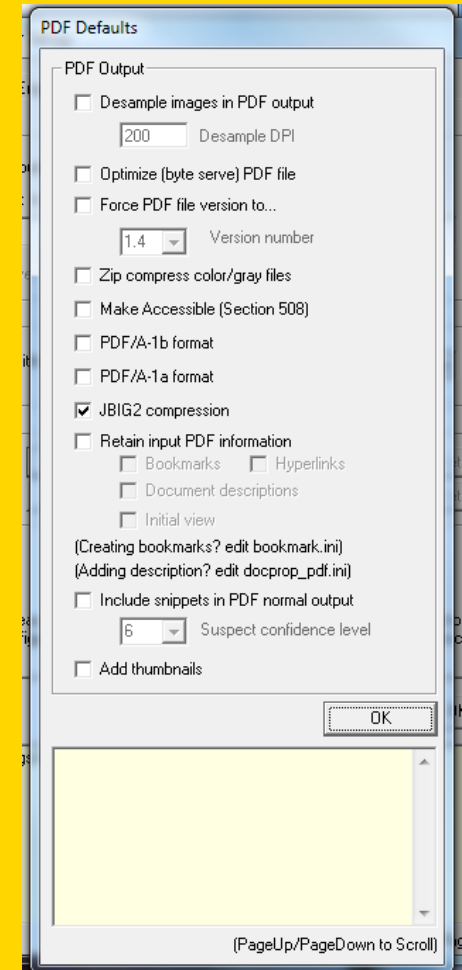
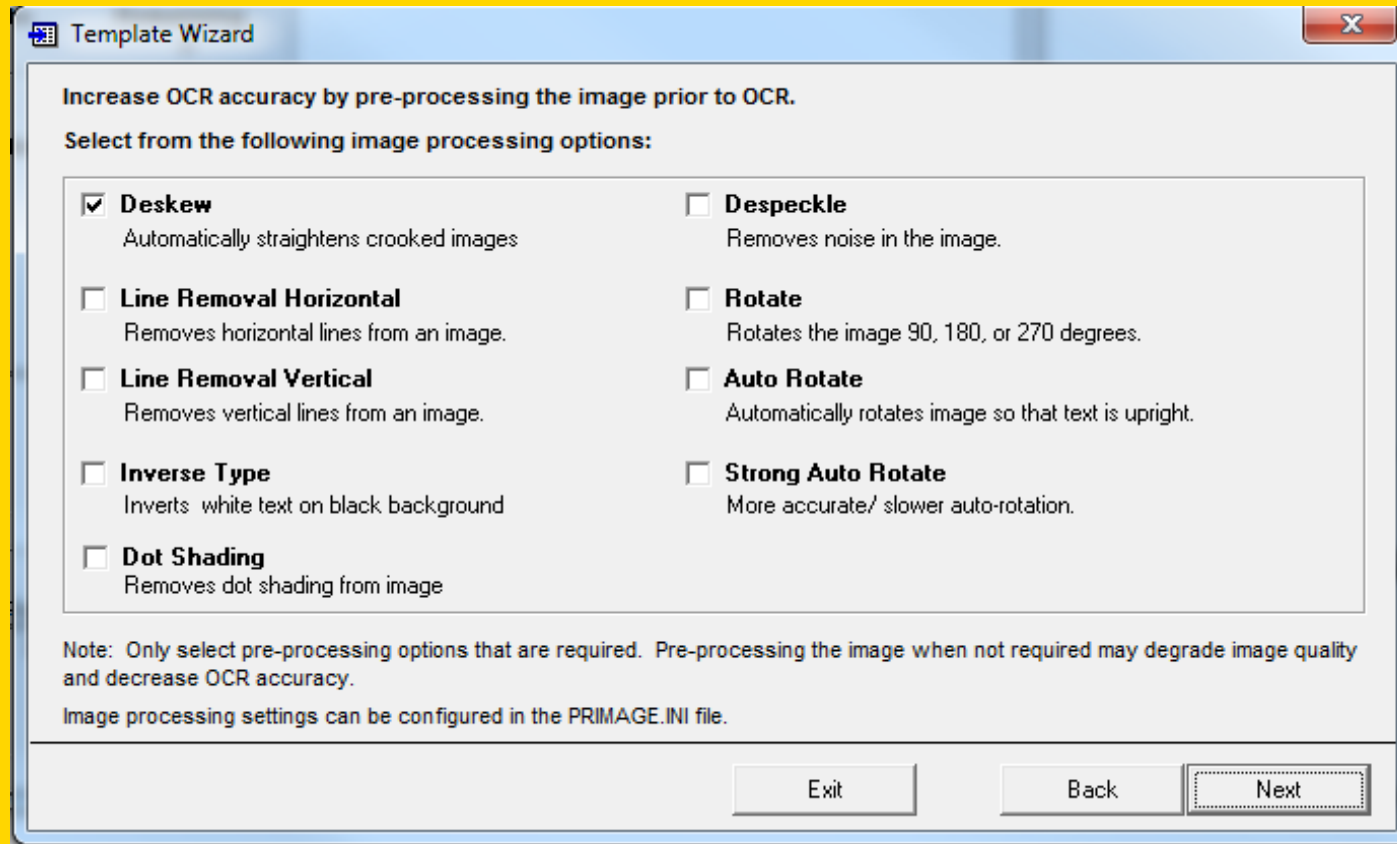
**Image File:** Any TIFF/JPEG/PDF/ etc file that can be accessed by the PC via a file path (can be on a remote server).

**Job File:** Describes a path to the images to be processed and a path to the template file to be used for processing.

**Job Directory:** The PrimeOCR Job Server will search this directory for job files and will process them. Location of the job directory can be changed in Setup/Input.



# Prime OCR



# Prime OCR Accuracy

## 2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

PRINT TRANSACTIONS FOR WHICH KEY?:

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

## 2, 4A. TRANSACTIONS BY KEY

A. This program will generate a list of all posting transactions by one key type.

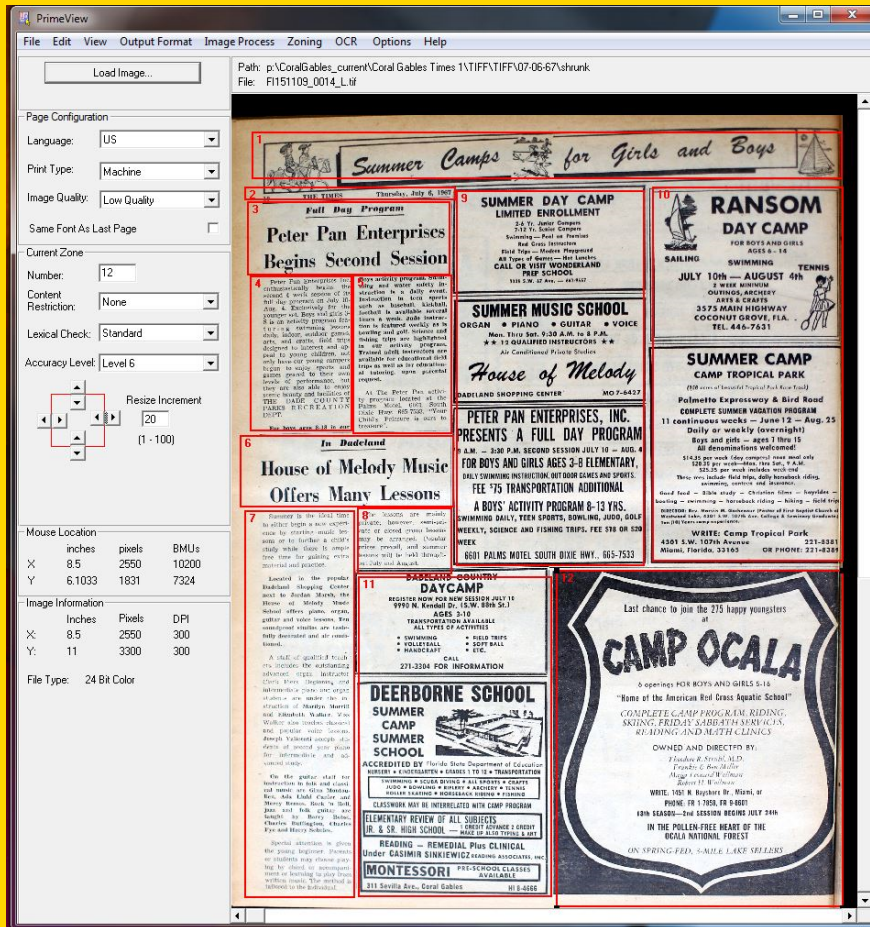
B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

(2,4A-1)

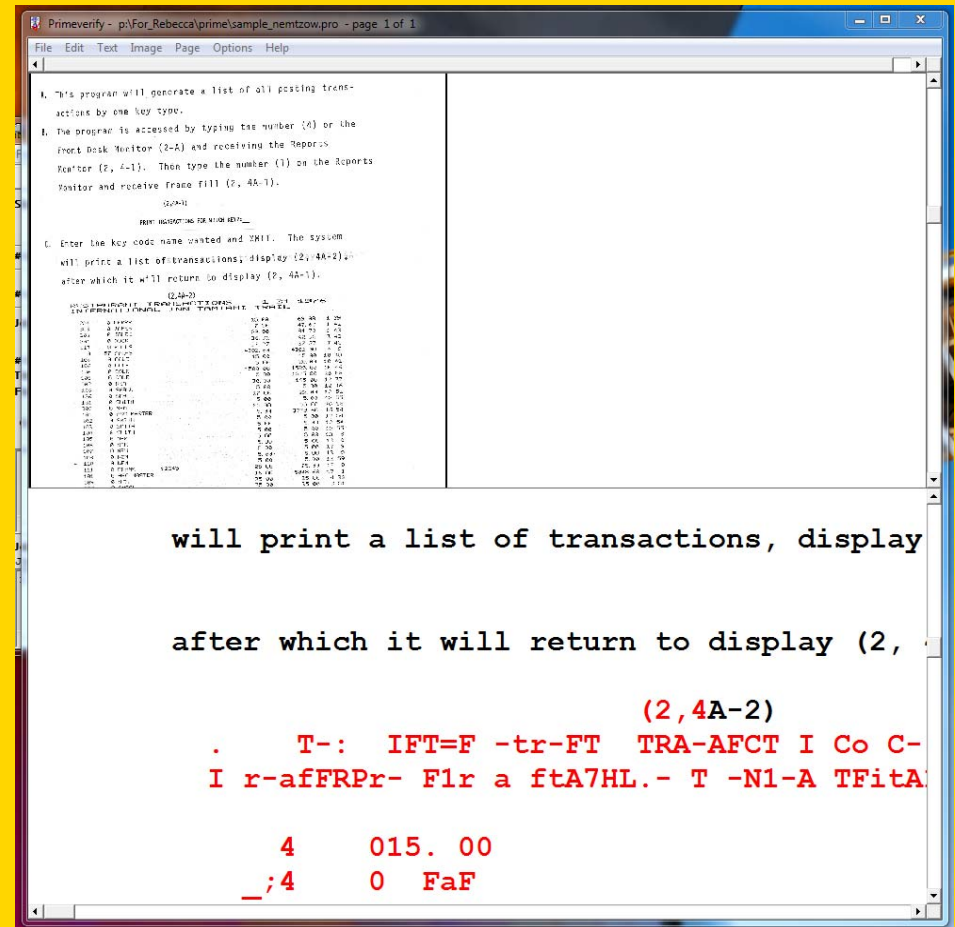
PRINT TRANSACTIONS FOR WHICH KEY?:

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1).

# PrimeView










# PrimeVerify





# File size comparison

Name	Size ^	Kind
 abbyy_finereader11.pdf	30 KB	Adobe...cument
 primeOCR-compressed.pdf	68 KB	Adobe...cument
 adobe clearscan ocr.pdf	135 KB	Adobe...cument
 abbyy mac finereader.pdf	230 KB	Adobe...cument
 adobe searchable image ocr.pdf	254 KB	Adobe...cument
 adobe searchable image exact ocr.pdf	287 KB	Adobe...cument
 primeOCR.pdf	316 KB	Adobe...cument

# IRIS

The screenshot displays the IRIS OCR software interface. At the top, the user is logged in as 'Limb - Rolando'. The main window shows a document titled 'FW Lovejoy\_000018'. The interface is divided into several sections:

- Top Bar:** Includes 'Dashboard', 'New Project', and 'Log Out' buttons. The 'LIMB' logo is prominently displayed.
- Navigation and Controls:** A toolbar with 'Current', 'Finish', and 'OCR - Lovejoy' tabs. Below this, there are options for 'Sort', 'Percent Accuracy' (set to 90.00), and 'Display Mode'.
- Document View:** The central area shows a scanned page of text. The text is as follows:

ness, but modest and unassuming. Studious and thoroughly reliable, he was liked and trusted by our group."  
In 1890, Frank Lovejoy went on to the Massachusetts Institute of Technology, where he was graduated in 1894 with the degree of B.S. in chemical engineering. Throughout his four years at M. I. T., he was a student of outstanding ability. He took full part in undergraduate activities but, characteristically, never sought prominence in them. "My pleasantest memory is of the intercourse with class and course mates and the friendships formed among Institute men," he wrote in M. I. T.'s "Class of 1894 Decennial Catalogue." "My most painful is of the hours spent in profitless 'grinding' that might have been spent in profitable pleasure."  
Following his graduation from M. I. T., the young chemical engineer went to work in a sugar refinery at West Baton Rouge, Louisiana. He had been at the refinery only six months, "doing the routine work of the sugar chemist, which on a sugar plantation begins before sunrise and not infrequently ends long after sunset," when he accepted an offer of employment with Curtis Davis & Company, soap manufacturers of Cambridgeport, Massachusetts. This company, incidentally, later became the nucleus of the present Lever Brothers plant. Frank Lovejoy had spent two years with the soap company, the first as draftsman and the second as chemist, when he was offered a position with the Eastman
- Thumbnail Grid:** On the left, a grid of thumbnails shows previous pages of the document, labeled 'FW Lovejoy\_000012' through 'FW Lovejoy\_000017'. Each thumbnail includes a small image of the page and a corresponding text snippet.
- Text and Statistics Panel:** On the right, a panel displays 'Text' and 'OCR Statistics' for the current page.

# OCR Freebies

- Tesseract
- Google Docs
- Evernote
- Web based

## 2, 4A. TRANSACTIONS BY KEY

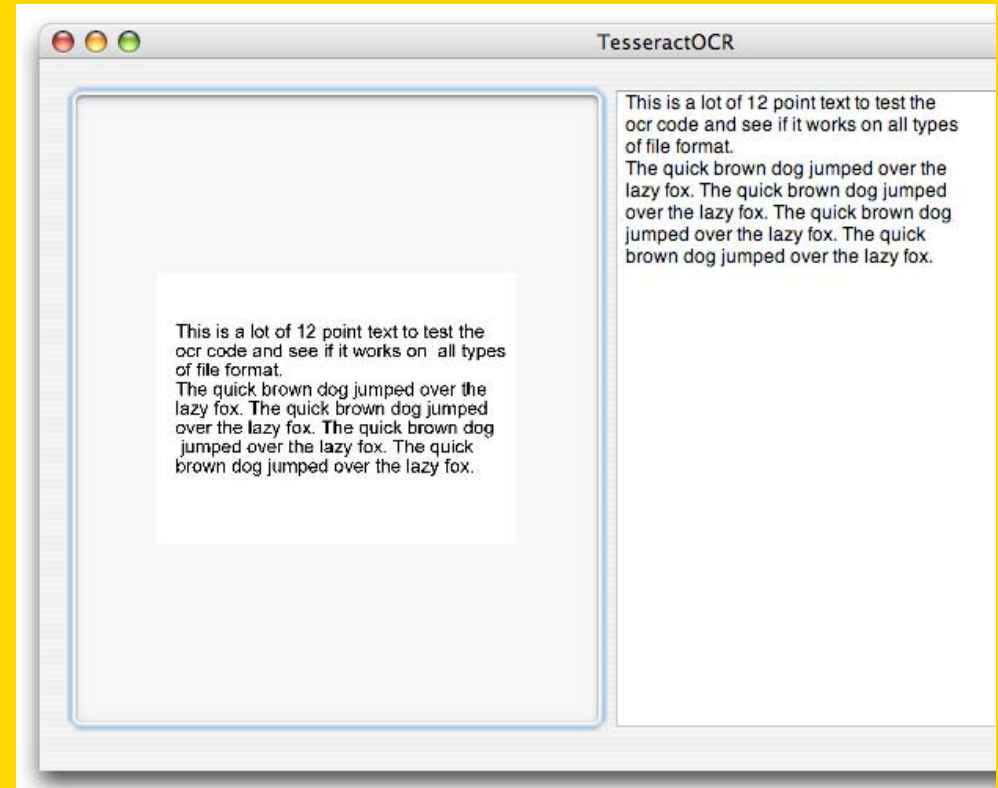
L. This program will generate a list of a) posting transactions by one key type.

B. The program is accessed by typing the number (4) on the Front Desk Monitor (Z-A) and receiving the Reports Monitor (2, 4-1). Then type the number (T) on the Reports Monitor and receive frame

ummn

PRINT TRANSACTIONS FOR WHICH KEY?:

C. Enter the key code name wanted and XMIT. The system will print a list of transactions, display (2, 4A-2), after which it will return to display (2, 4A-1). (2,4A-2)



# Accuracy – calculated at character level

2, 4A. TRANSACTIONS BY KEY

- A. This program will generate a list of all posting transactions by one key type.
- B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

2, 4A. TRANSACTIONS BY KEY

- A. This program will generate a list of all posting transactions by one key type.
- B. The program is accessed by typing the number (4) on the Front Desk Monitor (2-A) and receiving the Reports Monitor (2, 4-1). Then type the number (1) on the Reports Monitor and receive frame fill (2, 4A-1).

- Accuracy varies depending on image and text quality
- 99% accuracy = 1 out of 100 characters are recognized as uncertain
- Industry standard for OCR: 98% post-1950's, 95% 1900-1950, 85% pre-1900
- Character accuracy does not tell word accuracy or the usefulness of the text output

# Accuracy

## 90% Accuracy

- An average newspaper page has 1,000 words and 5,000 characters
- 90% accuracy = 500 incorrect characters
- Word accuracy would likely be between 500 - 900 correct words
- Depending on the number of "significant words" rendered correctly, the search results could vary widely

**May Be Abolished**

## Free Golf, City Cars Draw Commission Fire

**By PAT MURPHY**  
 Coral Gables city commissioners may swing their axes and end the controversial widespread use of free golf passes, and staff cars for certain city hall officials.

Such practices came under severe self-criticism from commissioners who considered them as not only troublesome but costly to taxpayers.

More than 20 city cars are

assigned to city hall for use by certain officials, some of whom use the vehicles on a round-the-clock basis.

As for free golf passes, the commission last year approved 140 dollars for those who, ex-copers, ministers, University of Miami and Veterans Administration personnel and directors of the Coral Gables Country Club, among others.

The initial attack on the golf passes came from

Commissioner W. Kirk Phillips Jr., who called the passes troublesome and said the commission should "take a staunch stand" to eliminate the privileges.

"If I had my way, it would only be given to former commissioners and mayors who are the forgotten men of this community," Phillips said.

"It would be a sign of continued recognition."

Commissioner Joe Murphy agreed, and said he is "not

of being pleased by friends who want to use his pass.

But Commissioner W. L. Philbrick said he would fight to keep ministers on the list for four passes, saying they didn't have as much money as golf as other persons.

Phillips shot back at Philbrick, contending he wanted to cut taxes but was unwilling to trim "frivolous" expenses such as golf passes.

At this point, Phillips

delivered some criticism at the Coral Gables Times, saying it had reported the commission with making budget cuts which actually were not cuts.

He also said he was opposed to increased golf fees, stadium taxes, higher sewer charges and higher fees at Venetian Park.

At this point, Phillips asked to be excused from the meeting. He said he came to

(See Commission, page 3)

**Betsy Sees The Monkees**

See Page 16



**Coral Gables IMES**

THE CITY BEAUTIFUL

Vol. 48 No. 109 Thursday, July 13, 1967 28 Pages 10 Cents

**Kids Need Fireworks**

See page 6

### A Resort Tax For Gables?

Will be an expected line to elementary funds and reluctant to let businessmen with more money. Coral Gables this week said it may join Miami Beach in imposing a two per cent "resort tax" on hotel, apartment, restaurant and tavern bills.

Coral Gables City Commissioner Joe Murphy suggested adopting such a tax, and asked that it be referred to the city's charter committee.

A special election is to be held August 14 to decide whether the city should eliminate liquor laws. (See A, Page 25)

### Quick on the Draw Are Dennis and Judy Cox

... South Dale couple are performing at Pioneer City.

## These Two Are Looking For People Who Like To Draw

By STACEY COLEMAN

My husband is Duke County fireman. I used to shoot some before we were married," Mrs. Cox explained, "and now we've both taken up the sport."

The Cox's are a husband and wife Irish shooting team.

The duo have been shooting for about three years as a hobby, and in the last two months have formed their own gun-act for hire.

### Capital Improvements

## Metro-City Co-Op Law Hearing Set For July 25th

An amended law to insure that cities will coordinate all of their capital improvement projects with Metro was discussed Monday and set for a July 25 public hearing by Commissioners during an antitrust hearing.

The new ordinance requires that all governmental units submit information about their planned six-year capital improvement projects to Metro Commission Clerk Ed Peltan by Oct. 1 of every year.

Purpose of their requirements is that projects can be studied with an eye toward coordination and compatibility with Metro's general land use master plan.

County Manager Porter Homer will discuss the projects with the Planning Department and the Planning Advisory Board.

If recommendations which state the effect of such projects upon each other, will be submitted to the Metro Commission before Jan. 15 of each year.

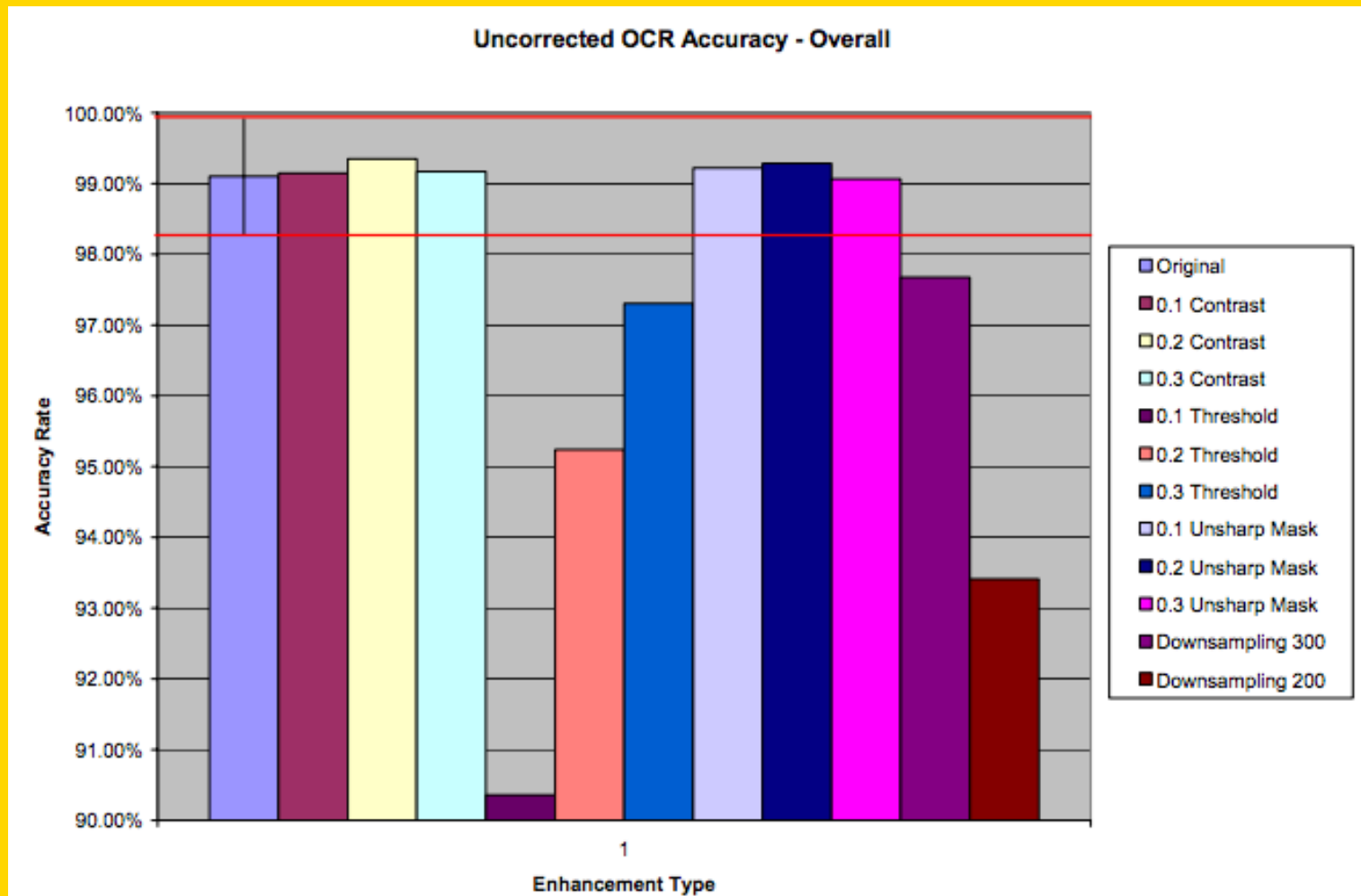
In other actions, Metro Commission:

**APPROVED** road design for six-lane dual road from SW 87th to SW 57th Avenue.

**AGREED** to issue certificate of individuality for \$1,200,000 to purchase Palmieri Country Club.

**AUTHORIZED** County Manager Porter Homer to file applications for federal

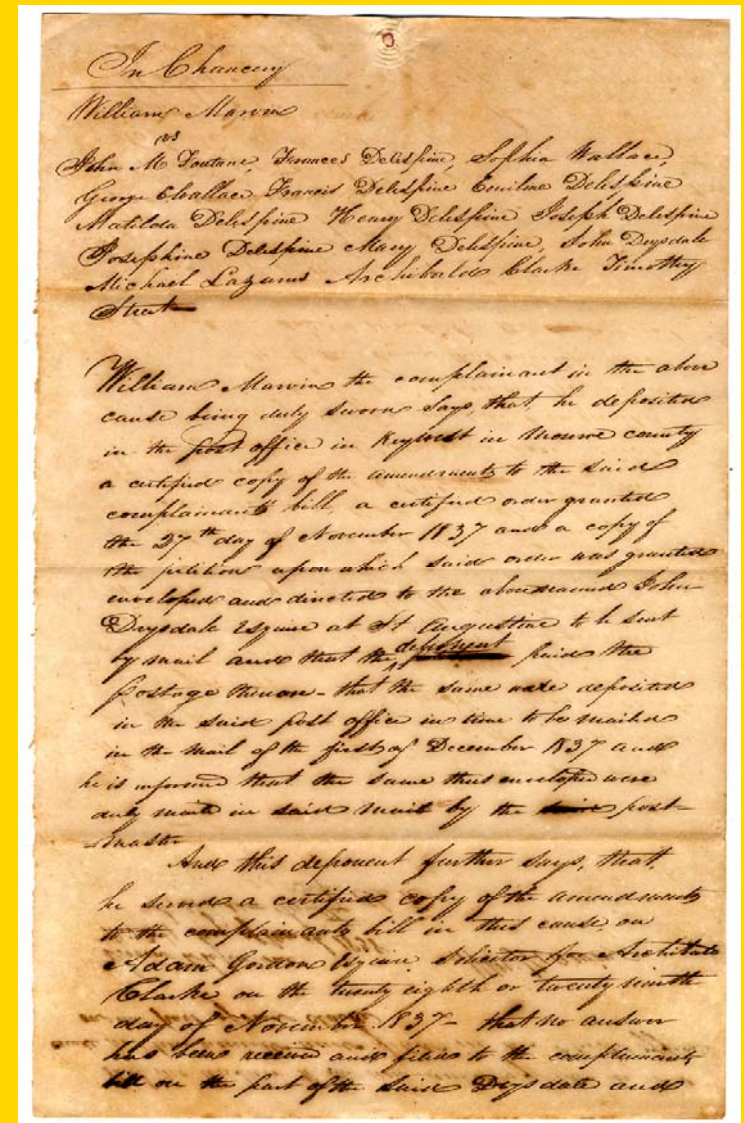
# Image processing and OCR Accuracy



“No single type of file enhancement, applied individually, improves character recognition rates for OCR.”  
*U.S. Government Printing Office, Washington, DC*

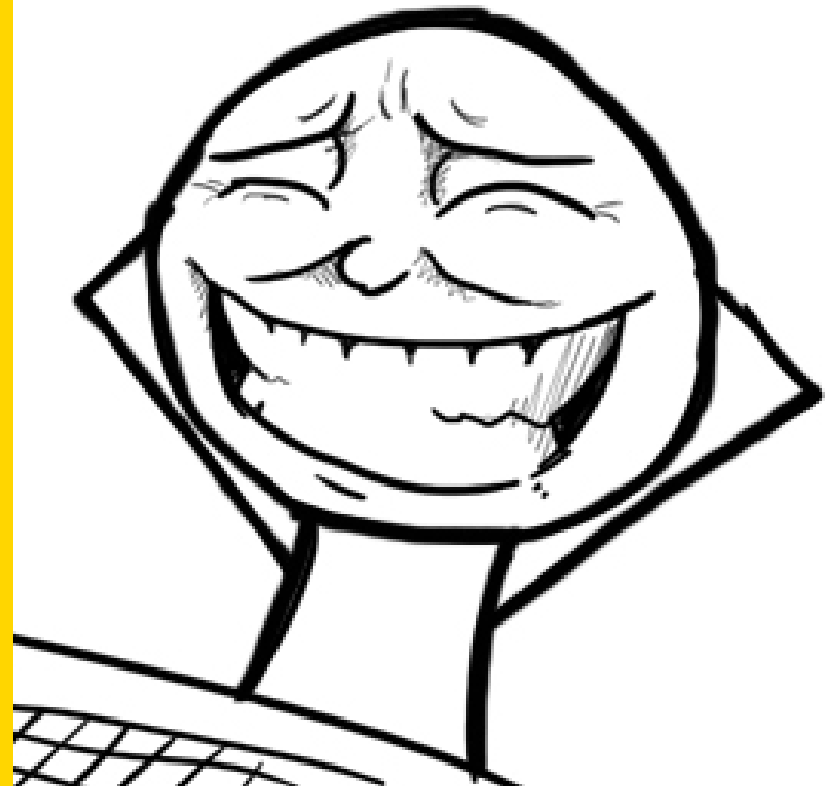
# Handwriting recognition

- Very low accuracy – especially with cursive, foreign languages
- Works best with forms or templates
- Intelligent Character Recognition



Congratulations!

**YOU'RE AN  
OCR EXPERT!**





# Resources

- Acrobatusers.com,. "Better PDF OCR: Clearscan Is Smaller, Looks Better". N.p., 2016. Web. 12 Jan. 2016.
- Borovikov, Eugene. "A Survey Of Modern Optical Character Recognition Techniques". *Arxiv.org*. N.p., 2014. Web. 11 Jan. 2016.
- Dlib.org,. "Measuring Mass Text Digitization Quality And Usefulness: Lessons Learned From Assessing The OCR Accuracy Of The British Library's 19Th Century Online Newspaper Archive". N.p., 2016. Web. 11 Jan. 2016.
- Lib.umich.edu,. "Digital Conversion OCR & Text Encoding | U-M Library". N.p., 2016. Web. 12 Jan. 2016.
- Library.illinois.edu,. "5.0 Best Practices For Optical Character Recognition". N.p., 2016. Web. 11 Jan. 2016.
- Tanner, Simon. "When The Data Hits The Fan!: Text Capture And Optical Character Recognition 101".*Simon-tanner.blogspot.com*. N.p., 2015. Web. 11 Jan. 2016.
- U.S. Government Publishing Office,. "Optimizing OCR Accuracy On Older Documents: A Study Of Scan Mode, File Enhancement, And Software Products". N.p., 2016. Web. 11 Jan. 2016.