

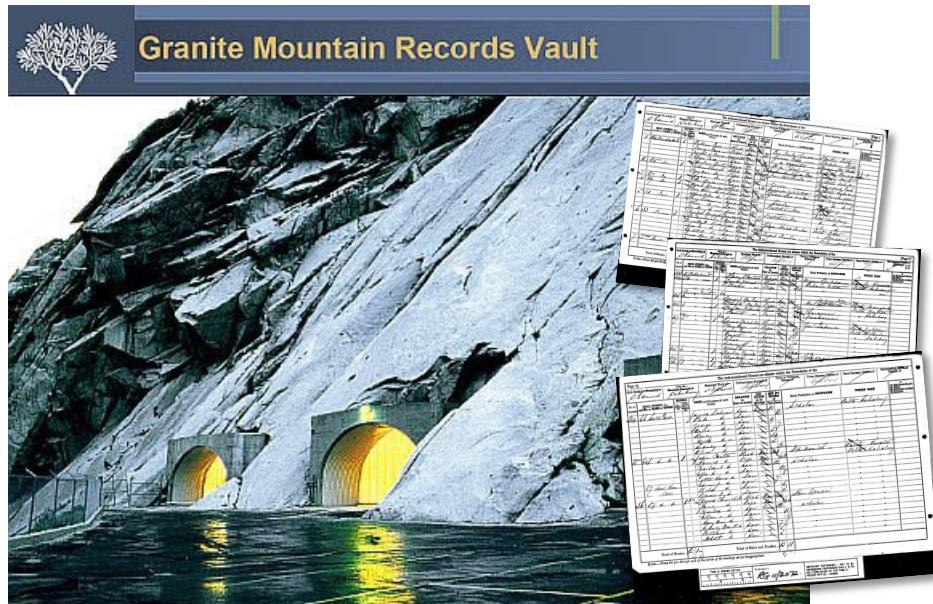


25 Years of Research in Family History Technologies at BYU: *Where we have been* *Where we are going*

Bill Barrett

Brigham Young University

The Microfilm Problem



- Early '90's
- Brittle Technology
- Poor Resolution

Granite Mountain Vault



Scanning the Granite Vault

2.5 Million rolls of microfilm



~ 1300 images per roll

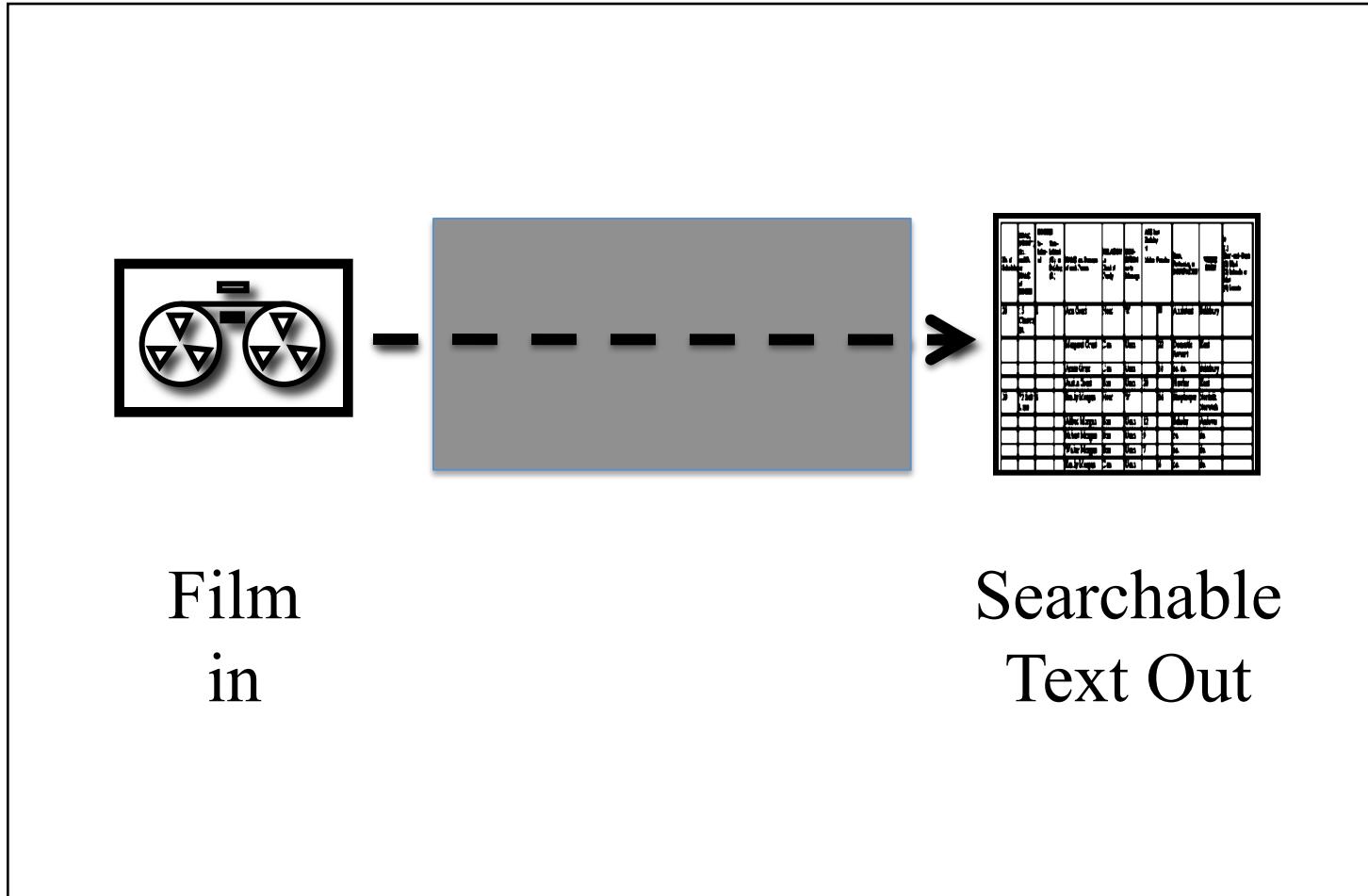
Up to 25 Megabytes per image

$1300 \times 25,000,000 \times 2,500,000$

= 81.25 Petabytes



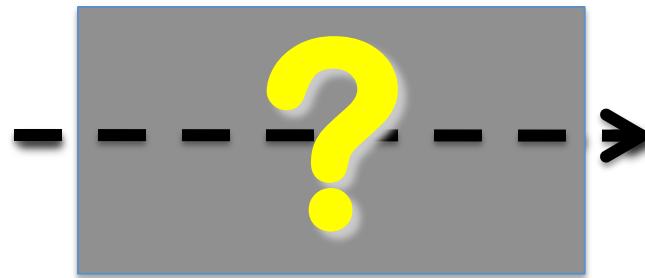
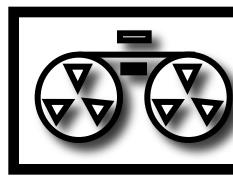
The Digital Microfilm Pipeline*





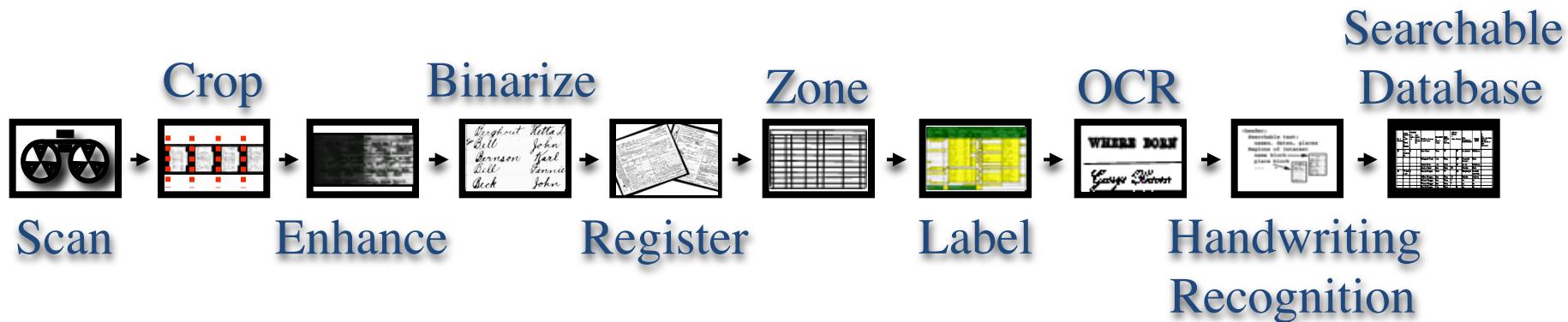
The Digital Microfilm Pipeline

What's inside the box?



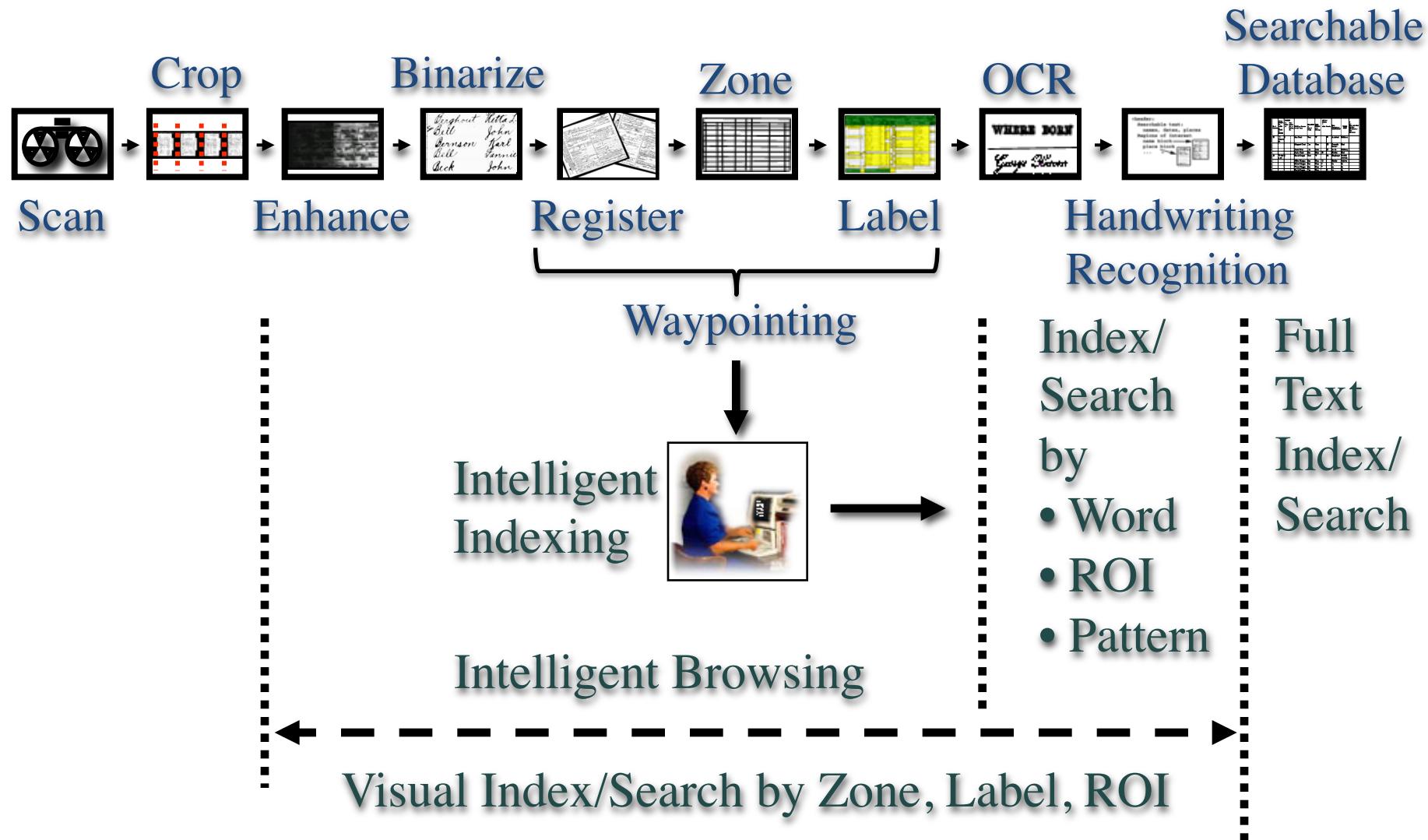
What processing steps have to happen in between?

The Digital Microfilm Pipeline

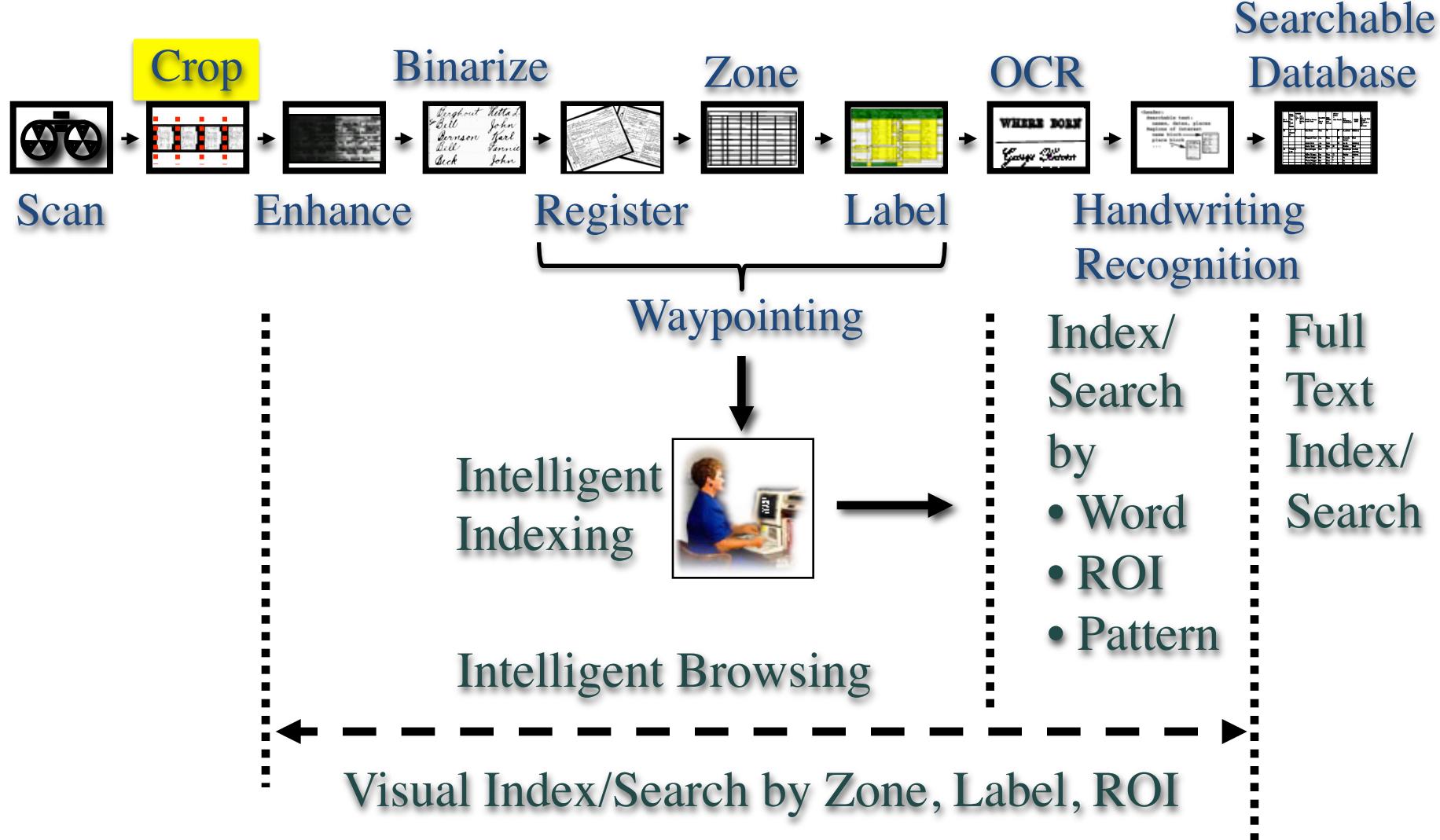


We envisioned a sequence of processing steps that needed to happen
 – like the Henry Ford Assembly line

The Digital Microfilm Pipeline



The Digital Microfilm Pipeline



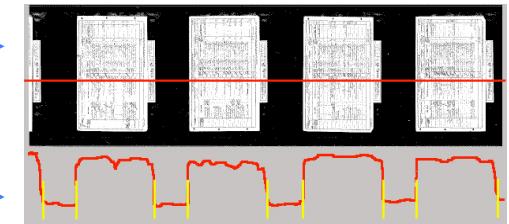


Cropping - 1994

Civil Parish (or Township) <i>R. May</i>		Municipal Borough of <i>Thensley</i>		Municipal Ward of <i>Wobh.</i>		Parliamentary Borough of <i>Thensley</i>		Town or Village or Hamlet of <i>Thensley</i>		Urban Sanitary District of <i>Thensley</i>		Rural Sanitary District of <i>Thensley</i>		Ecclesiastical Parish District <i>R. May</i>	
No. of Schedules	ROAD, STREET, &c. and No. or NAME of HOUSE	HOUSES	NAME and Surname of each Person	RELATION Head of Family	ONE- DIRECTION as to Marriage	AGE last Birthday of Male Females	Rank, Profession, or OCCUPATION			WHERE BORN				(1)Ded- (2)Ded- (3)Ded- (4)Ded- (5)Ded-	
4	1, Charlotte St. ("Queen's Arms")	1	Michael Morrison	Head	Mar.	31	Licensed Victualler			Middleton; Pington				11	
			Mary J. Do.	Wif.	Mar.	29		Selp.; Condore							
			Ellen Do.	Daus.		26		Selp.; Thensley						Lanc.	
			Elizabeth Morrison	Mother	W.	58	Servant	Selp.; Thensley							
			Tom Fox	Fir.	Con.	28	General Ser.	Selp.; Thensley		Hort.; Hobart					
			Catharine Doyle	Fir.	Con.	24	Barmaid	Pend							
5	2, Charlotte St.	1	Leander Weston	Head	Mar.	59	Grocer; (wastes, employing 2 men)	Cumberland; Wigton							
			Emma Do.	Wife	Mar.	56		Do.; Langnowra							
			William Do.	Fon		42	Soldier	Selp.; London							
			Frances Do.	Daus.		9		Selp.; Do.						Do-and- Lanc.	
			George Weston	Servtman	Con.	22	Grocer & Shopman	Middlesex; Fleetwood							
			Anna Cook	Servt.	Con.	22	General Serv.	Pend							
6			James F. Phillips	Head	Mar.	41	Barder's Clerk	Yorkshire; Leeds							
			Carroll Do.	Wif.	Mar.	39		Do.; Bradford							
			Sophia White	Fon.	Con.	19	General Serv.	Selp.; Bridgwater							
7	Do.	1/6													
7	2, Bird Lane	1	William Frampton	Head	Mar.	72	Couch Turner	Hertfordshire; Bidden							
			Anna Do.	Wif.	Mar.	69		Do.; Tameworth							
8	2, Bird Lane	1	Thomas Johnson	Head	Wob.	68	Printed Grocer	Dorset; Blandford							
			Henry Johnson	Fon	Con.	59	Cyanist	Selp.; Thensley						Bath.	
			Emma Do.	Daus.		44	Cust. Males	Middlesex; St. Pancras							
			Jane Frampton	Apprentice	Con.	18	Cust. Males (Apprentice)	Selp.; London							
9			Walter Campbell	Lodger	Con.	53	Ship Carpenter (out of employ)	Durham; Sunderland							
			Do.												
							End 1	St. Mi	last	Codicetion District					
Total of Houses...		1 U					Total of Males and Females...	9	13						
Total of Houses...		2 B													

PUBLIC RECORD OFFICE						Reference:-	COPYRIGHT PHOTOGRAPH - NOT TO BE REPRODUCED PHOTOGRAPHICALLY WITHOUT PERMISSION OF THE PUBLIC RECORD OFFICE, LONDON	
1	2	3	4	5	6		RG 11/2072.	
						11	4	

- Scan continuously in Ribbons →
 - Crop in batch, exploit consensus →



1920 Utah Census Example

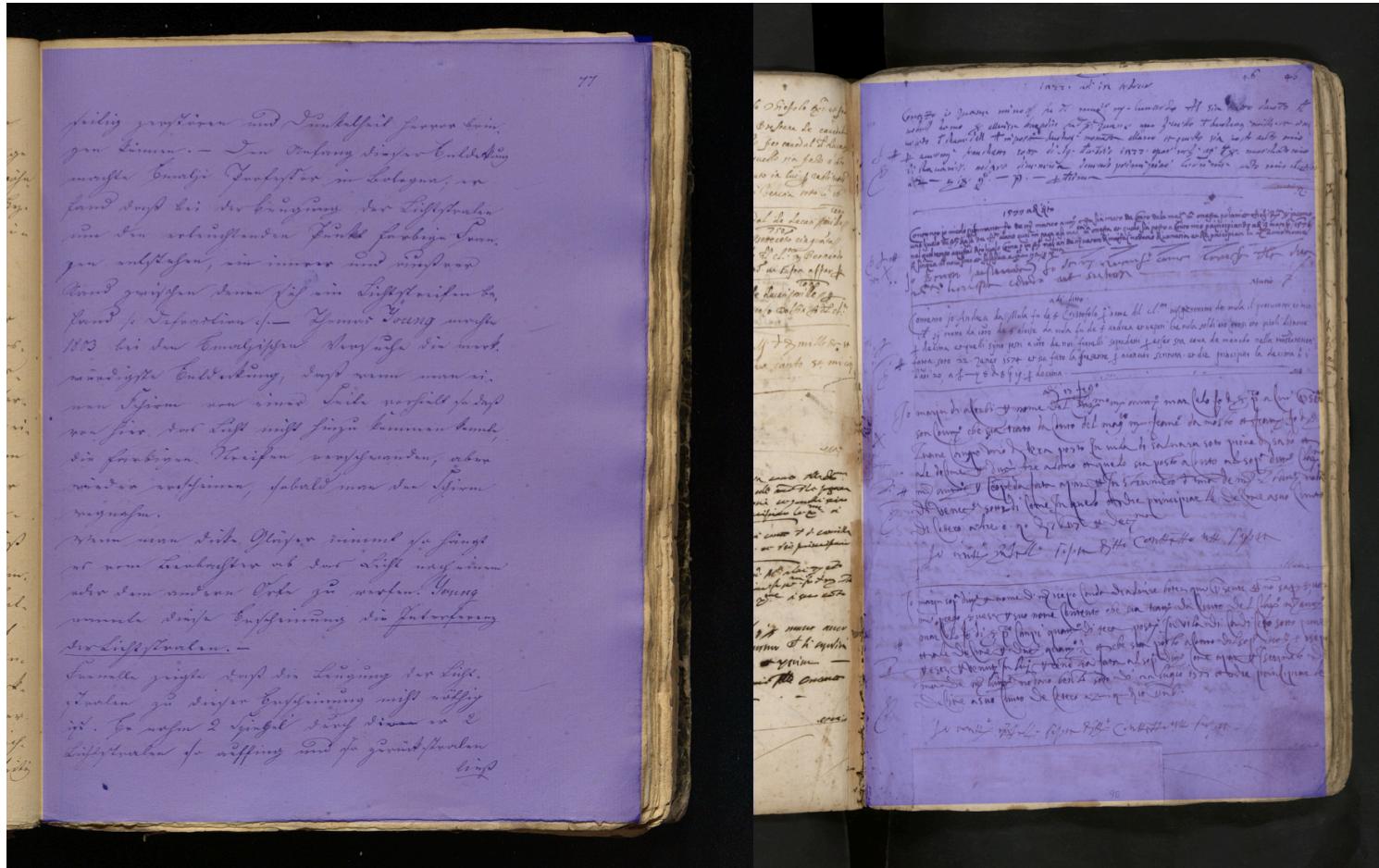


25 MB (~6K x 4K) images

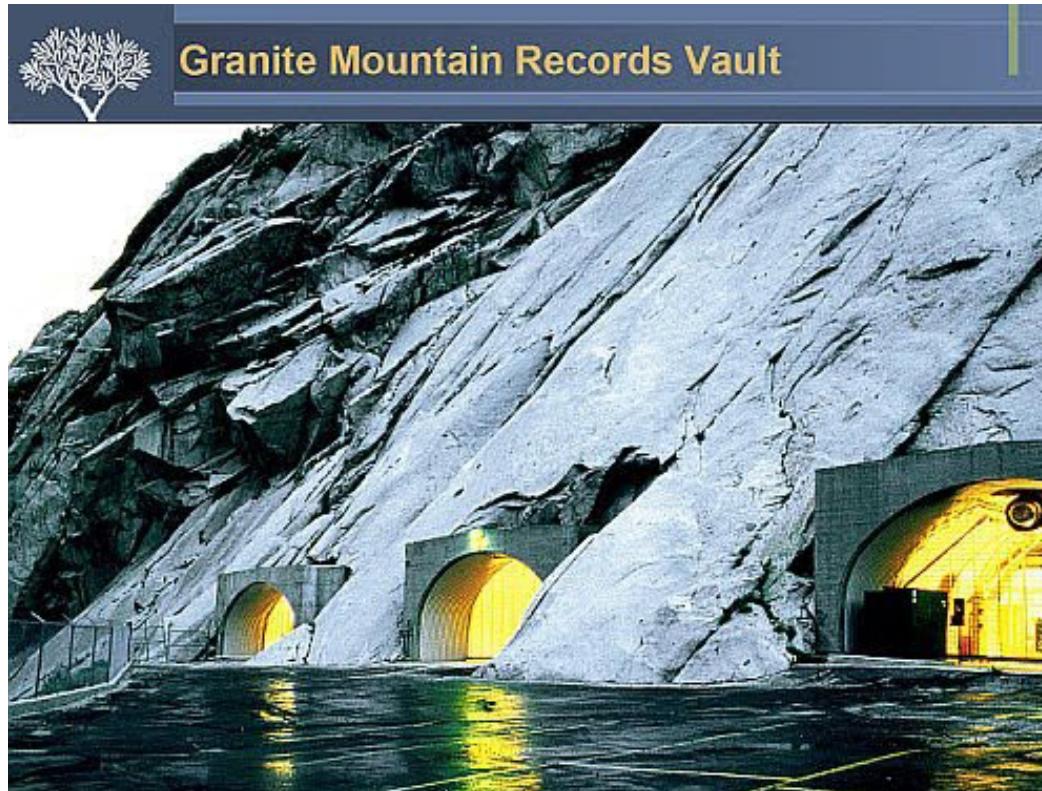
Cropping – 2017 Tensmeyer et al



PageNet – Fully Convolutional Network

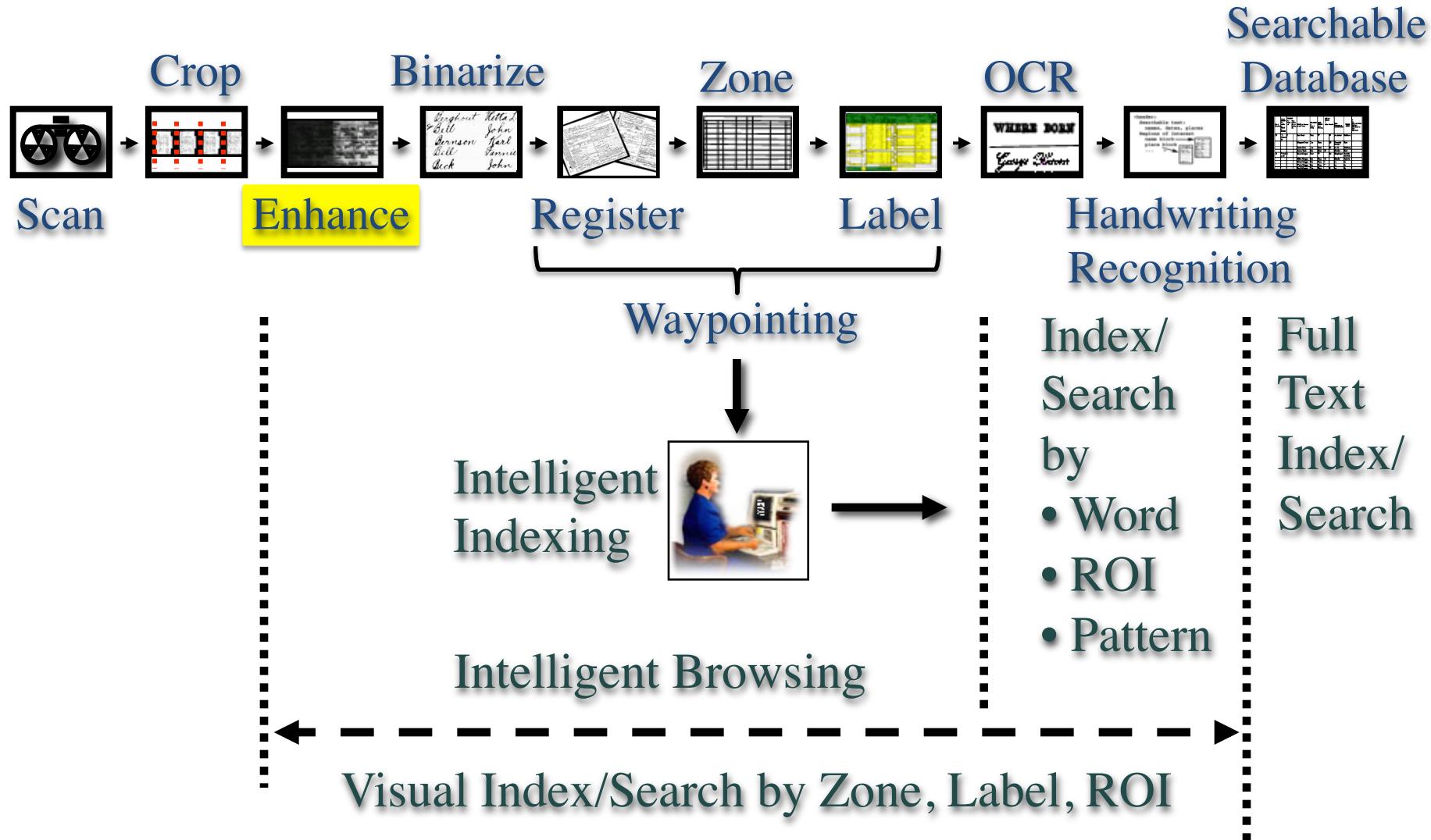


The Digital Microfilm Pipeline

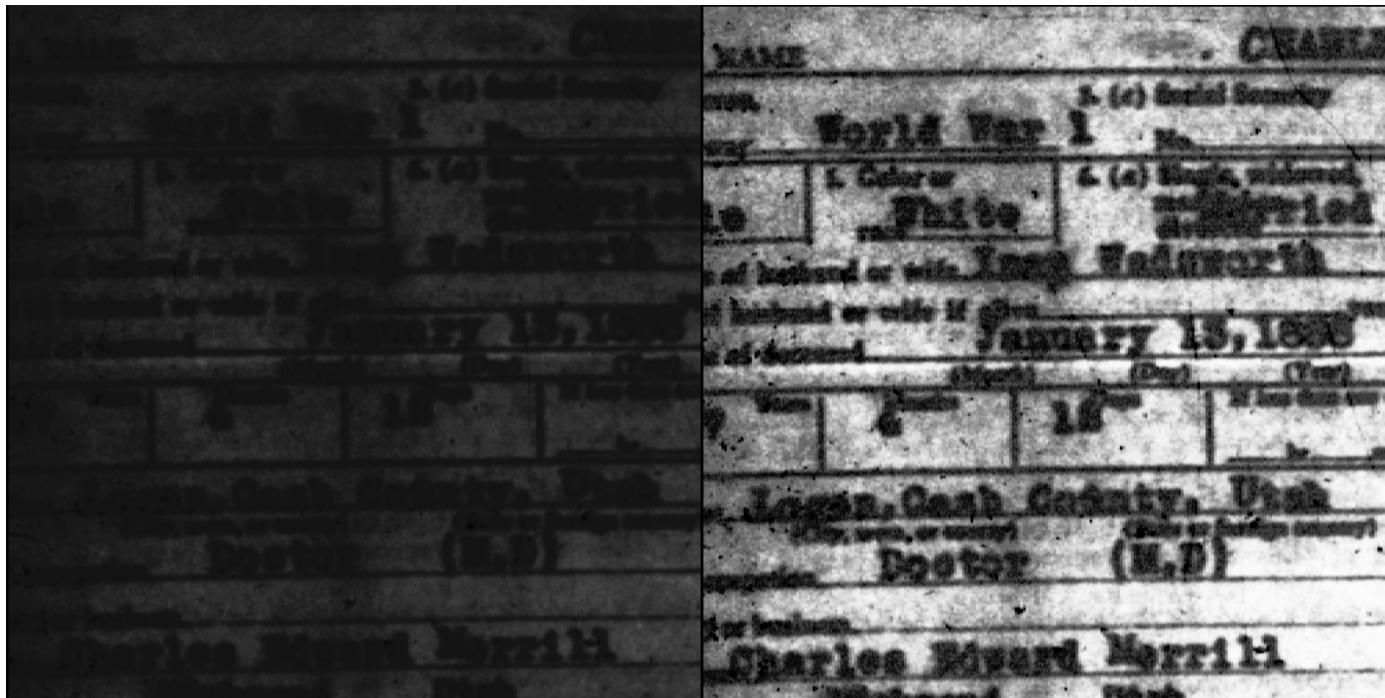


10 Terabytes/day

The Digital Microfilm Pipeline



Histogram Enhancement



(Left) Original scanned record (Right) After Enhancement

Threshold Enhancement

Miller, Riley is listed
Moore, Elizabeth J.
Moore, Ira the hill went
Moore, Small the way to
Mosier, Infant Son and
Mosier, Beryl he is now
Mosier, James R. Rich cr
Mosier, Jessie Baile you
Mosier, Rebecca Christia
Mosier, Susan cemetery
Mosier, Thomas P. off f
Mosier, Warren Leo above

Miller, Riley
Moore, Elizabeth J.
Moore, Ira S.
Moore, Wm.
Mosier, Infant Son
Mosier, Beryl
Mosier, James R.
Mosier, Jessie B.
Mosier, Rebecca Christia
Mosier, Susan
Mosier, Thomas F.
Mosier, Warren Leo

(Left) Original scanned record (Right) After thresholding for Bleed-through removal, with Enhancement, and Antialiasing

Stroke Enhancement



(Left) Original scanned record (Right) After
Matched Filter Stroke Enhancement



Family History Technology Workshop (fhtw.byu.edu) -2001

Secure | <https://fhtw.byu.edu>

FH TW Home Call for Papers Local Information Program Archive Registration

BRIGHAM YOUNG UNIVERSITY
Family History TECHNOLOGY WORKSHOP

27 February 2018, Hinckley Alumni Center at BYU

The Future of Family History

Research talks, developer presentations and lightning talks. Handwriting recognition, automated record transcription, data modeling, machine learning, natural language processing, visualization, human interaction, user experience. Any technology that makes family history better.

The first screenshot shows a handwritten pedigree chart for the Benson Martha family, with a red box highlighting a 'Son' entry. The second screenshot shows a digital pedigree chart for Leonidas Hartline Kammard, featuring a tree diagram with green and red branches, and a detailed profile for Leonidas Hartline Kammard. The third screenshot is a conceptual data model diagram showing entities like Person, Child, Spouse, and various attributes like BirthDate, BirthPlace, DeathDate, DeathPlace, Name, Gender, MarriageDate, MarriagePlace, BloodType, and Child.

About the Workshop

The Family History Technology Workshop has been held for 17 years, at both BYU and at RootsTech in Salt Lake City. Attendees include researchers, software developers, and professionals, brought together by their shared passion for improving family history technology.



Family History Technology Workshop ([fhtw.byu.edu](https://rootstech.ce.byu.edu)) - 2001

Secure | <https://rootstech.ce.byu.edu>

BYU Continuing Education

PROGRAMS EMAIL LIST ABOUT Q

RootsTech

HOME ROOTSTECH CONFERENCE WEBSITE [↗](#) FAMILY HISTORY & GENEALOGY [↗](#)

BYU created conference with FamilySearch: RootsTech 



February 28– March 3, 2018

Salt Palace Convention Center
Salt Lake City, Utah

For Non-Credit or Non-BYU Employees
You may register at rootstech.org [↗](#)

The Brigham Young University logo, which is a stylized tree with many branches.

Brigham Young University is pleased to participate with FamilySearch in redesigning three former conferences, the Conference on Computerized Family History, the Technology Workshop, and the FamilySearch Developers Conference, all together into the new RootsTech and offering even more to those interested in family history.

RootsTech

[Register Now](#) [↗](#)

Online Registration Closes: Mar 3, 2018

[Manage My Continuing Education Account](#) [↗](#)

Employees may qualify for free admission.

For more information,
visit rootstech.org [↗](#)

BRIGHAM YOUNG UNIVERSITY

Family History Technology Lab (fhtl.byu.edu) - 2001



The screenshot shows the homepage of the BYU Family History Technology Lab. At the top, there is a dark navigation bar with the text "BYU FHTL" on the left and "Home" "People" "Contact" "Donate" on the right. Below the navigation bar is a large, dark, abstract background graphic with glowing blue and yellow lines forming a stylized brain or network pattern. In the center, there is a dark rectangular box containing the text "Family History Technology Lab" in white. To the left of this text is a white icon of a circuit board or network connections. To the right is a block of text: "The BYU Family History Technology Lab is a non-profit student research lab sponsored by Brigham Young University's Computer Science department." Below this text is another block of text: "We're dedicated to creating simple, usable, and enjoyable family history apps." At the bottom right of the central box is a button labeled "Watch a Training Video". To the right of the central box is a vertical sidebar with several circular icons: a network icon, a "RF" monogram, a power plug icon, the letters "OPG", the American and German flags, a "G!" icon, a circular color wheel, and a smiley face icon.

BYU FHTL

Home People Contact Donate

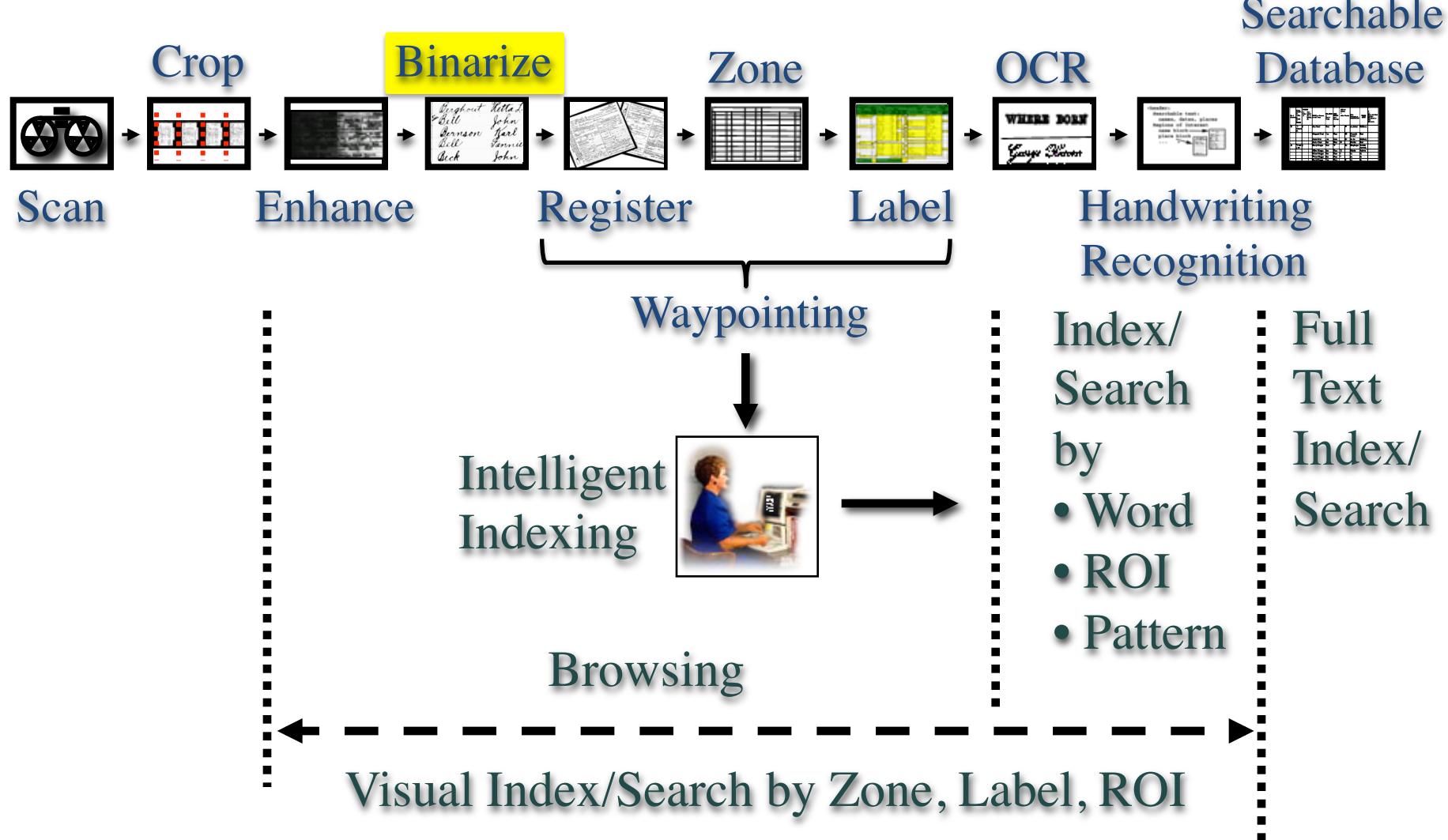
Family History Technology Lab

The BYU Family History Technology Lab is a non-profit student research lab sponsored by Brigham Young University's Computer Science department.

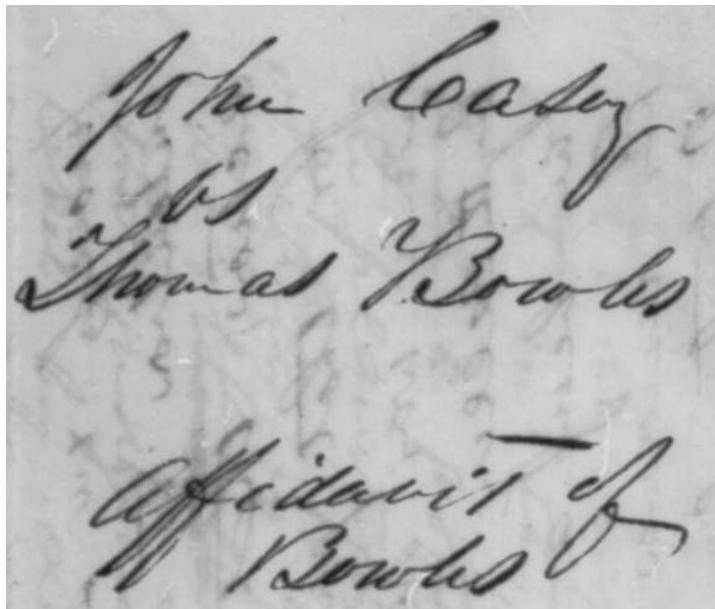
We're dedicated to creating simple, usable, and enjoyable family history apps.

Watch a Training Video

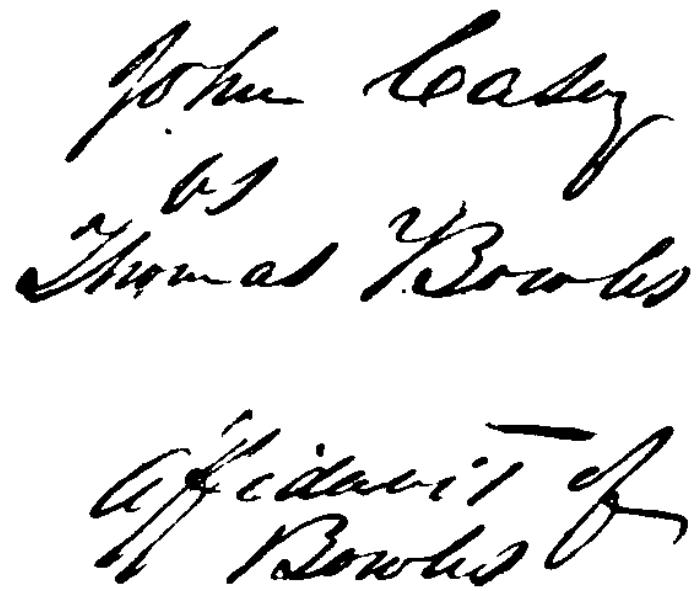
The Digital Microfilm Pipeline



Binarization



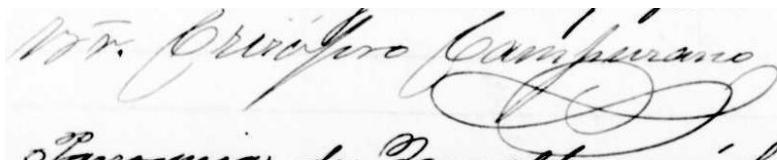
(a) Original Image



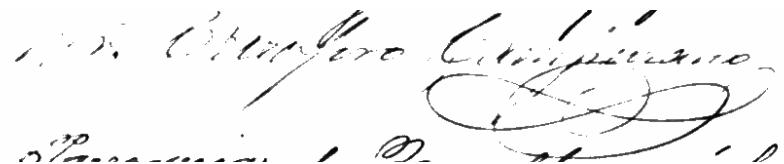
(b) Binarized Result

Median Filter Subtraction for background elimination

Binarization



(a) Original with background removed



(b) First set of text pixels (S_1)



(c) Second set of text pixels (S_2)



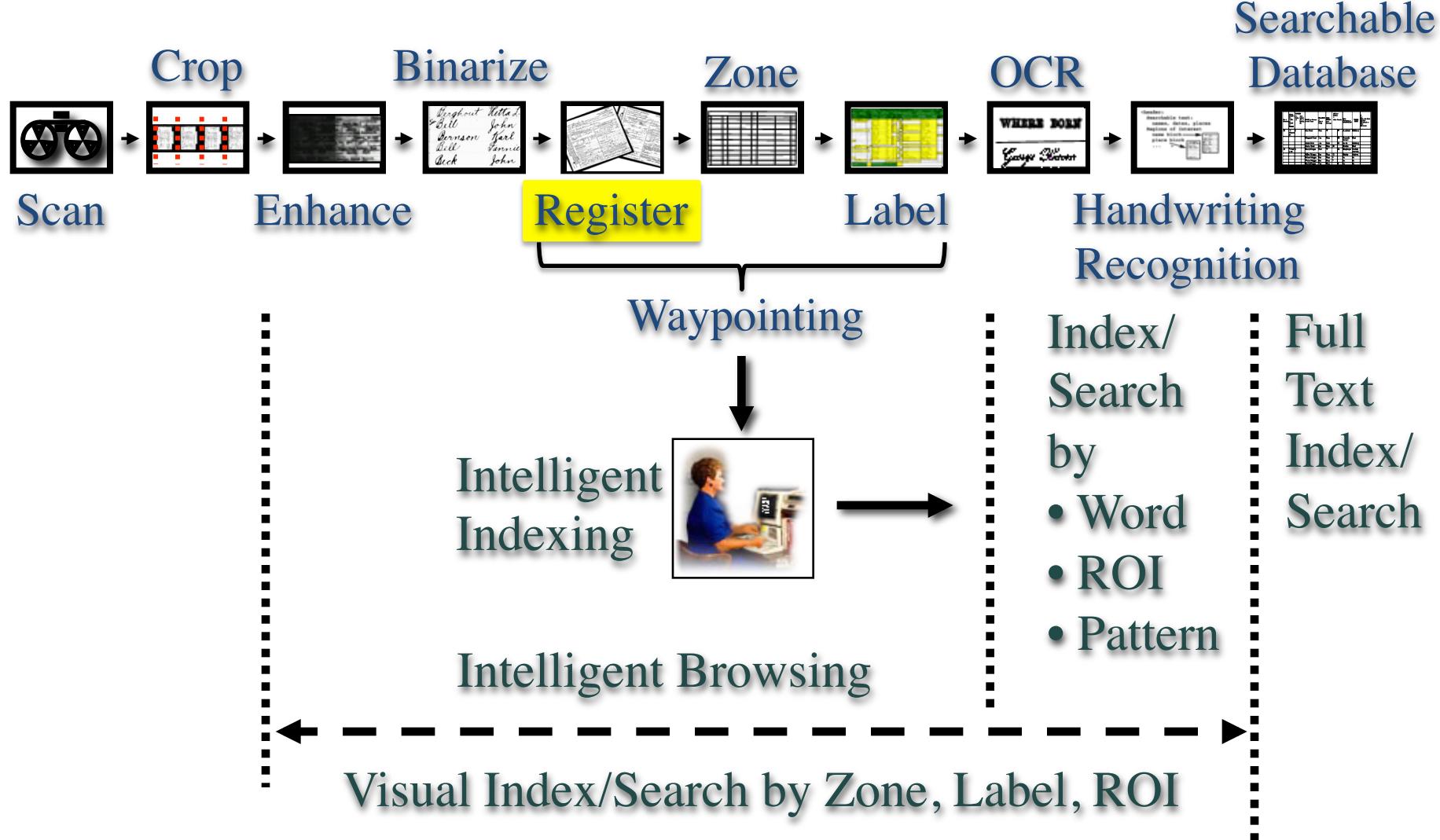
(d) Third set of text pixels (S_3)



(f) Result ($S_1 + S_2 + S_3 + S_4$)

Recursive Otsu Intensity Parsing and Accumulation

The Digital Microfilm Pipeline





Registration

NAME and Surname of each Person		R	Head
Grace Kelly Lawrence			
Mary Ann			
Carlton C			
Agnes G			
11.11.11.11.11.			

+

NAME and Surname of each Person		R	Head
Walter Gouverneur			
Edmund d.			
John K Farquhar			
Camelline d -			
11.11.11.11.11.			

+

NAME and Surname of each Person		R	Head
William Packino			
Anne d -			
Albert d			
John Stergold			
11.11.11.11.11.			

=

NAME and Surname of each Person		R	Head
Walter Gouverneur			
Edmund d			
John K Farquhar			
Camelline d -			
John Stergold			
11.11.11.11.11.			

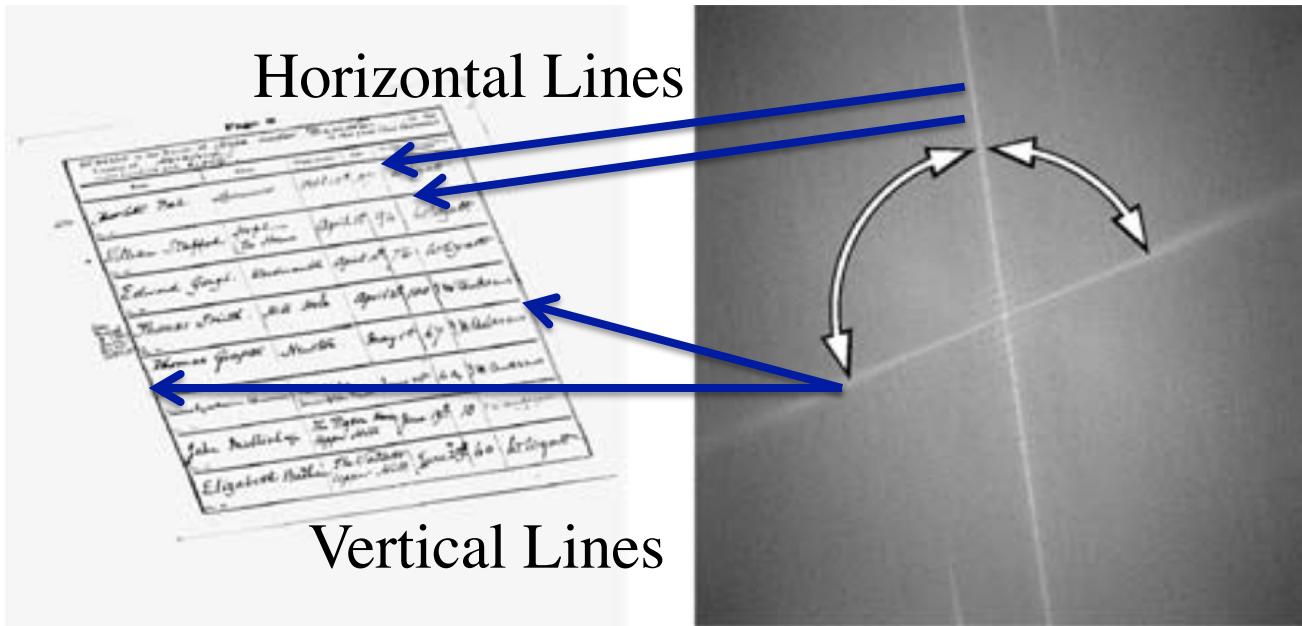
Not registered

NAME and Surname of each Person		R	Head
Walter Gouverneur			
Edmund d -			
John K Farquhar			
Camelline d -			
11.11.11.11.11.			

Registered

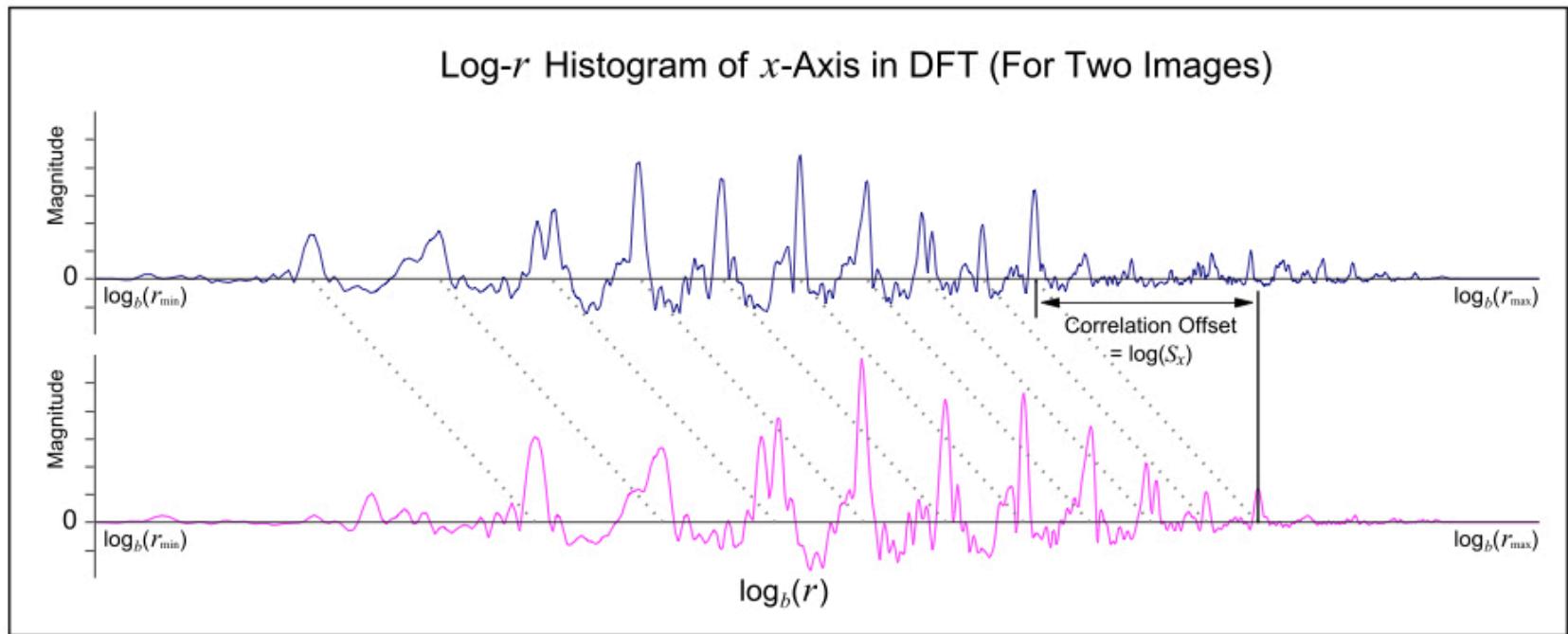
Documents must be aligned, rotated and scaled to register them

Registration – Hutchison 2003



Fourier-Mellin Transform for detection of rotation and shear

Registration – Hutchison 2003



Fourier-Mellin Log Correlation, Scale Correction



Registration – Hutchison 2003



Register

Alignment, rotation and scaling using the Fourier-Mellin Transform allows registration to sub-degree and sub-pixel accuracy.



Registration – Hutchison 2003

Unregistered
Mediaan

West Virginia State Department of Health
DIVISION OF VITAL STATISTICS

CERTIFICATE OF DEATH

Register No. _____

WEDNESDAY, FEBRUARY 27, 1974

1. PLACE OF DEATH (Dist. No. _____)
County _____
District _____
or
Town or City _____ No. _____

2. FULL NAME
PERSONAL AND STATISTICAL PARTICULARS

3. SEX COLOR OR RACE SINGLE
 MARRIED
 DIVORCED
 SEPARATED
(Name the wife)

4. DATE OF BIRTH
Month Day Year

5. AGE LESS than
 ONE year
 ONE year
 ONE year
 ONE year

6. OCCUPATION
 Farmer, agriculturist
 General laborer
 General worker of industry
 General merchant or salesman
 Other employer or manager

7. BIRTHPLACE
(Name of country)

8. NAME OF FATHER

9. BIRTHPLACE OF FATHER
(Name or country)

10. MAIDEN NAME OF MOTHER

11. BIRTHPLACE OF MOTHER
(Name or country)

12. IS THE ABOVE A STRANGER TO THE PLACE OF DEATH
Unknown _____
Yes _____
No _____

13. DATE _____ REGISTRATION NO. _____

14. DATE OF DEATH
Month Day Year

15. I HEREBY CERTIFY, That I attended deceased from _____, 19____, to _____, 19____, that I last saw him/her on _____, 19____, and that death occurred, on the date stated above, at _____, the CAUSE OF DEATH was as follows:
(Describe)

16. CONTRACTOR
(Signature) _____ Date _____

17. SIGNATURE _____ Date _____

NOTE: State the Disease Contract Disease, Isolated from Vacant Coffin, Dead House or Cemetery, and whether Autopsied, Buried, or Buried.

18. LENGTH OF RESIDENCE FOR HOSPITAL, INSTITUTION, TENANTS OR RECENT RESIDENTS
At place of death _____ to the time _____
Where was disease recovered
If not at place of death _____

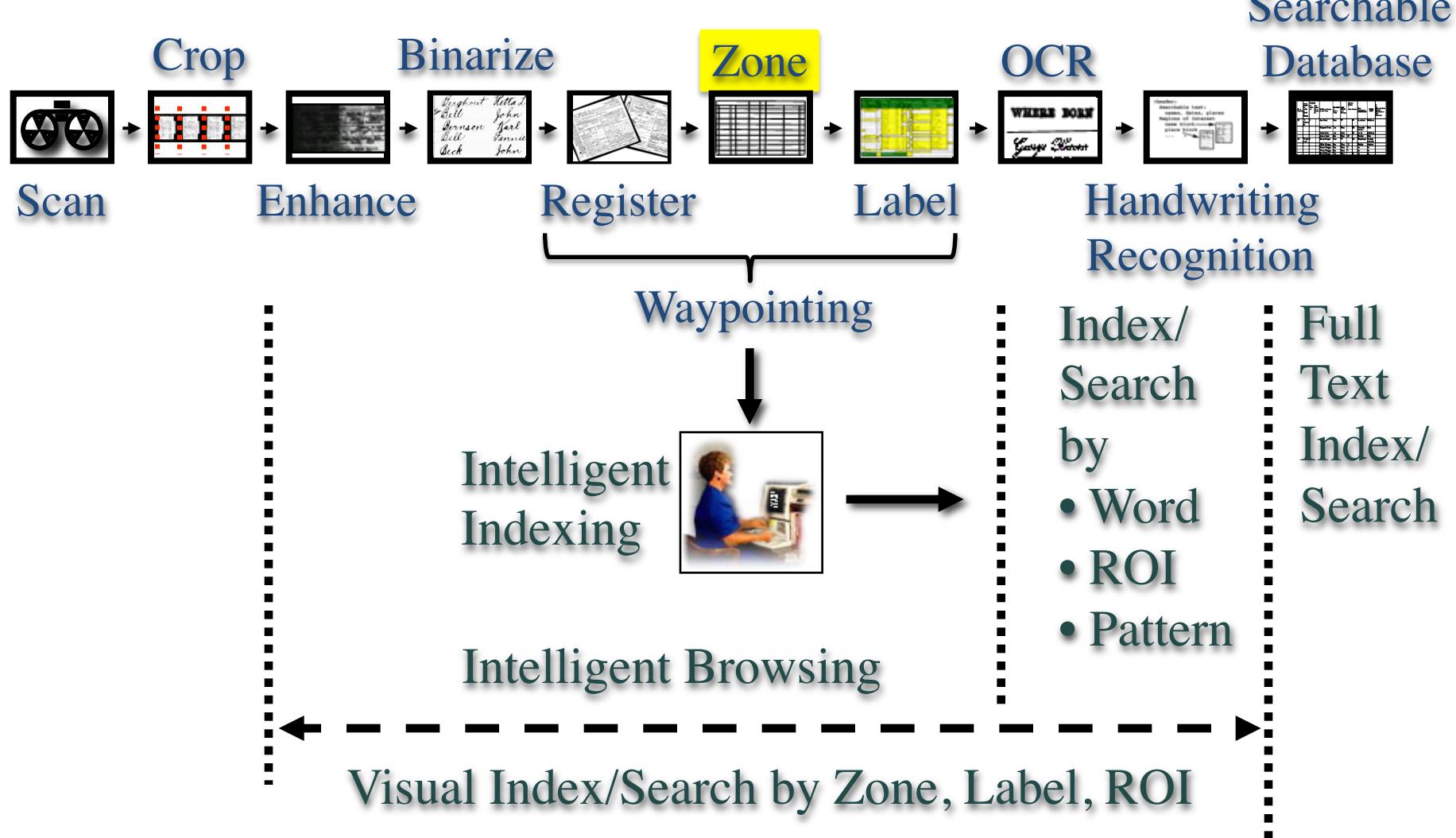
19. FORM OF DEAD REMAINS

20. PLACE OF BURIAL OR BURIED IN _____ DATE OF BURIAL _____

21. CEMETERY NUMBER _____

445 Images

The Digital Microfilm Pipeline



Segmentation & Zoning

- Nielson 2003



The undermentioned Houses are situate within the Boundaries of the								
Civil Parish [or Township] of Hugglesstone St Peter		City or Municipal Borough of	Municipal Ward of	Parliamentary Borough of Lepton	Town or Village or Hamlet of Birstall	Urban Sanitary District of	Rural Sanitary District of	Ecclesiastical Parish or District of
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES In which dwelt or is dwelling (B.)	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of Male Female	Rank, Profession, or OCCUPATION	WHERE BORN
4	continued		Susan Parr	Servant	Mar.	21	Butler	London Bridgwater
			Frances Morris	Servt	Mar.	25	Postman	Newport Smt'
5	Victory Cottage	1	James Lake	Head	Mar.	52	Postman	Leeds Market Cambridge
10	10.	1	Eliza Lampard	Widow	Mar.	45	Postman	Leeds Market Cambridge
			Eliza Lampard	wife	Mar.	45	Postman	Leeds Market Cambridge
			Emily Jane Lee	Dauch	Mar.	25	Pilot Seafarer	Birstall 10.
			Anna Lee	Dauch	Mar.	10	Scholar	Birstall 10.
7	10.	1	Edward Lampard	Brood	Mar.	48	Postman	Hugglesstone St Peter
10	Leaven Cottage	1	Elizabeth S. Souter	Heads	Mar.	34	Labourer	Leeds Market Cambridge
			Anna Lee	wife	Mar.	34	Labourer	Leeds Market Cambridge
			Henry Lee	Son	Mar.	13	Labourer	Leeds Market Cambridge
			Eliza Lee	Dauchter	Mar.	10	Scholar	Leeds Market Cambridge
			Ed	Son	Mar.	10	Postman	Leeds Market Cambridge
			Eliza	Son	Mar.	11	Postman	Leeds Market Cambridge
			Anna	Son	Mar.	11	Postman	Leeds Market Cambridge
			Eliza	Son	Mar.	11	Postman	Leeds Market Cambridge
			Eliza Lee	Son	Mar.	11	Postman	Leeds Market Cambridge
			Eliza Lee	Son	Mar.	11	Postman	Leeds Market Cambridge
			Eliza Lee	Son	Mar.	11	Postman	Leeds Market Cambridge
			Eliza Lee	Son	Mar.	11	Postman	Leeds Market Cambridge
			Eliza Lee	Son	Mar.	11	Postman	Leeds Market Cambridge
6	10.	1	Isabella Sugs	Head	Mar.	32	6 Green	Leeds Market Cambridge
			Elizabeth Sugs	wife	Mar.	32	Postman	Leeds Market Cambridge
			Anna Lee	Son	Mar.	10	Scholar	Leeds Market Cambridge
			John Lee	Son	Mar.	10	Postman	Leeds Market Cambridge
			Eliza Lee	Son	Mar.	10	Postman	Leeds Market Cambridge
10	Church Lane	1	Maria Swindman	Widow	W	60	Postale Officatu	Leeds Market Cambridge
			Eliza Staples	Servt		7	Servants	Birstall 10.
6	Total of Houses...	11	Total of Males and Females...	11	12			

NOTE.—Draw the pen through such of the words of the headings as are inappropriate.

Candidate Line Identification

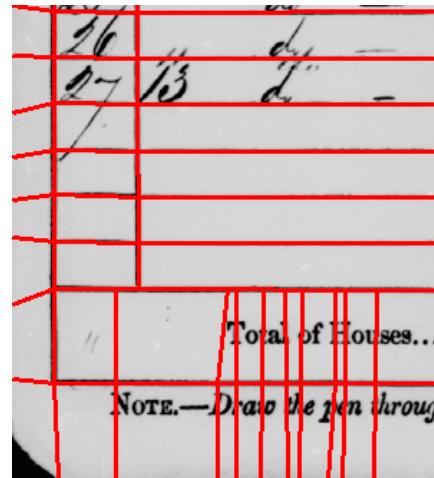
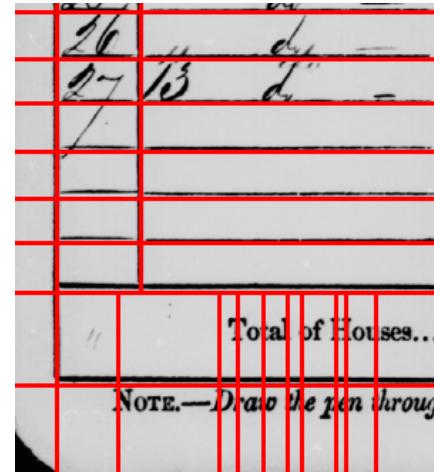
- Nielson 2003



Local “Snapping”

- Adjust the position of the mesh line to the digitized line
- For each line segment in the mesh, compute its filtered profile
- Calculate line strength

$$s(x) = e^{-\frac{(p_g - x)^2}{2s^2}} \cdot lp_f(x)$$
- “Snap” to the location maximizing $s(x)$



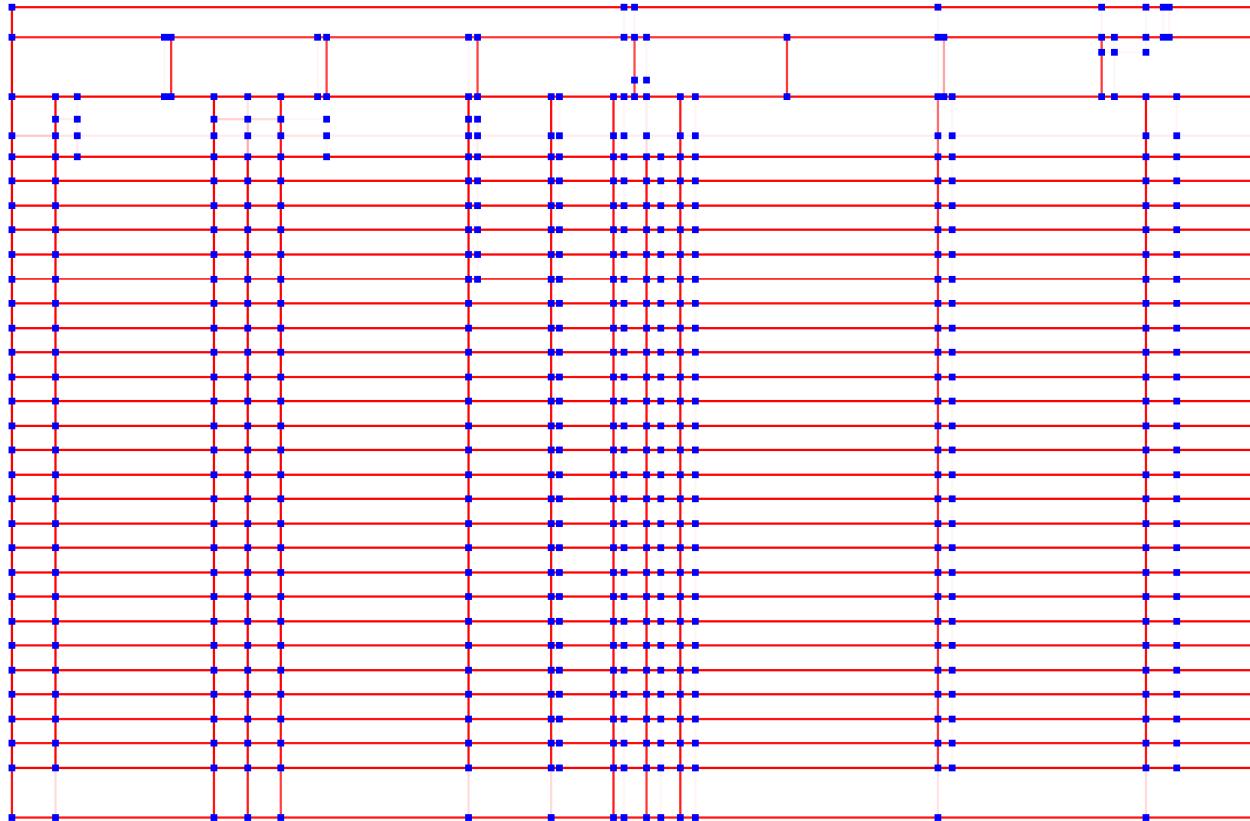


Zoning

The undermentioned Houses are situate within the Boundaries of the										
Civil Parish [or Township] of P. Edmund.		City or Municipal Borough of P. Edmund	Municipal Ward of P. Edmund	Parliamentary Borough of P. Edmund	Town or Village or Hamlet of P. Edmund	Urban Sanitary District of P. Edmund	Rural Sanitary District of P. Edmund	Ecclesiastical Parish or District of		
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES Un- in- habited (U.) or Occupied (O.)	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last BIRTHDAY of Male Inhabitants	Rank, Profession, or OCCUPATION	WHERE BORN	II (1) Dumb-Dumb (2) Mind (3) Imbecile or Idiot (4) Lunatic	
8	Collected	104								
9	29 Main Street	1	Charles Hollaway Head	X	Wife	50	Saxton	Hartlepool		
			Robert H. d.	Wife	Widow	50	Harold L. Hollaway	London	P. Marshall	
			Eliza d.	Daughter	Widow	40	d.	d.	W. Hollaway	
			Frances d.	Daughter	Widow	38	d.	d.	d.	
			William d.	Son	Widower	37	W. d.	d.	d.	
			Frances d.	Daughter	Widow	36	W. d.	d.	d.	
			Willie C. b.	son	Widower	36	W. d.	d.	d.	
			John d.	son	Widower	35	W. d.	d.	d.	
			Elizabeth Hart	Daughter	Widow	35	W. d.	d.	d.	
			Willie d.	Son	Widower	32	W. d.	d.	d.	
			Charles Edmund	Head	Widower	30	Schoolmaster	W. d.	d.	
			Eliza May d.	Wife	Widow	30		Carter		
			Hannah C. d.	Wife	Widow	29	Former Wife	Denton		
			Eliza May Mother	Daughter	Widow	17		W. d.		
			Joseph d.	son	Widower	16		d.	d.	
			Willie d.	son	Widower	8		d.	d.	
			George d.	Daughter	Widow	7		d.	d.	
			Elizabeth d.	Daughter	Widow	7		d.	d.	
			Gaura d.	Daughter	Widow	16		d.	d.	
Total of Houses...		11	Total of Males and Females...		8	15				

Line detection and local refinement for document zoning

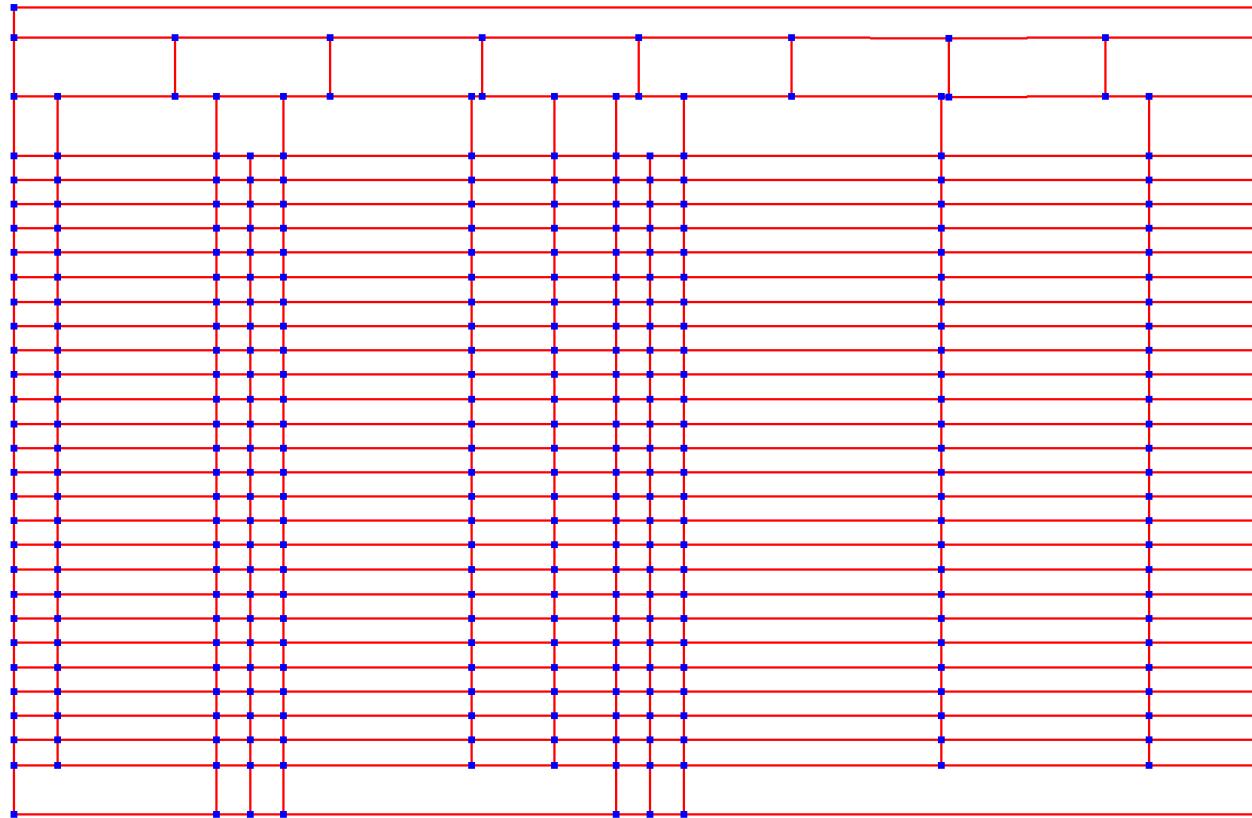
Zoning by Consensus



Consensus from 28 frames



Zoning by Consensus



Final Zone Template

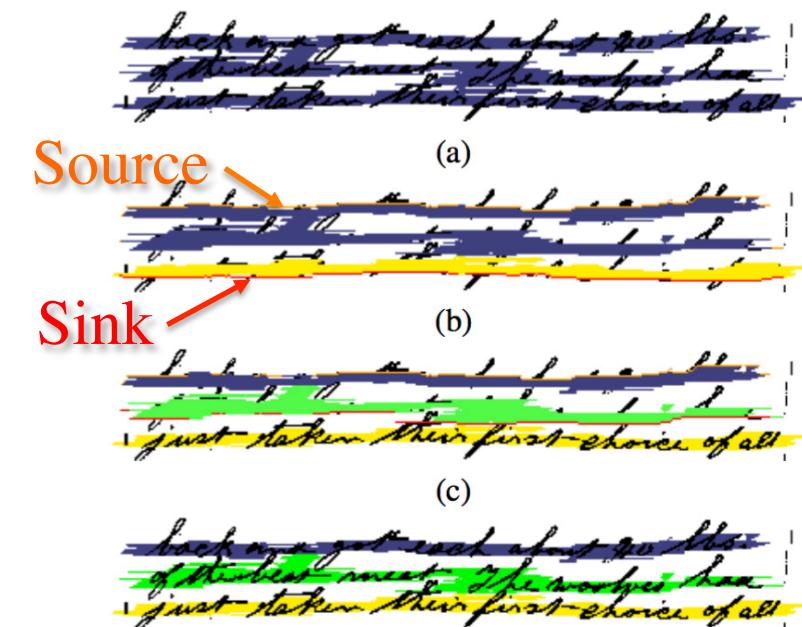
Free-form Line Detection

- Kennard, 2006

and all hearts merry, but none more glad than ours.

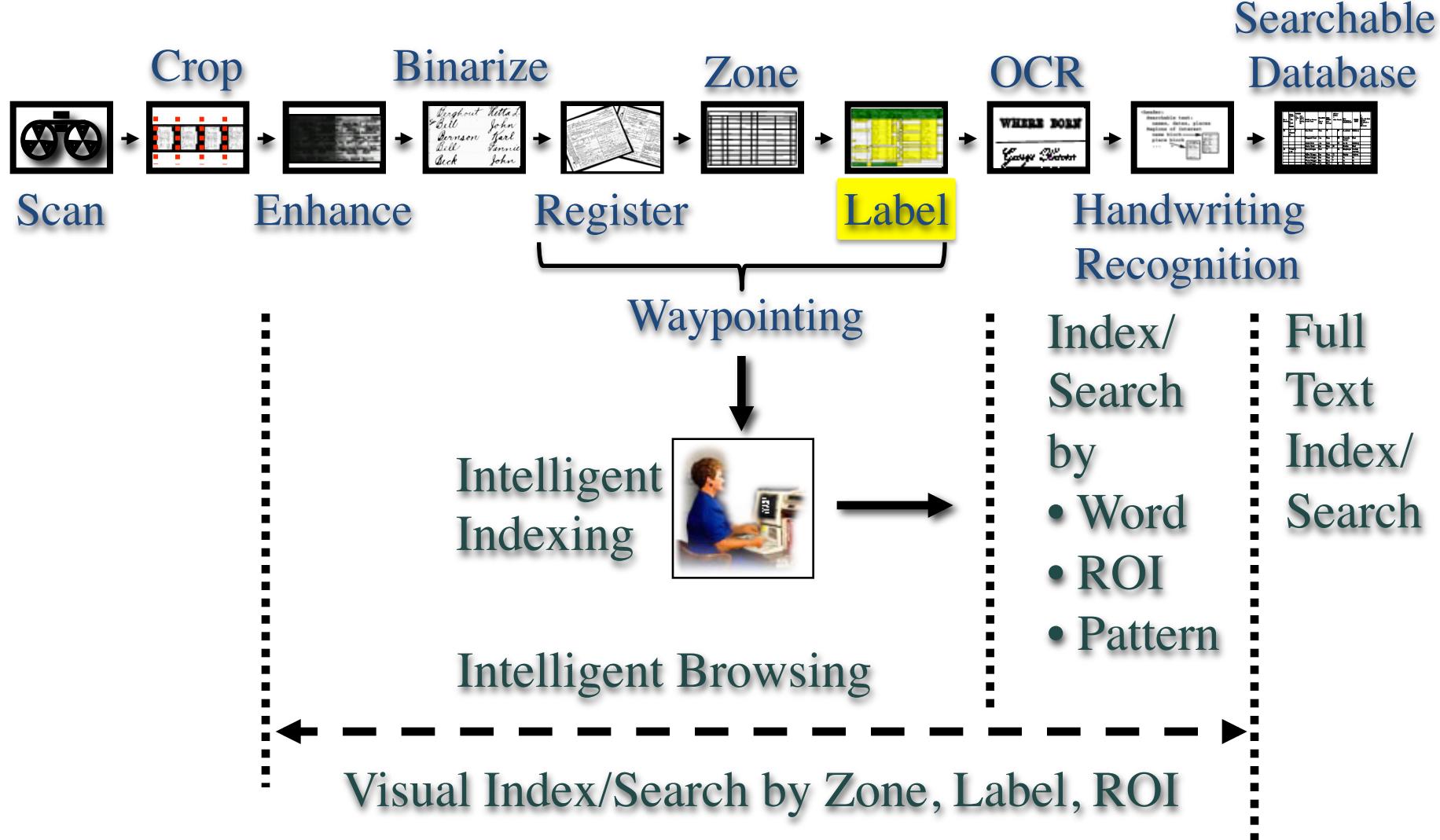
Next morning 3 of us went back and got each about 40 lbs. of the best meat. The wolves have just taken their first choice of all game; the insides.

The day we reached the upper crossing of the Platt, Aug. 4th a high mountain lay in faint view, through the smoke which bounded the atmosphere, which some guessed to be one, some two, some three miles distant, and so



Recursive Graph Cut Line Segmentation

The Digital Microfilm Pipeline





Label Machine Print

The undomesticated Horses are those within the boundaries of the											
Cattlemaster's Name		County		Municipal District		Parishment District		Union Boundary District		State Boundary District	
John Edward Bokhbay		Benton		Benton		Benton		Benton		Benton	
No. of MALES, FEMALEs, & NEUTERED HORSES		RELATION		NAME AND BOUNDARY OF each		RELATION		NAME AND BOUNDARY		RELATION	
MALES		FEMALEs		NAME AND BOUNDARY		FEMALEs		NAME AND BOUNDARY		NEUTERED	
Page 2											
The undomesticated Horses are those within the boundaries of the											
Cattlemaster's Name		County		Municipal District		Parishment District		Union Boundary District		State Boundary District	
John Edward Bokhbay		Benton		Benton		Benton		Benton		Benton	
No. of MALES, FEMALEs, & NEUTERED HORSES		RELATION		NAME AND BOUNDARY OF each		RELATION		NAME AND BOUNDARY		RELATION	
MALES		FEMALEs		NAME AND BOUNDARY		FEMALEs		NAME AND BOUNDARY		NEUTERED	
Page 3											
The undomesticated Horses are those within the boundaries of the											
Cattlemaster's Name		County		Municipal District		Parishment District		Union Boundary District		State Boundary District	
John Edward Bokhbay		Benton		Benton		Benton		Benton		Benton	
No. of MALES, FEMALEs, & NEUTERED HORSES		RELATION		NAME AND BOUNDARY OF each		RELATION		NAME AND BOUNDARY		RELATION	
MALES		FEMALEs		NAME AND BOUNDARY		FEMALEs		NAME AND BOUNDARY		NEUTERED	
Total of Horses 45 - Total of Males and Females 914											
NOTE.—Draw the pen through all of the words of the heading as and disappears.											

Label Handwriting



Labeled Image

[Example]

The undermentioned Houses are situate within the Boundaries of the									18
Civil Parish [or Township] of	Municipal Borough of	Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of	Urban Sanitary District of	Rural Sanitary District of	Ecclesiastical Parish or District of		
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES In- habited (Or not in- habited (B.)	NAME and Surname of each Person	RELATION to Head of Family	CON- DICTION as to Marriage	AGE last Birthday of Males Females	Rank, Profession, or OCCUPATION	WHERE BORN	II (1) Deaf-and-Dumb (2) Blind (3) Imbecile or (4) Lunatic
4	7, Charlotte St. ("Queen's Arms")	1	Michael Morrison Mary J. Do. Ellen Do. Elizabeth Morrison Ann Fox Catherine Doyle	Head Wife Daus. Mother Sew. Sew.	Mar. Mar. Mar. W. Unm. Unm.	31 29 30 58 28 24	Licensed Victualler Salo; Condev Salo; Shrewsbury Annuitant General Serv. Batmaid	Middlesex; Tolington Salop; Condev Salop; Shrewsbury Lunatic	
5	8, Charlotte St.	1	Lambeit Newton Emma Do. William Do. Henrietta Do. George Bacon Jane Cook James F. Phillips Harriet Do. Sophia White	Head Wife Son Dau. Shopman Sew. Head Wife Sew.	Mar. Mar. Mar. Dau. Unm. Unm. Mar. Unm.	39 36 12 9 19 22 41 29 16	Grocer; (master, employing 2 men) Scholes Do. Grocer's Shopman General Serv. Barker's Clerk General Serv.	Cambridgeshire; Weyton Do.; Longtown Salop; Ludlow Do.; Do. Middlesex; Shoreditch Scotland Yorkshire; Leeds Do.; Bradford Salop; Bridgnorth	
6	9, Do.	1/6							
7	1, Bird Lane	1	William Frampton Anne Do. Thomas Johnson Henry Johnson Emma Do. Jane Palmer Walter Campbell	Head Wife Head Son Wife Apprentice Lodger	Mar. Mar. Wid. Unm. Unm. Unm. Unm.	72 69 68 39 41 18 23	Coach Tuner Coach Tuner Peculiar Grocer Organist Cosey Makers Cosey Makers (Apprentice) Ship Carpenter (out of employ)	Hertfordsh.; Bidden Do.; Tamworth Devon; Honiton Salop; Shrewsbury Middlesex; St Pancras Salop; Ludlow Durham; Sunderland	
8	2, Bird Lane	1							Blind
9	Do.	2/5							
Total of Houses...		4 1/6 2/5	Total of Males and Females...		9	13	End of St. Michael Ecclesiastical District		

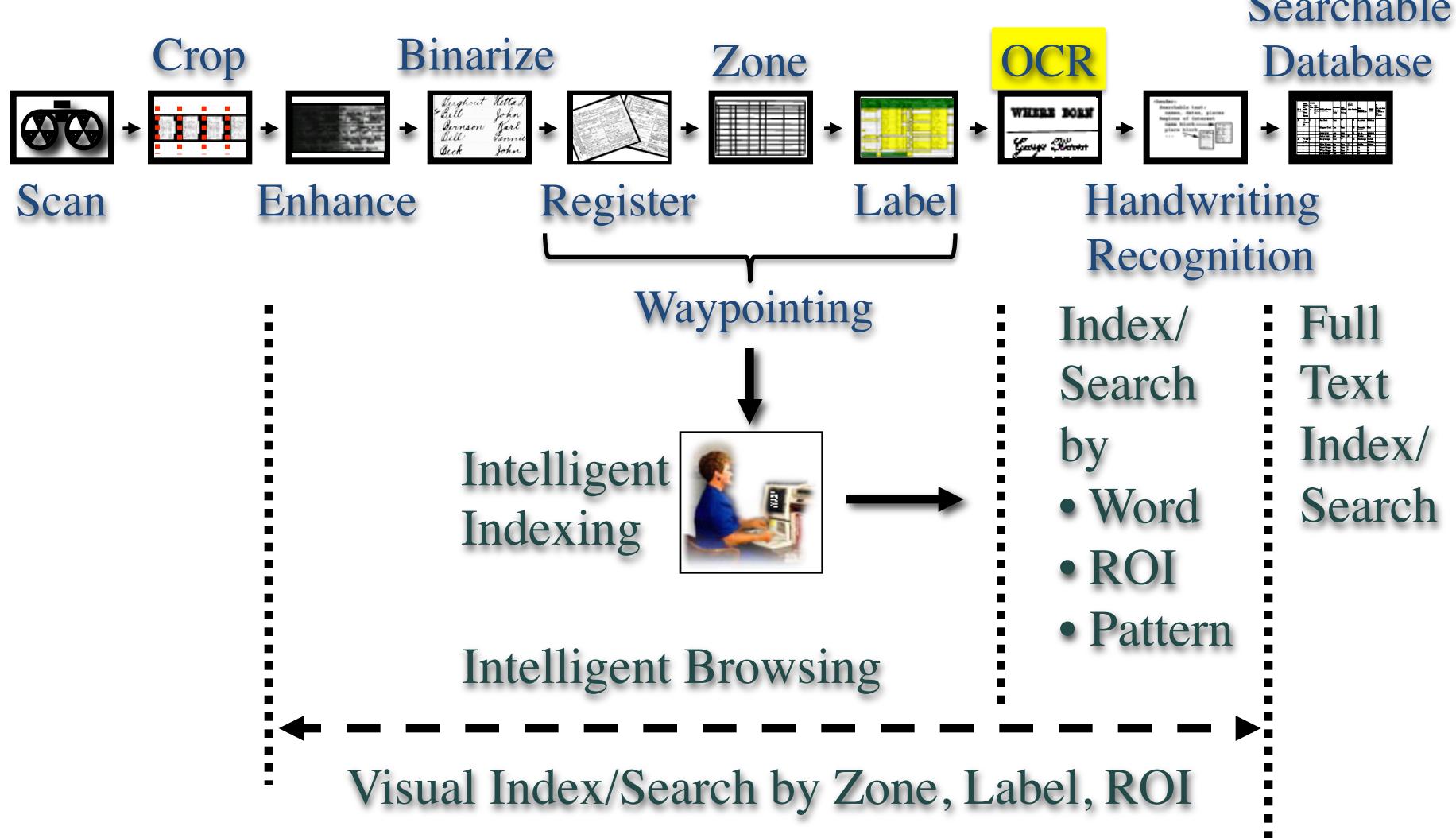
PUBLIC RECORD OFFICE						Reference:-
1	2	3	4	5	6	RG 11/2072.
1				2		

COPYRIGHT PHOTOGRAPH - NOT TO BE REPRODUCED PHOTOGRAPHICALLY WITHOUT PERMISSION OF THE PUBLIC RECORD OFFICE, LONDON



Zoned and Labeled Image

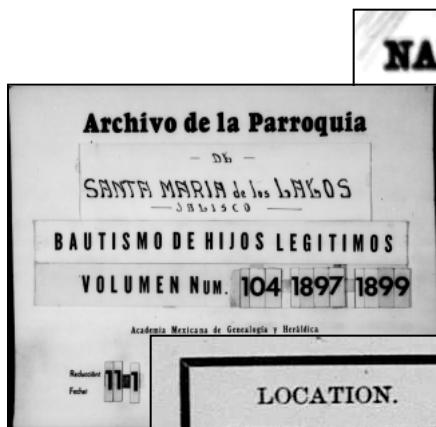
The Digital Microfilm Pipeline





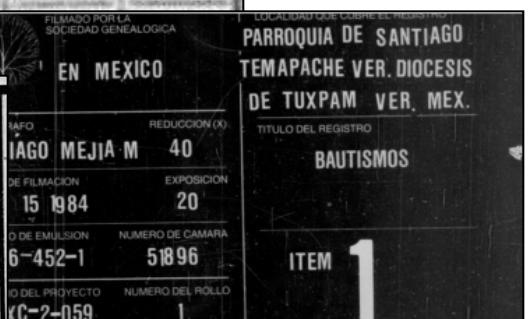
Machine-Printed OCR

A solved problem?



LOCATION.			NAME	RELATION.
IN CITIES.		Number of dwelling-house, in the order of visitation.	of each person whose place of abode on June 1, 1900, was in this family.	Relationship of each person to the head of the family.
Street.	House Number.	Number of family, in the order of visitation.	Enter surname first, then the given name and middle initial, if any. <small>INCLUDE every person living on June 1, 1900. OMIT children born since June 1, 1900.</small> <i>Anthony Lucas family 13901 over from Shrub</i>	
		1	3	4
		2		

PLACE OF ABODE			
Street, town, etc.	House number (in cities or towns)	Number of dwelling houses in order of visitation	Number of family in order of visitation
		1	2



Must ***detect*** Machine Print before recognizing it



Machine vs. Hand Print

- Pinson 2005, 2011

Your marriage record is vital
Be sure the information you give is complete and accurate

PLEASE PRINT – USE BLACK INK

GROOM (MALE)
 1 FIRST NAME OF GROOM **DAVID**
 Zip Code **73203** 3 USUAL RESIDENCE STREET ADDRESS AND CIT
 5 FATHER FULL NAME FIRST MIDDLE LAST
 Home Ph # **621-4403** 6 MOTHER FULL NAME FIRST MIDDLE MAIDEN NA
 Office Ph # **533-4010** **JESSICA STEWART**
 7 FIRST NAME OF BRIDE

BRIDE (FEMALE)

Zip Code **etsy**
 9 USUAL RESIDENCE STREET ADDRESS AND CIT
 Weston
 11 FATHER FULL NAME FIRST MIDDLE LAST

Home Ph # **Jose** 12 MOTHER FULL NAME FIRST MIDDLE MAIDEN N.
 Office Ph # **Mar** **Th.**
 Blood relation hip of groom On what island do
 to bride **NONE** (O h H M K
 L M I K)

FORWARDING ADDRESS
 (After Marriage) **42 VALLEY VE**
CON

SUPPLEMENTARY DATA
 NUMBER OF THIS MARRIAGE
 FIRST SECOND BY DEATH DIVORCE IF PREVIOUSLY MARRIED LAST !
 ETC (SPECIFY) DISSOLUTION OR ANNULLMENT (specify)
 DATE ENDED MONTH / YEAR

25	FIRST	26	—	26b	—
----	--------------	----	---	-----	---

MA
 TO BE F
 (Plea

Your marriage record is vital
Be sure the information you give is complete and accurate

PLEASE PRINT – USE BLACK INK

GROOM (MALE)
 1 FIRST NAME OF GROOM **DAVID**
 Zip Code **73203** 3 USUAL RESIDENCE STREET ADDRESS AND CIT
 5 FATHER FULL NAME FIRST MIDDLE LAST
 Home Ph # **621-4403** 6 MOTHER FULL NAME FIRST MIDDLE MAIDEN NA
 Office Ph # **533-4010** **JESSICA STEWART**
 7 FIRST NAME OF BRIDE

BRIDE (FEMALE)

Zip Code **etsy**
 9 USUAL RESIDENCE STREET ADDRESS AND CIT
 Weston
 11 FATHER FULL NAME FIRST MIDDLE LAST

Home Ph # **Jose** 12 MOTHER FULL NAME FIRST MIDDLE MAIDEN N.
 Office Ph # **Mar** **Th.**
 Blood relation hip of groom On what island do
 to bride **NONE** (O h H M K
 L M I K)

FORWARDING ADDRESS
 (After Marriage) **42 VALLEY VE**
CON

SUPPLEMENTARY DATA
 NUMBER OF THIS MARRIAGE
 FIRST SECOND BY DEATH DIVORCE IF PREVIOUSLY MARRIED LAST !
 ETC (SPECIFY) DISSOLUTION OR ANNULLMENT (specify)
 DATE ENDED MONTH / YEAR

25	FIRST	26	—	26b	—
----	--------------	----	---	-----	---

MA
 TO BE F
 (Plea

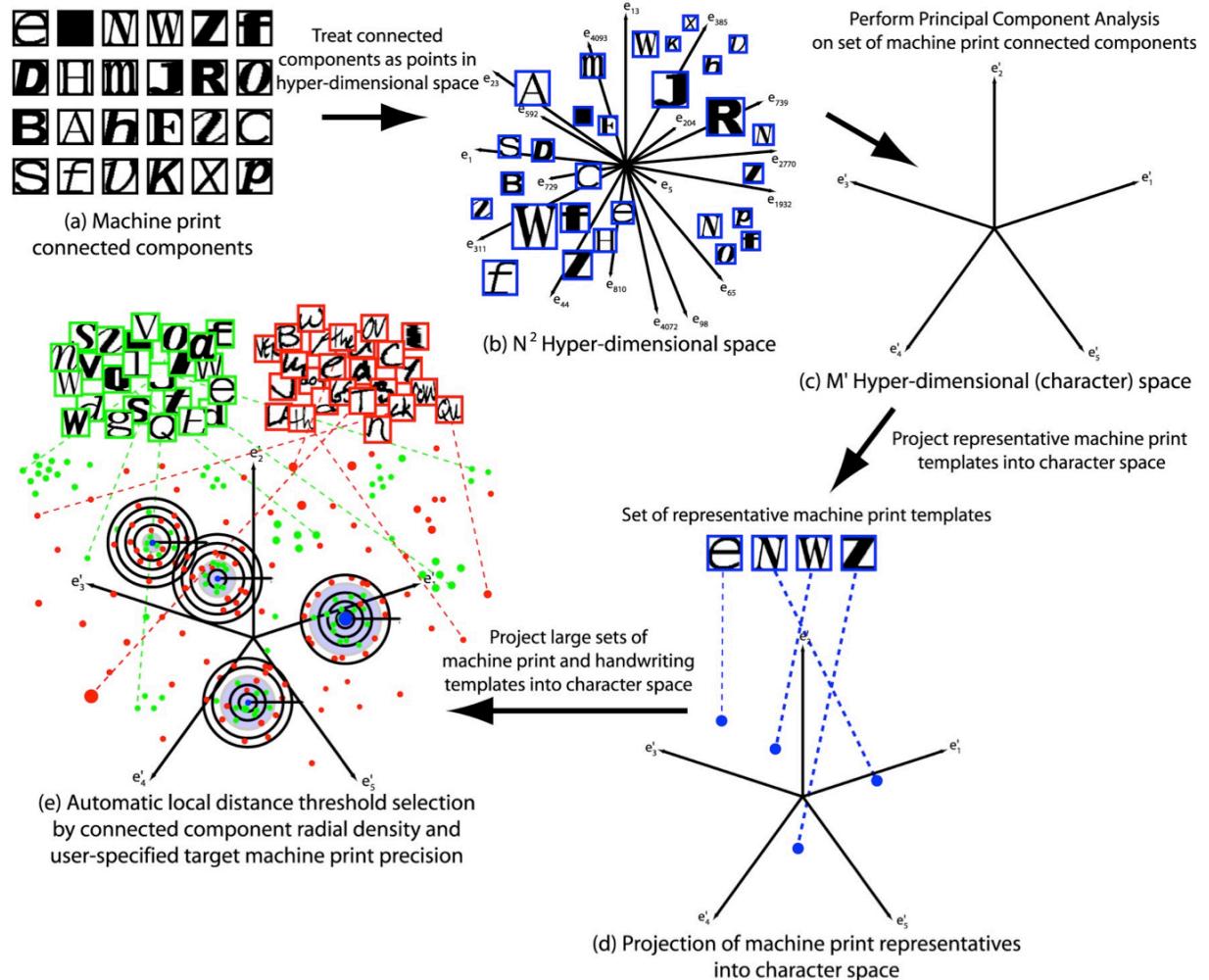
PCA Classification

Machine Print

- 26 fonts
- 273K CCs
- 6000 Templates

Hand Print

- 2100 writers
- ~600K CCs





Precision

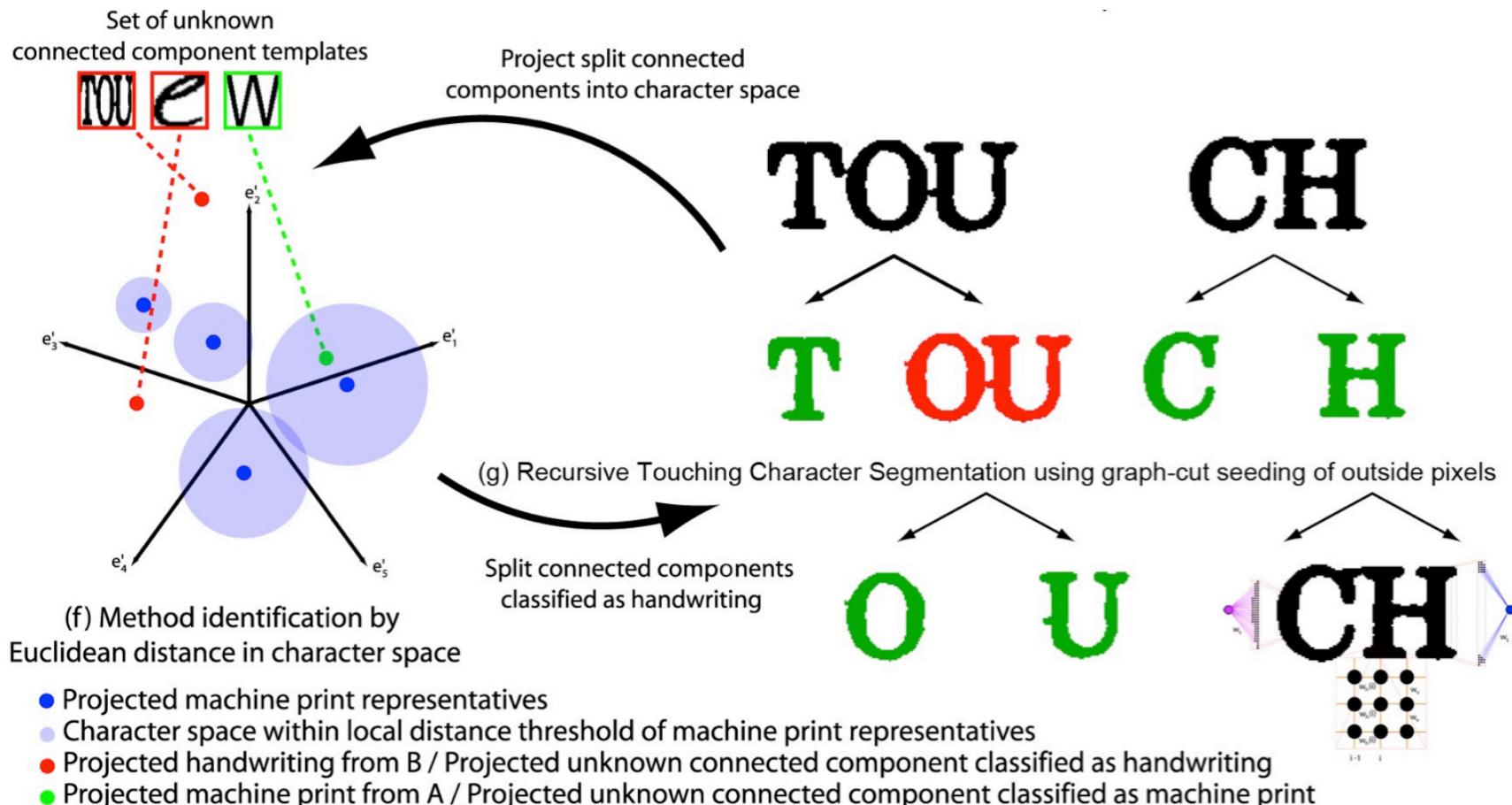
Machine Print – 94%

Hand Print – 89%

Many Errors from Touching characters:

quantitative method is Mean
hod takes note that an original
was interpolated using a magni
However as an enhancement
to areas of the image that have

Split Touching Characters





Machine-Printed OCR – Barzee '96

- Posting Paper on the Web

- First-generation print
- Font-specific OCR
- Grayscale attributes
- Multiple character prototypes for kerned and touching characters
- Lexicon/contextual error correction
- User interaction to recognize problem words

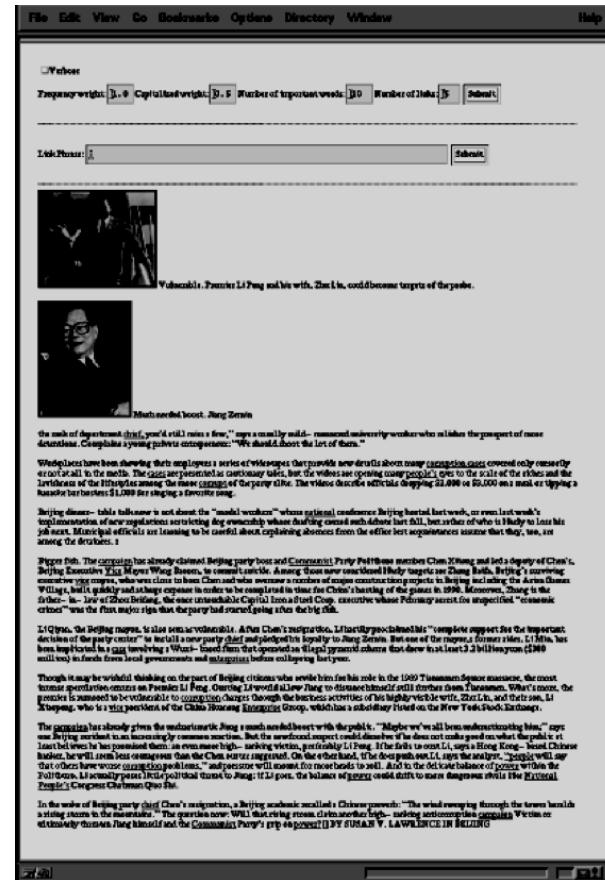
||

99.8% compared to 99.94% (human)

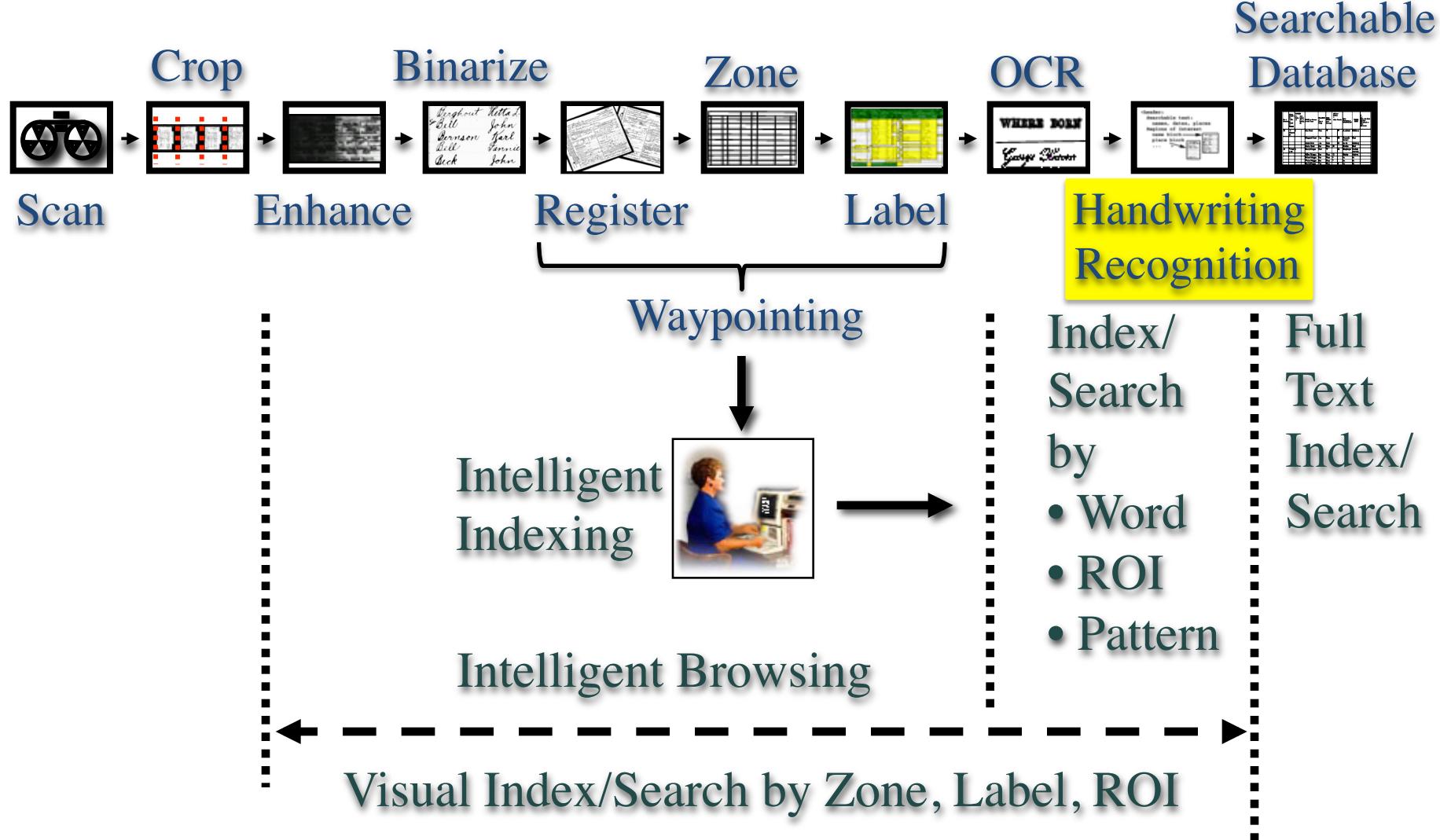
Machine-Printed OCR – Barzee '96

- Posting Paper on the Web

- Paper in -> HTML out
- Own OCR
- Graphics detection
- Captions detection
- Automatically generated links
- 99.8% OCR
- 99.96% - human proof reading



The Digital Microfilm Pipeline





Handwriting Recognition

NOT a solved problem

All current approaches must **segment** words/lines before recognition is attempted

Segmentation in tabular documents – Davis 2015



Patterson Thomas

A red double-headed horizontal arrow is positioned below the name "Patterson Thomas", indicating the segmentation boundary between the two words.

Gymnasium
Patterson Thomas
Barton George Jr.
Lizzy
Galph
Lemoyne
Barton Mariah
Williams Perce
Clift
Lester
Edith

Segmentation in tabular documents



~~Patterson Thomas~~

Labour norms

~~Boston Gazette.~~

Almond names

Bentley [↑] Grace A.

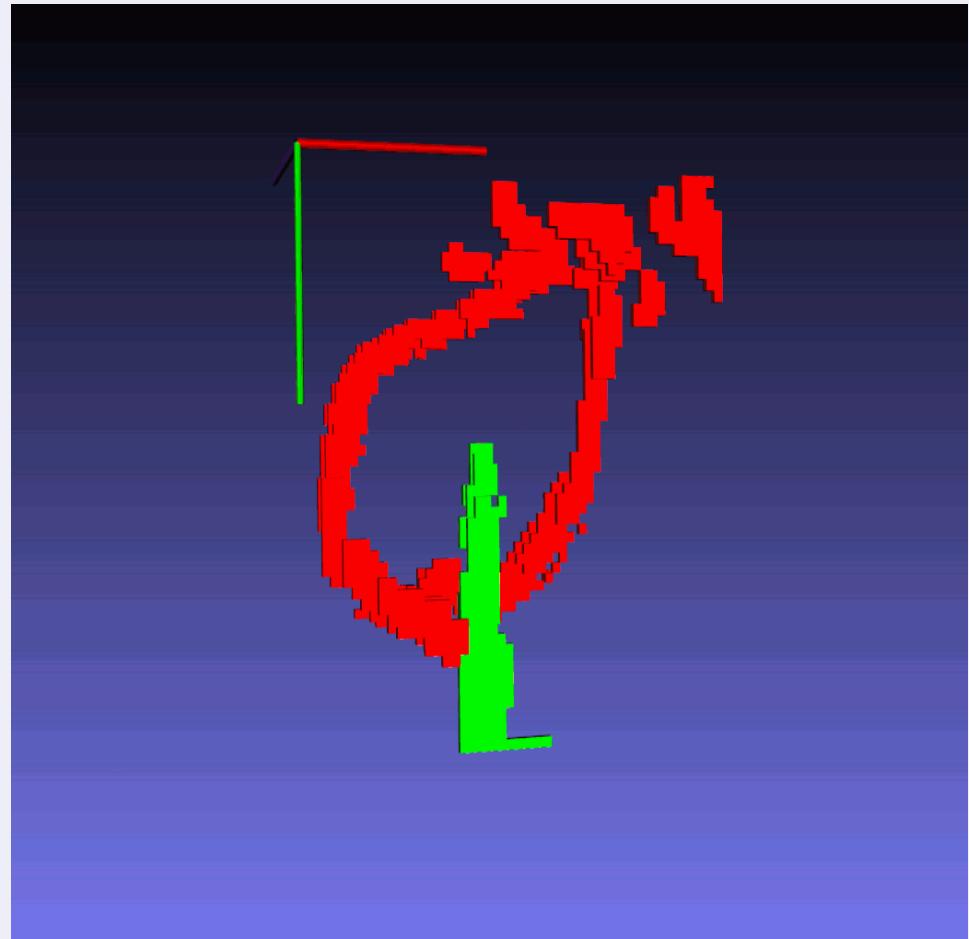
Tiffanie
2017

— Gyne
Patterson Thomas
Fector George A.
— Dijon
— Gallo
— Johnson
Farton Oppenish
Williams Rose
— Gill
— East
— Edith



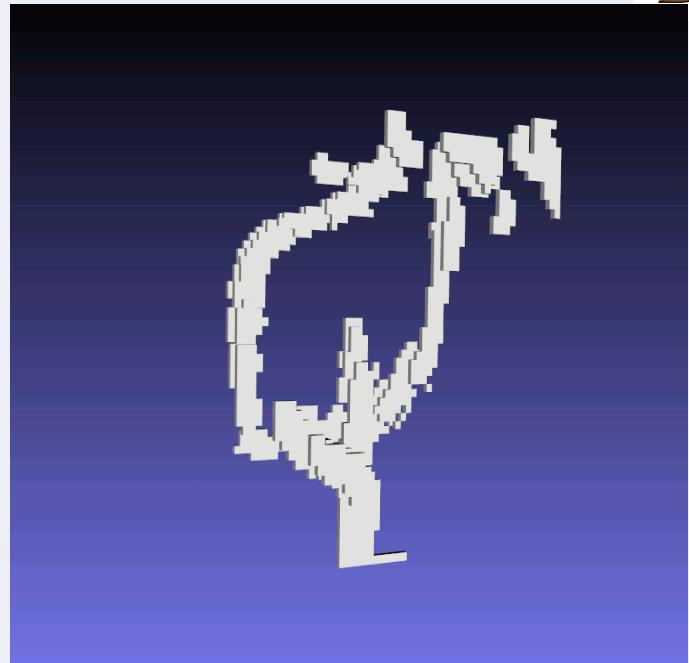
3D cut overview

- Take subsection
- Find slopes
- Project image to 3D using slopes
- **Graph cut**



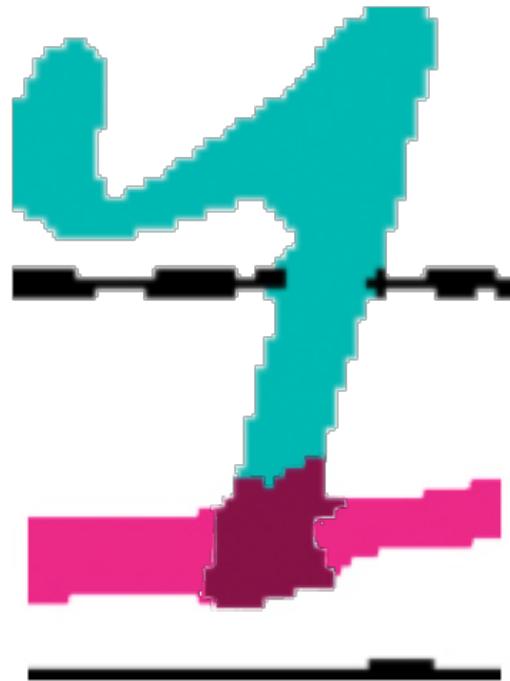
Graph creation

- Pixels are assigned slope values in a fill manner.
 - Pixels can receive multiple assignments.
- 3D distance map computed.
- Source and sink connections based on severing points.





Peal Apart Strokes



So how well does this all work?

Horizontal
segmentation:

- 92.6% accuracy on our application dataset.
- 89.5% accuracy on lines of the ICDAR2013 Handwriting Segmentation Contest

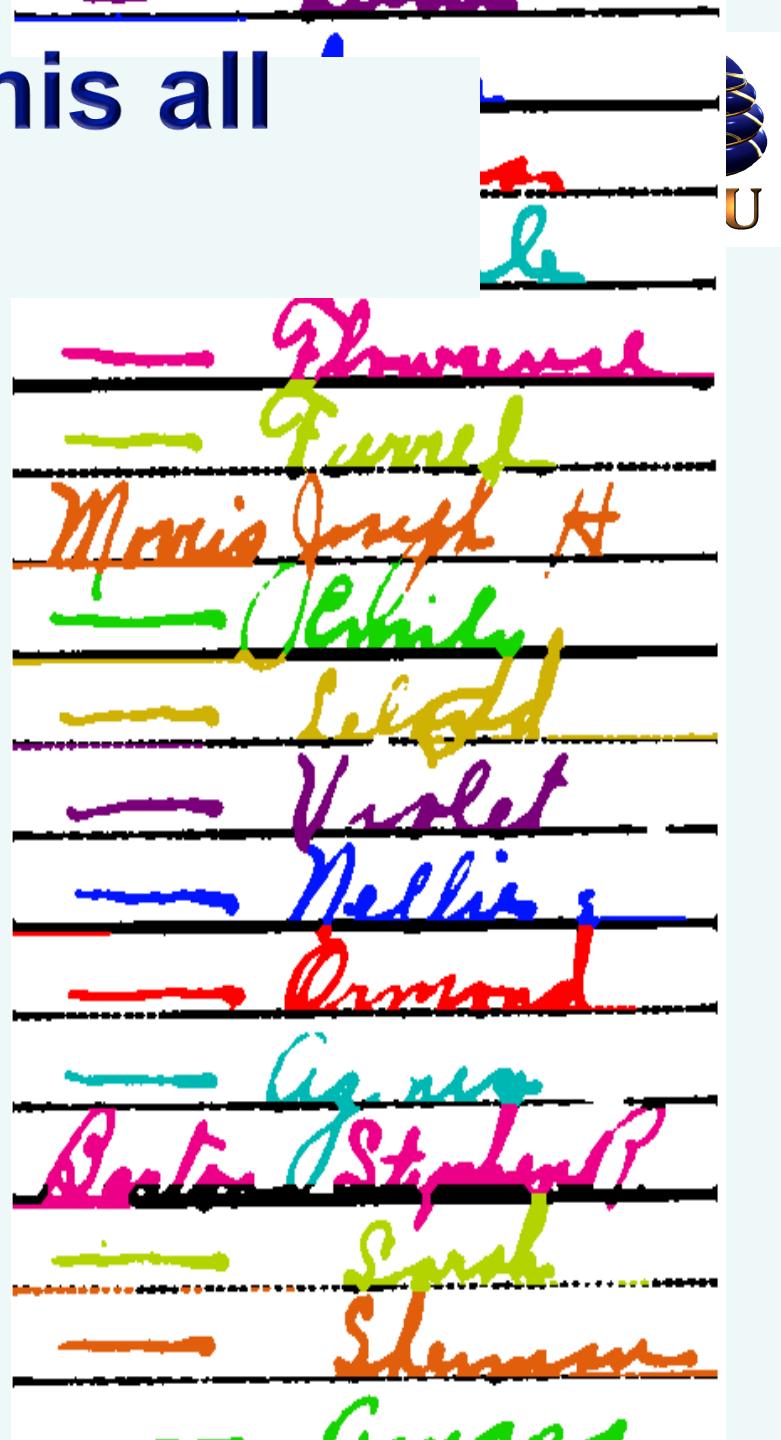


So how well does this all work?

Vertical segmentation

(without 3D cut):

- 80.7% accuracy on our application dataset.





Handwriting Recognition

Handwriting Recognition
by Word Warping
- *Doug Kennard* - 2012



Word Warping

brief

Monday



Handwriting Recognition

- *using Word Warping*

Single author

in vocabulary – 91-92%

Single author

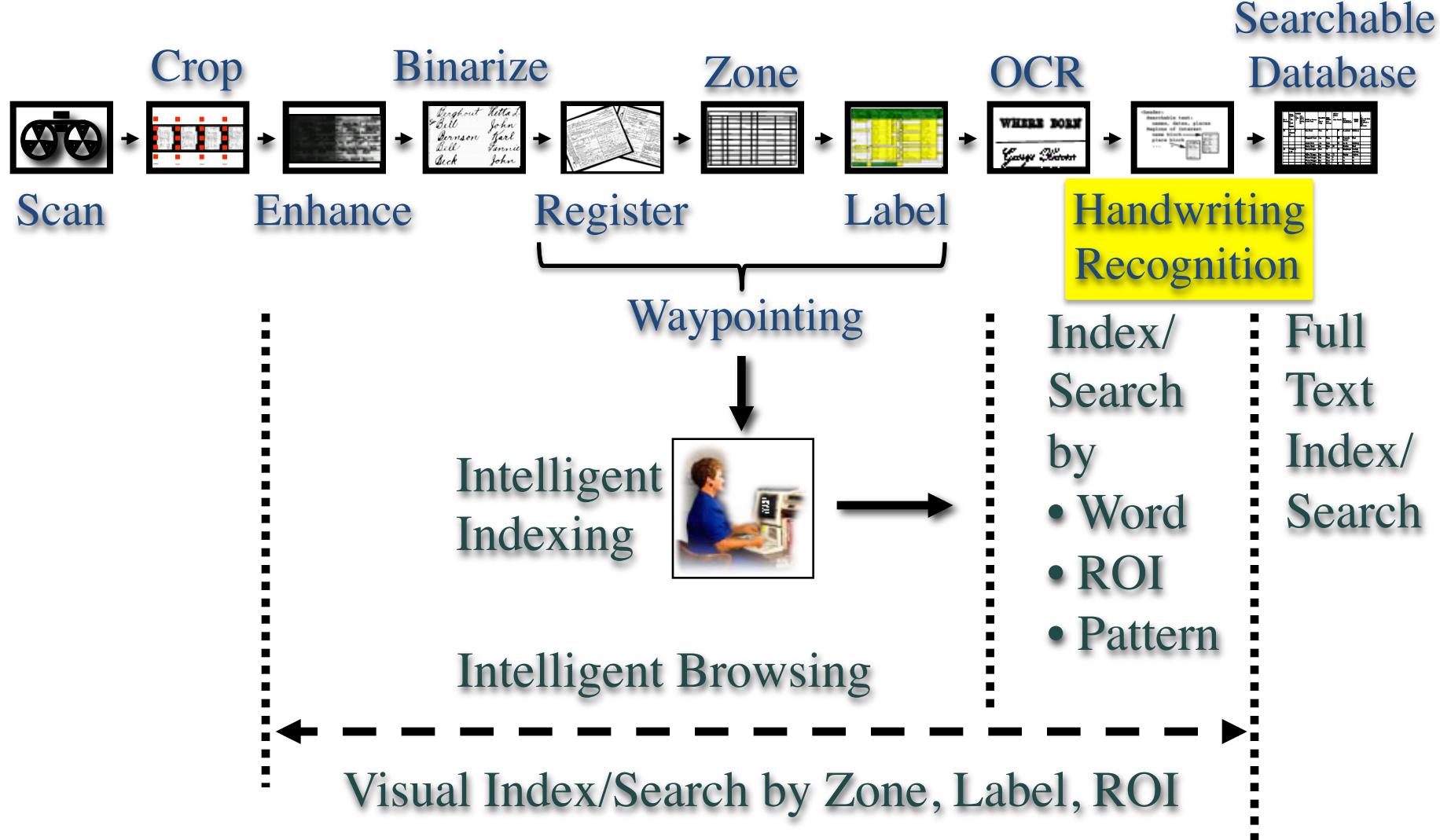
in vocabulary and

out of vocabulary – 69-72%

Multi-author *in vocabulary* – 66%

Multi-author *in+out* of vocab – 56%

The Digital Microfilm Pipeline





Handwriting Recognition with Deep Learning Networks – '16

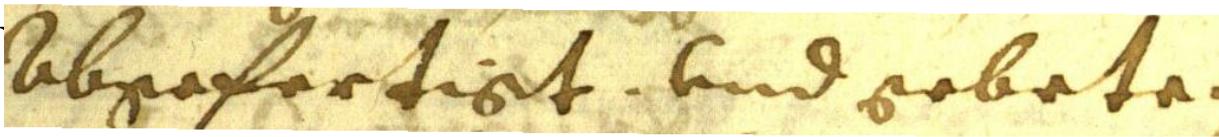
Wigington, Davis, Stewart

- The image shows a piece of aged, yellowish-brown paper with handwritten text in brown ink. The text appears to be in a Gothic script and reads "An die herrn gehaimen Rāth".
- Expert Transcription:
- An die herrn gehaimen Rāth
- Our system:
- An die herrn gehaimen Rāth



Handwriting Recognition with Deep Learning Networks – '16

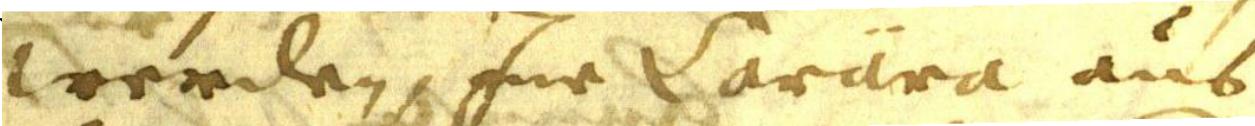
Wigington, Davis, Stewart

- Abgefertigt. vnd gebeten.
- Expert Transcription:
- Abgefertigt. vnd gebeten
- Our system:
- Abgefertigt. vnd gebete■



Handwriting Recognition with Deep Learning Networks – '16

Wigington, Davis, Stewart

- 
- Expert Transcription:
 - werden, Ine Carära aūs
- Our system:
 - werden, Ine Carara aūs

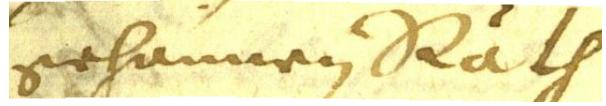
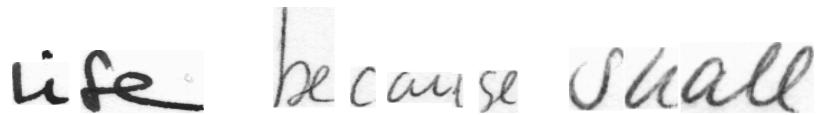


Handwriting Recognition with Deep Learning Networks – '16

Wigington, Davis, Stewart

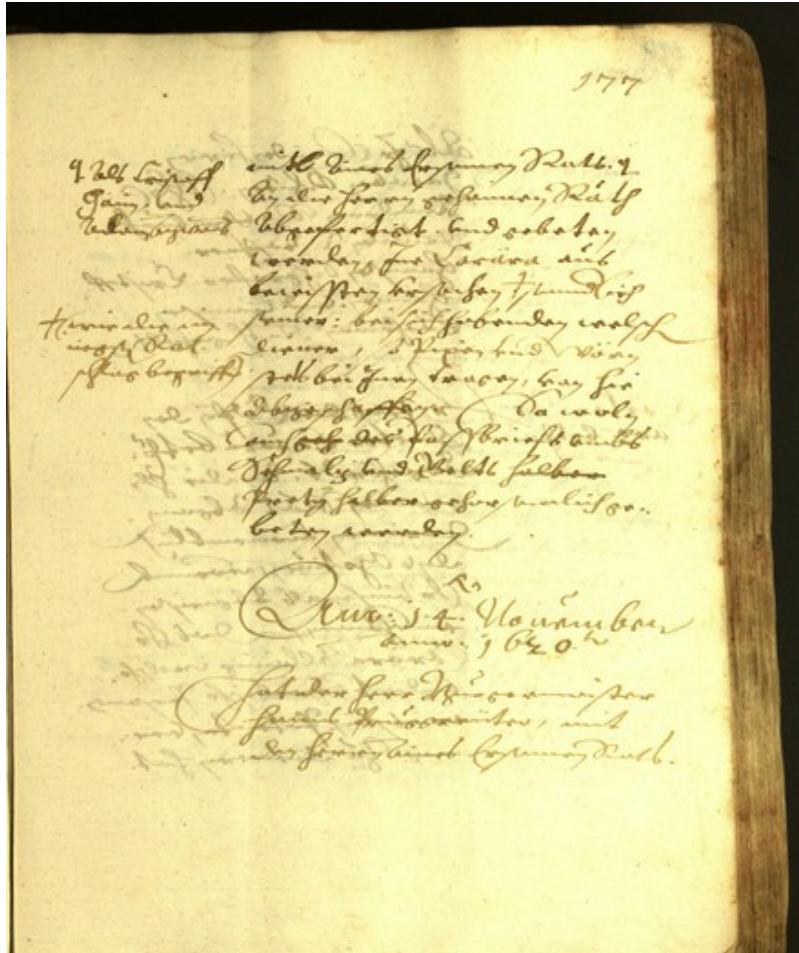
- A photograph of a handwritten document page showing several lines of cursive text in brown ink on aged paper. The text appears to be in a Gothic script, possibly Latin or Old German.
- Expert Transcription:
- seiner: bei sich habenden welsch
- Our system:
- seiner. bei sich habenden welsch

Handwriting Datasets

- German (15th-18th Century) 
- George Washington 
- Smith Journals 
- IAM (Multi-Author) 
- MNIST (Numbers) 



German



- 1470 - 1805
- Bleed Through
- Overlapping strokes
- 95% Character Accuracy
- 80.3% Word Accuracy



George Washington

Single Writer

- 99.3% Character Accuracy
- 98.4% Word Accuracy

Word Examples





Smith Journals

Single Writer

- 98.3% Character Accuracy
- 97% Word Accuracy

Word Examples

meabo mr me then
S.L. keep got funeral
company of attended
for retired
eve music 18
the read Elders
July Duranti
practised of salaafule Tuesday



IAM Handwriting Dataset

657 Writers

- 97.2% Character

Accuracy

- 95.03% Word

Accuracy

Word Examples

command

actions

descended

someone

common

succed succeeded

Powell

Government
in dominance

administration

SALO

different

rival

Government

Cler

In

ambassador

condemnation

commended

Alliances

National

central charge

reverer

Whitehall

called

Yousc

children



MNIST Handwritten Digits

500 Writers

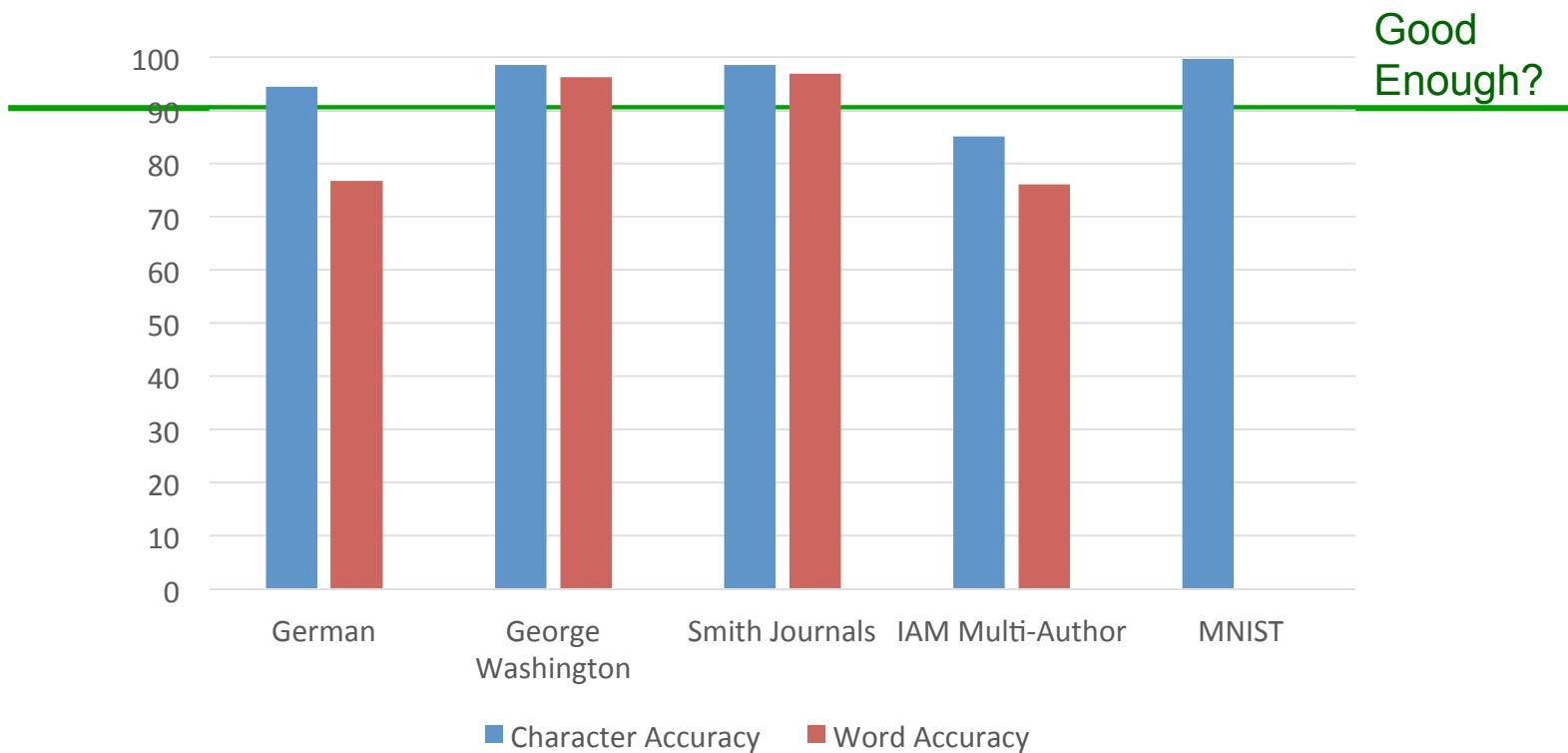
- 99.62% Character
Accuracy

Examples





Handwriting Datasets





But what about

Sehr lieben und begütigen auf dem einen
gesehenen für die Freigemeinde geistl.
und geistlichen Begeisterung
Ihr großer
E.

also die feldwirksame Zygiamen- und Salicis-
Stimmen, gäbe man hundt gäbe also ein
Haus gegen vierhundert Pfund. Wenn man
Längen und Breite, das am ehesten ist,
aufzugeben an Baumwollstoffe. Es ist
Sicher gäbe es geringer manch mein Vor-
handen ~~ist~~ ^{ist}. Also der Bau ist
der gestalt so breit und lüpferig, da er
nicht kann fürstlichkeiten gebraucht werden.
Die Stoffe, abgezogenen bezüglich bestimmt
Rathen gegen aufzunehmen, so wie die
Von Habsburgs aufzunehmen gäbe, bestigt
es nicht mit wissung gegen bestigt, wenn
dann aber nicht, so kann man's leicht
abholen.

également au Comité l'assurance des perspectives du commun
fort avantageuses. (Ancien nom de l'Amicale Féminine)

Sous couvert d'une grande discrétion, j'ai été dans l'ordre
arrivé pour la conférence avec des personnes très
connaissantes et un avis clair. J'y ai été assez tôt, mais je devais assister à
la réunion d'une femme responsable, aussi
longtemps qu'il y ait lieu il a pris la parole, il a déclaré en
ces termes : "Je vous conseille que nous restions
"pour le Comité de l'Amicale de la Ville que je suis
"l'Amicale, car je n'accepterais aucun programme
"en dehors des femmes" ; mais je dois dire que ma
"conscience fait faire mes aspirations. Aujourd'hui
"j'aimerais de plus que ce discours arrive à l'ordre du jour de
"mes conseils que m'impose ma volonté de longue date
"de collaborer avec lui" de conduire ensemble l'Amicale
"avec les compagnes qui sont prêtes à se charger de
"l'autorité de l'Amicale que nous manquons et sans contradiction.
"Si les organisations avec nous n'accepteraient pas, malgré
"cette grande volonté assurée depuis plusieurs mois,
"elles auraient application pour que nous nous puissions
"accepter comme élision dans le accord avec cette
"organisation. C'est à moins que l'on trouve un moyen de convaincre
"cette organisation que nous ne pouvons pas d'y joindre la main,
"pour suivre nos idées libres" —

"Assurez-vous si l'Amicale peut être déclarée sur une
"base d'un bon fonctionnement et de deux fois
"par..."

Souscrivent que dans l'acte à la fin, où le
cessionnaire a déploré de leur omission, comme
que leur absence d'abord, il approuve par le tout
de l'acte, et plus par l'assent des propriétaires de la localité,
que ces élections furent faites le 18 mars ou le 10 mai de l'an
passé, pour le Commerce des marchés, en une obligation
pour le budget du Comptoir, de verser le 2 millions de francs

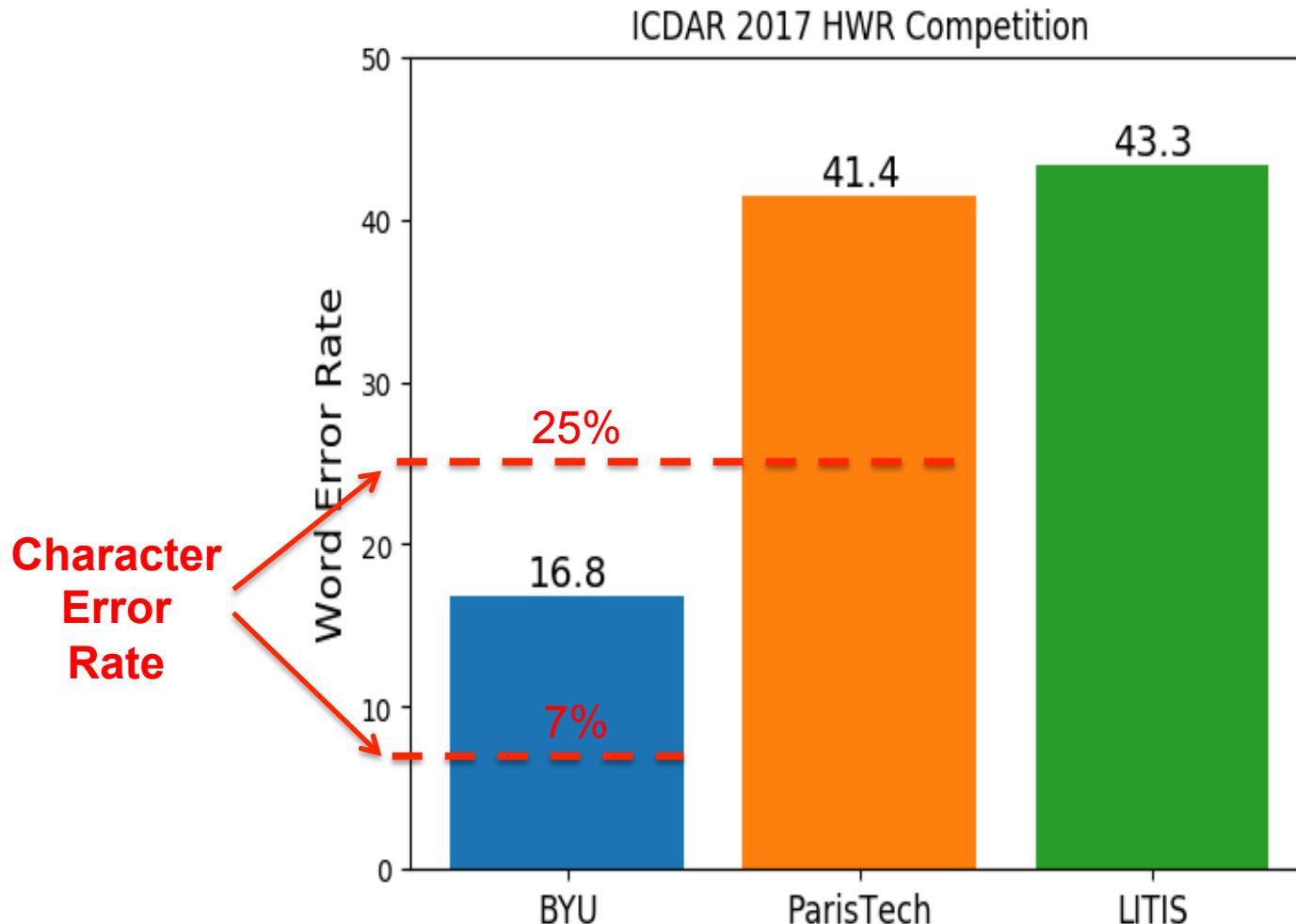


The READ Dataset

- ICDAR 2017 Competition

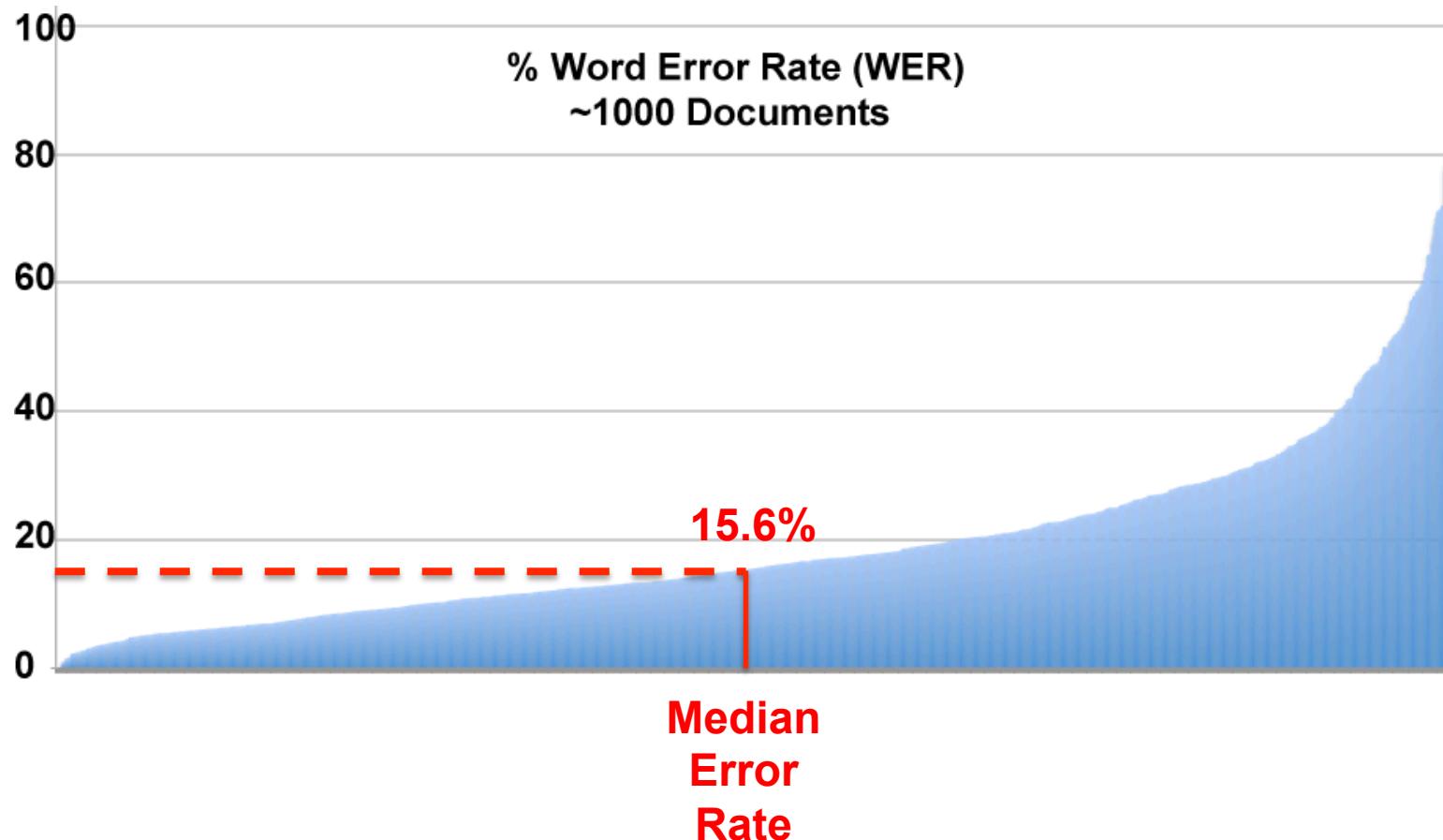
- Multiple languages
 - *German*
 - *French*
 - *English*
- No word-level segmentations
- 50 training images with line-level segmentations and transcriptions
- 10,000 training images with only transcriptions
- 100-200 year old open format letters

Handwriting Recognition:

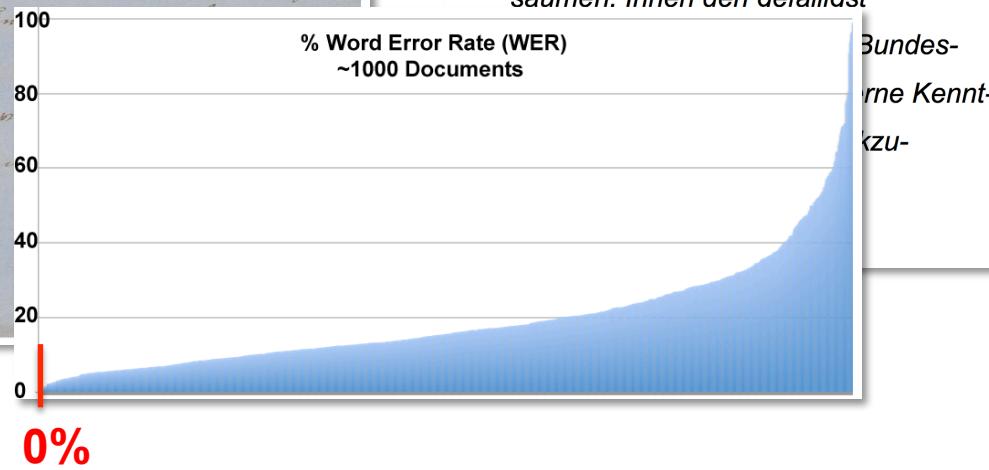
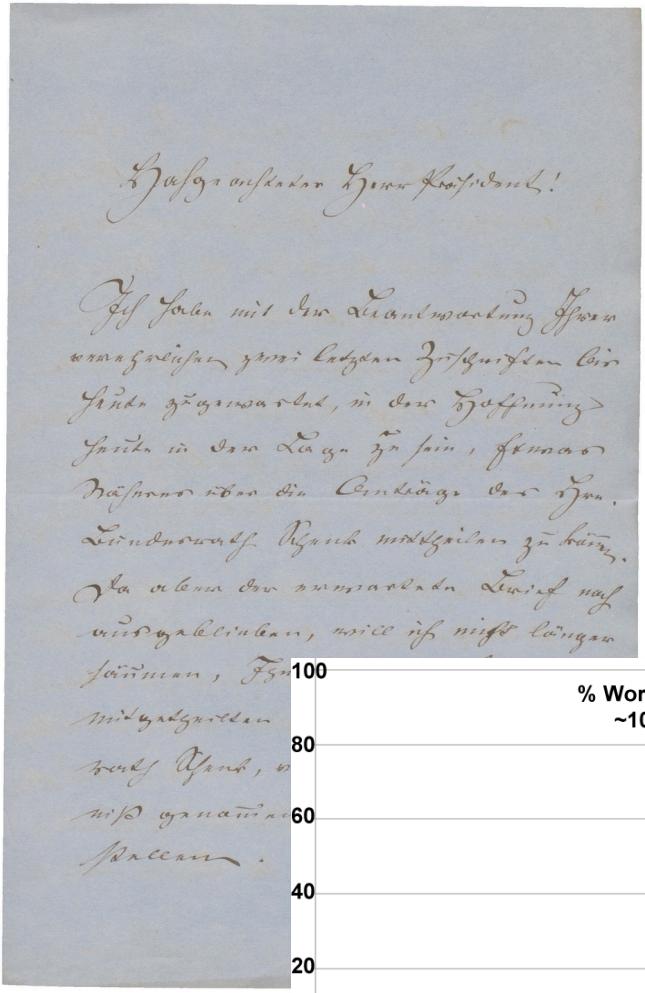




Handwriting Recognition:



0 Errors German Handwriting Recognition



Automated Transcription

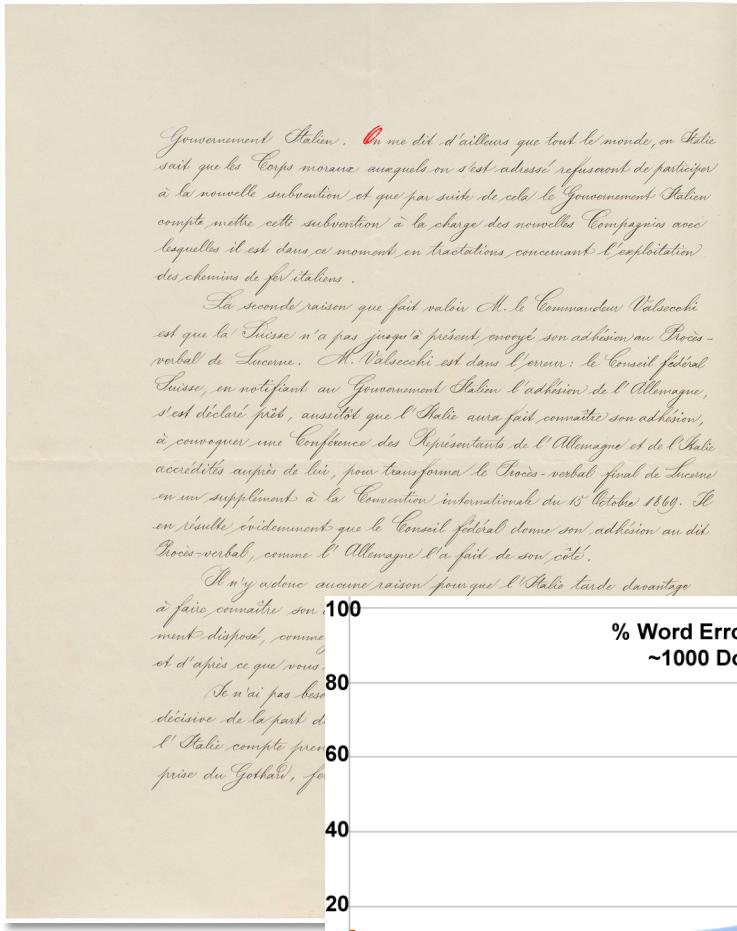
Hochgeachteter Herr Präsident!

Ich habe mit der Beantwortung Ihrer
verehrlichen zwei letzten Zuschriften bis
heute zugewartet, in der Hoffnung
heute in der Lage zu sein, Etwas
Näheres über die Anträge des Hrn.
Bundesrath Schenk mittheilen zu können.
Da aber der erwartete Brief noch
ausgeblieben, will ich nicht länger
säumen. Ihnen den aefälliaſt

Bundes-
reine Kennt-
kzu-

French Handwriting Recognition

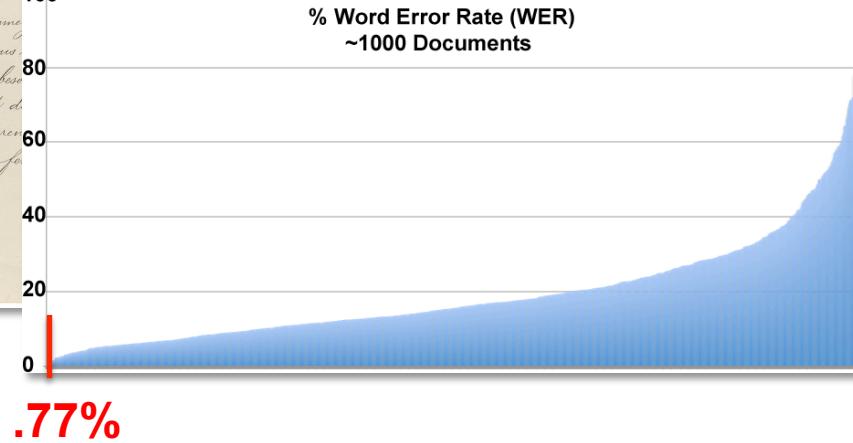
0.1% CER .77% WER



La seconde raison que fait valoir M. le Commandeur Valsecchi est que la Suisse n'a pas jusqu'à présent envoyé son adhésion au Procès-verbal de Lucerne. M. Valsecchi est dans l'erreur: le Conseil fédéral Suisse, en notifiant au Gouvernement Italien l'adhésion de l'Allemagne, s'est déclaré prêt, aussitôt que l'Italie aura fait connaître son adhésion, à convoquer une Conférence des Représentants de l'Allemagne et de l'Italie accrédités auprès de lui, pour transformer le Procès-verbal final de Lucerne en un supplément à la Convention internationale du 15 Octobre 1869. Il en résulte évidemment que le Conseil fédéral donne son adhésion au dit Procès-verbal, comme l'Allemagne l'a fait de son côté.

Il n'y a donc aucune raison pour que l'Italie tarde davantage à faire connaître son adhésion, comme et d'après ce que vous.

Je n'ai pas less decisive de la part de l'Italie, compte prise du Gotthard, fo



Automated Transcription

Gouvernement Italien. En me dit d'ailleurs que tout le monde, en Italie sait que les Corps moraux auxquels on s'est adressé refuseront de participer à la nouvelle subvention et que par suite de cela le Gouvernement Italien compte mettre cette subvention à la charge des nouvelles Compagnies avec lesquelles il est dans ce moment en tractations concernant l'exploitation des chemins de fer italiens.

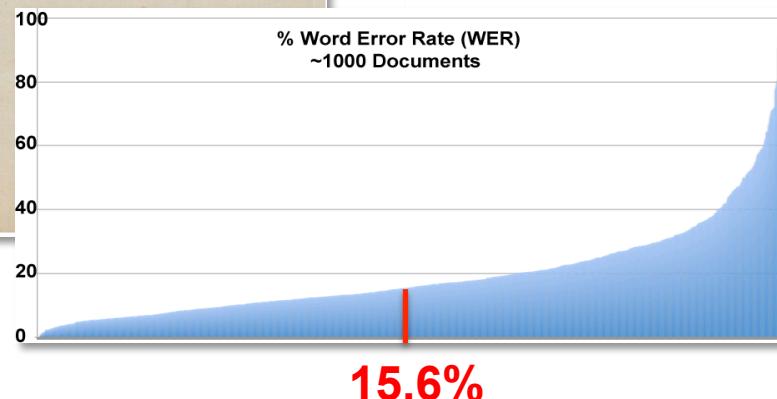
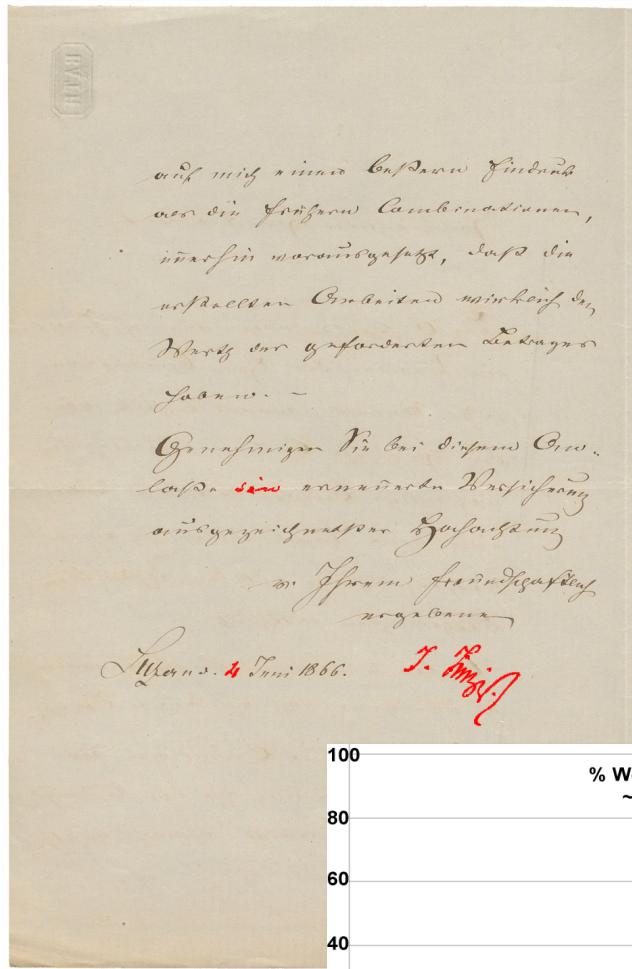
La seconde raison que fait valoir M. le Commandeur Valsecchi est que la Suisse n'a pas jusqu'à présent envoyé son adhésion au Procès-verbal de Lucerne. M. Valsecchi est dans l'erreur: le Conseil fédéral Suisse, en notifiant au Gouvernement Italien l'adhésion de l'Allemagne, s'est déclaré prêt, aussitôt que l'Italie aura fait connaître son adhésion, à convoquer une Conférence des Représentants de l'Allemagne et de l'Italie accrédités auprès de lui, pour transformer le Procès-verbal final de Lucerne en un supplément à la Convention internationale du 15 Octobre 1869. Il en résulte évidemment que le Conseil fédéral donne son adhésion au dit Procès-verbal, comme l'Allemagne l'a fait de son côté.

Il n'y a donc aucune raison pour que l'Italie tarde davantage

Ministère est favorable. ès tout ce qui s'est passé mantes et maintes fois. claration prompte et concernant l'attitude que réorganisation de l'entre- a meilleure impression.

German Handwriting Recognition

5.8% CER 15.6% WER



Automated Transcription

auf mich einen beßern Eindruck
als die früheren Combinationen,
immerhin vorausgesetzt, daß die
erstellten Arbeiten wirklich den
Werth des geforderten Betrages
haben. —

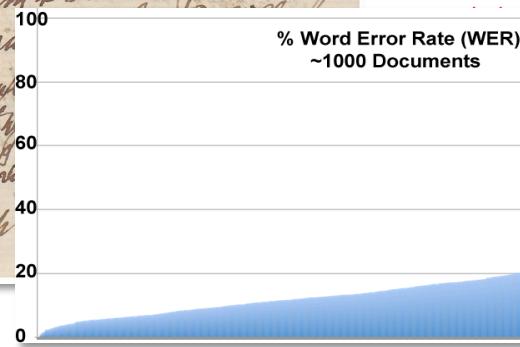
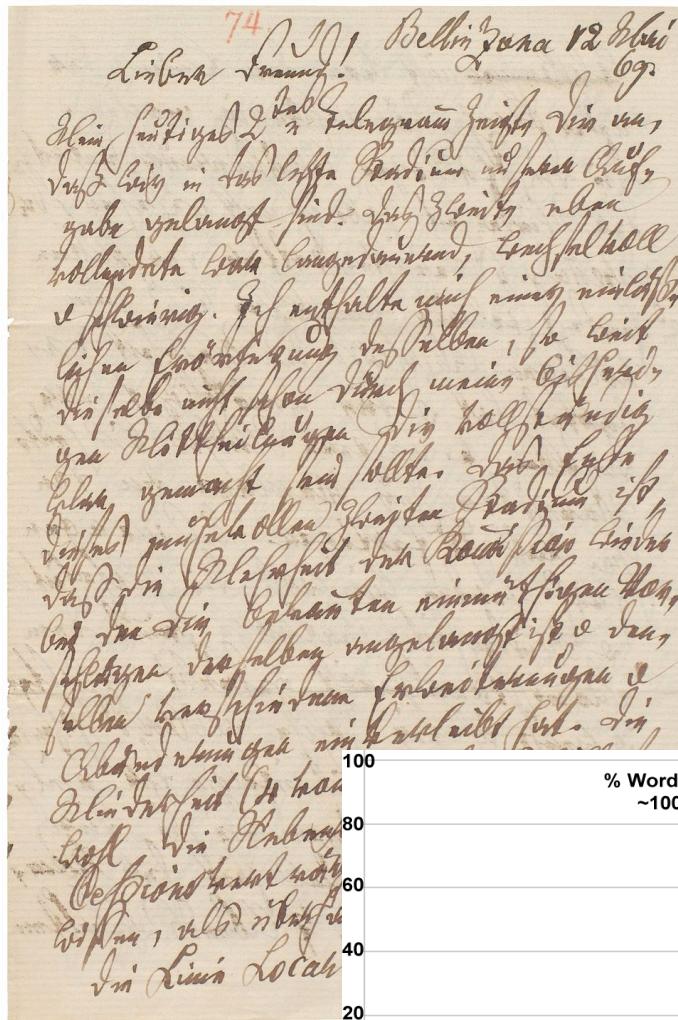
Genehmigen Sie bei diesem An-
laße den erneuerte Versicherung
ausgezeichnetster Hochachtung

v. Ihrem freundschaftlich
ergebenen

Luzern d. 20 Juni 1866. J. Zingg.

German Handwriting Recognition

49.25% CER 71.93% WER



Automated Transcription

G. Bellinzona 12 Mai
69.

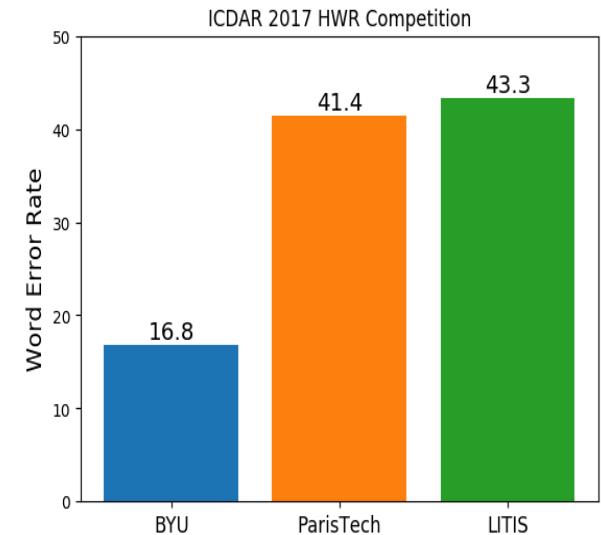
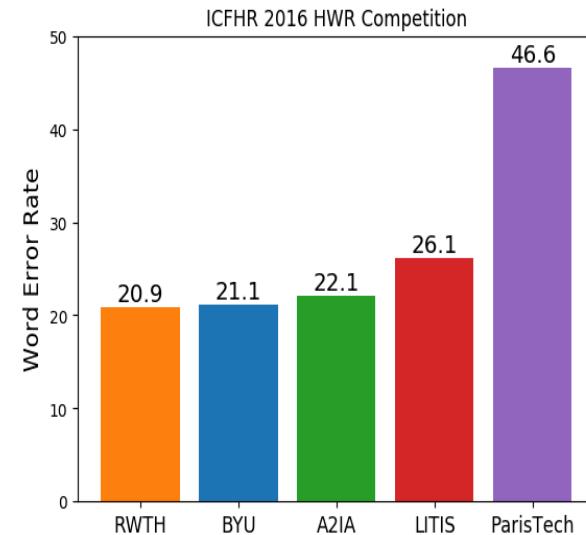
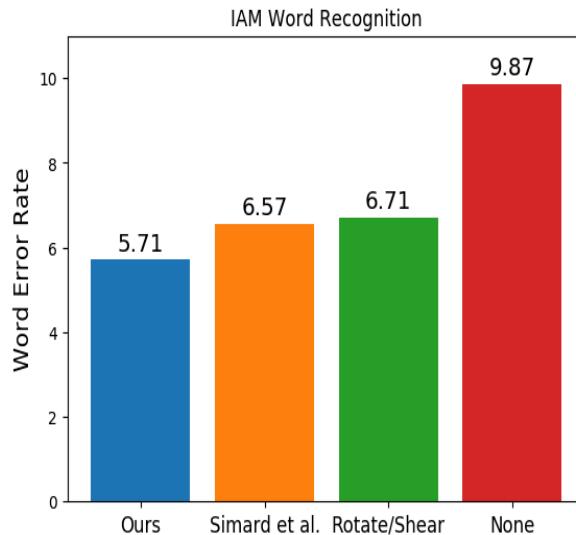
Lieber Freund!
A. Stelle mit
Rhein heutiges Telegramm zeigte Dir an-
bei diesem Anlass unser Stich-
daß wir in das letzte Stadium unserer Auf-
gabe gelangt sind. Das zweite eben
vollendete war langedauernd, wechselvoll
& schwierig. Ich enthalte mich einer einläß-
sich nicht unterung desselben, so weit
lichen Erörterungen in Bem
willf. nicht schon durch meine Aufe.
der unterblieben bin vollständig
Mittheilungen zeigen zu beste
ten gemacht sein sollte. Dann für
klar ist den zweiten Stadium, ist
daß unser allen zweiten Teßin wieder
selbst der Sonnelles der
& die Mehrheit der Sie sich
von den die Erwarten einmüthigen Vor-
schlägen denselben angelangt ist & den-
bleiben. Erweiterungen &
selben verschiedene einverleibt hat. Die
Abänderungen ein kann man die
so befinden
in den

ahnens insbesondere

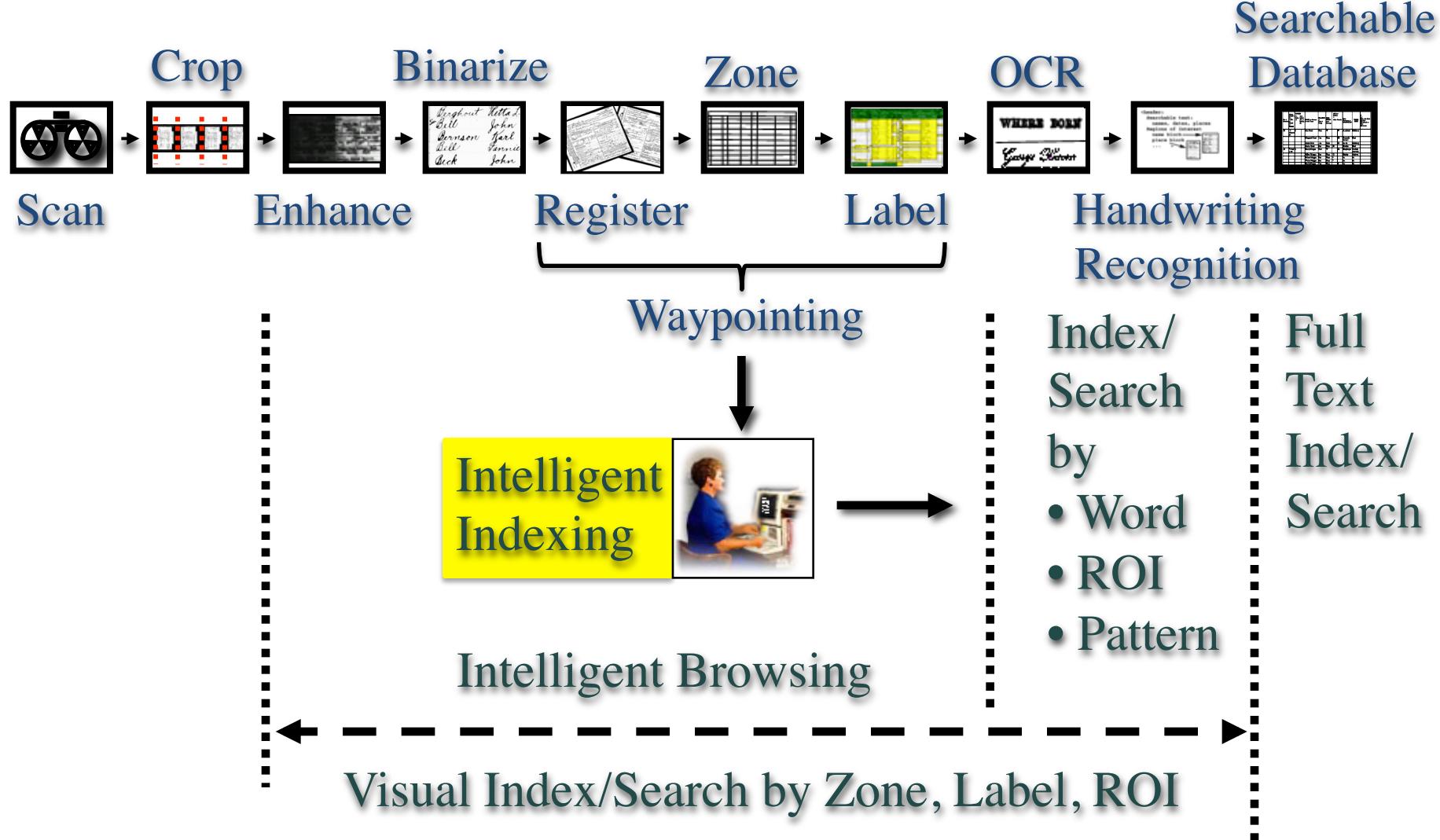
71.93%

Handwriting Recognition:

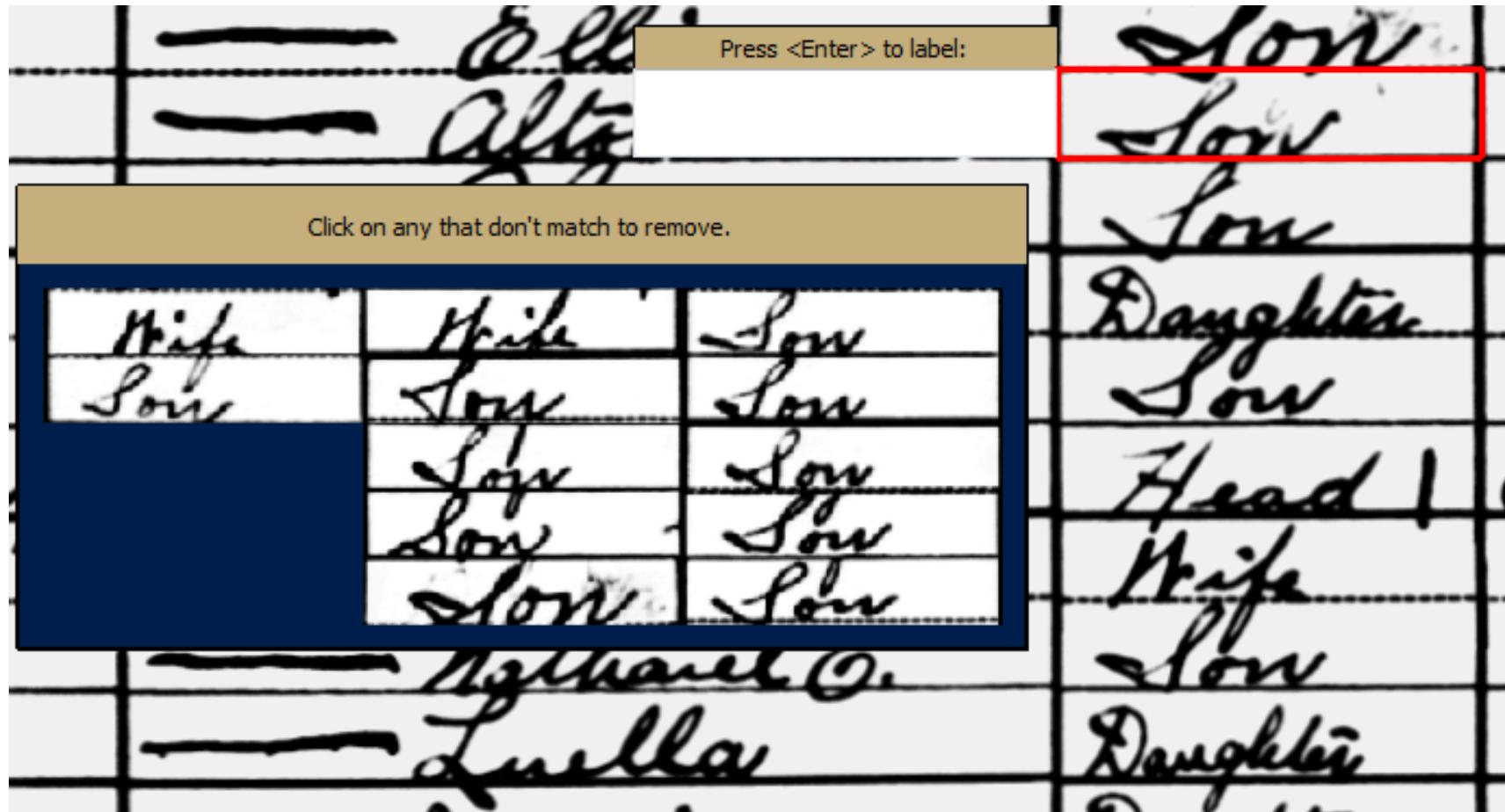
- “Data Augmentation for Recognition of Handwritten Words and Lines using a CNN-LSTM Network”, ICDAR 2017
- Competitions: ICFHR 2016 and ICDAR 2017



The Digital Microfilm Pipeline



Recommendation System





Intelligent Indexing

- *Robert Clawson 2014*

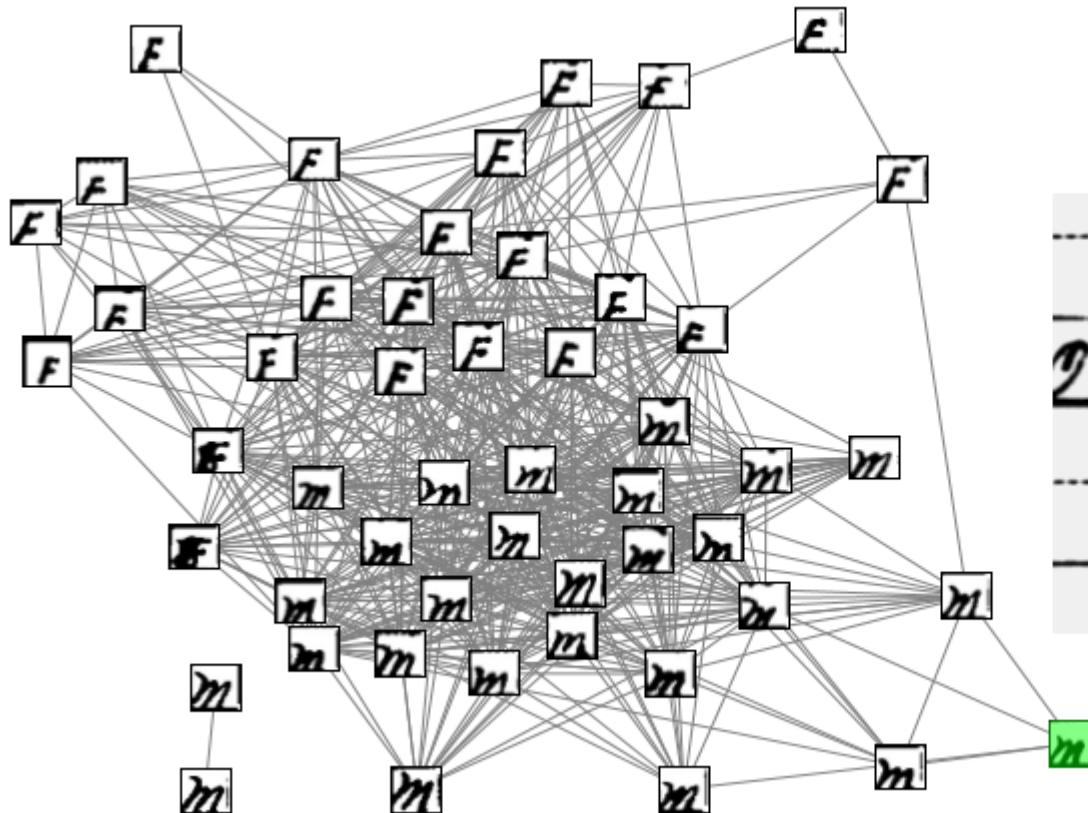
[Intelligent Indexing Video](#)



Intelligent Indexing

Enter names here, then the given name and middle initial, if any.		Relationship of the person to the head of family.		Name entered	If married,	Date of birth	Age at last day.	Year of birth then to the age of death	Relationship of death	If married	Year of death	Entered in Index, 1-10	Entered Index	Entered Index
		T	S	O	R	10	11	12	13	14	15	16	17	18
Johnson	I	Head	I	O F M W 42 M	1891 na	1904						yes	yes	
Joan M		wife		F W 38 M								no	yes	
Georgia		daughter		F W 14 S								yes	yes	yes
Sylvia		daughter		F W 12 S								yes	yes	yes
John		daughter		F W 9 S								yes		
Marie		daughter		F W 1½ S								yes		
Noble Sarah C		Head	I	O F F W 6½ WD 1812 na	1864							yes	no	
Peterson Joseph		Head	I	O F M W 43 M								yes	yes	
Laura G		wife		F W 38 M								yes	yes	
Merrill		son		M W 16 S								yes	yes	yes
Laura		daughter		F W 14 S								yes	yes	yes
Marie		daughter		F W 19 S								yes	yes	yes
Fra		daughter		F W 9 S								yes		

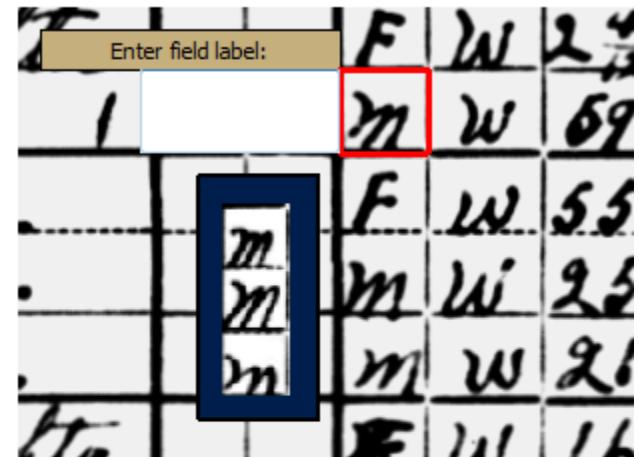
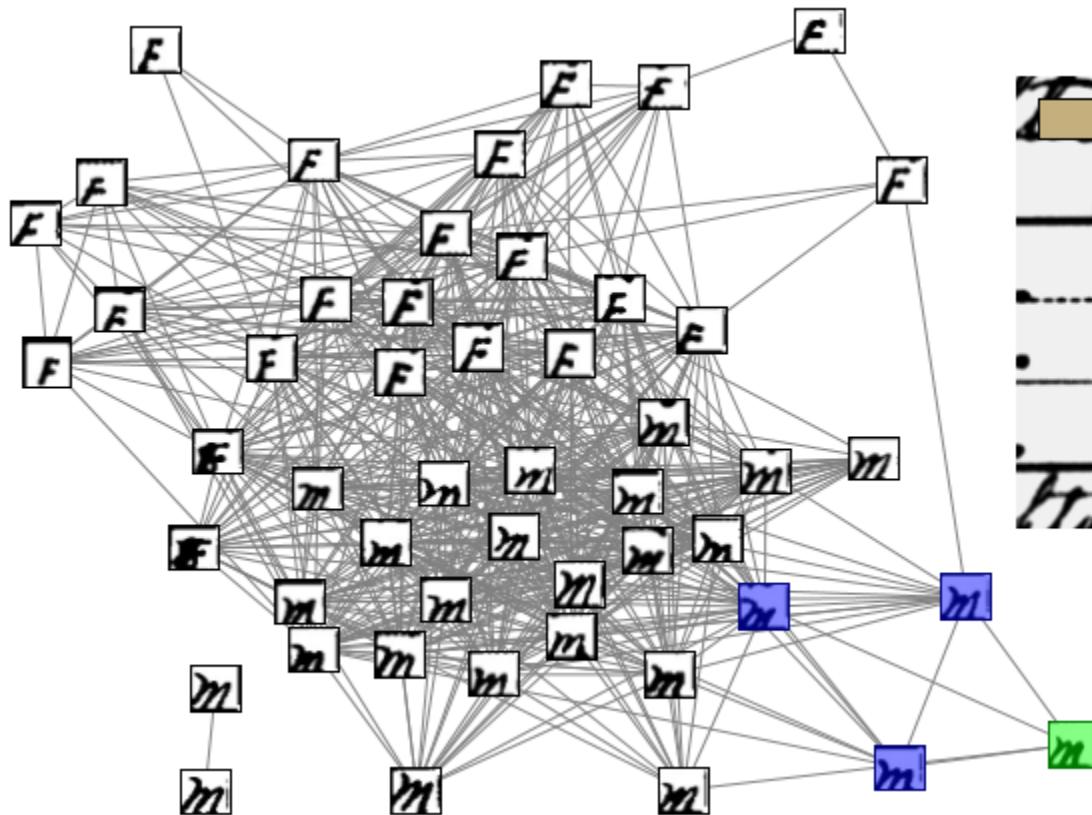
Recommendation System



	m	w	2	s
F	w	2	s	
F	m	w	69	n
F	w	55	n	
m	w	25	s	
m	w	21	s	

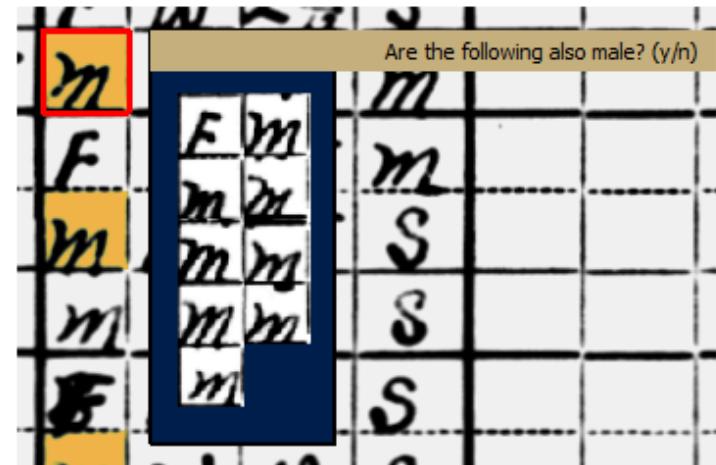
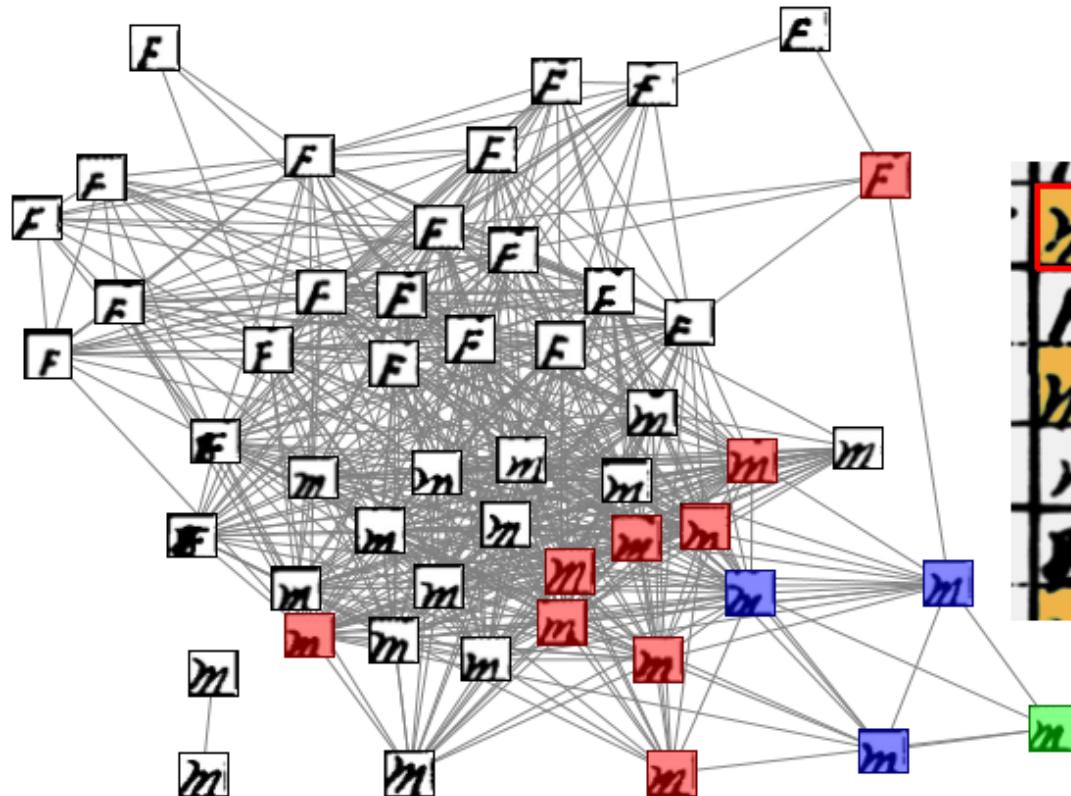
Leveraging the Label

Recommendation System



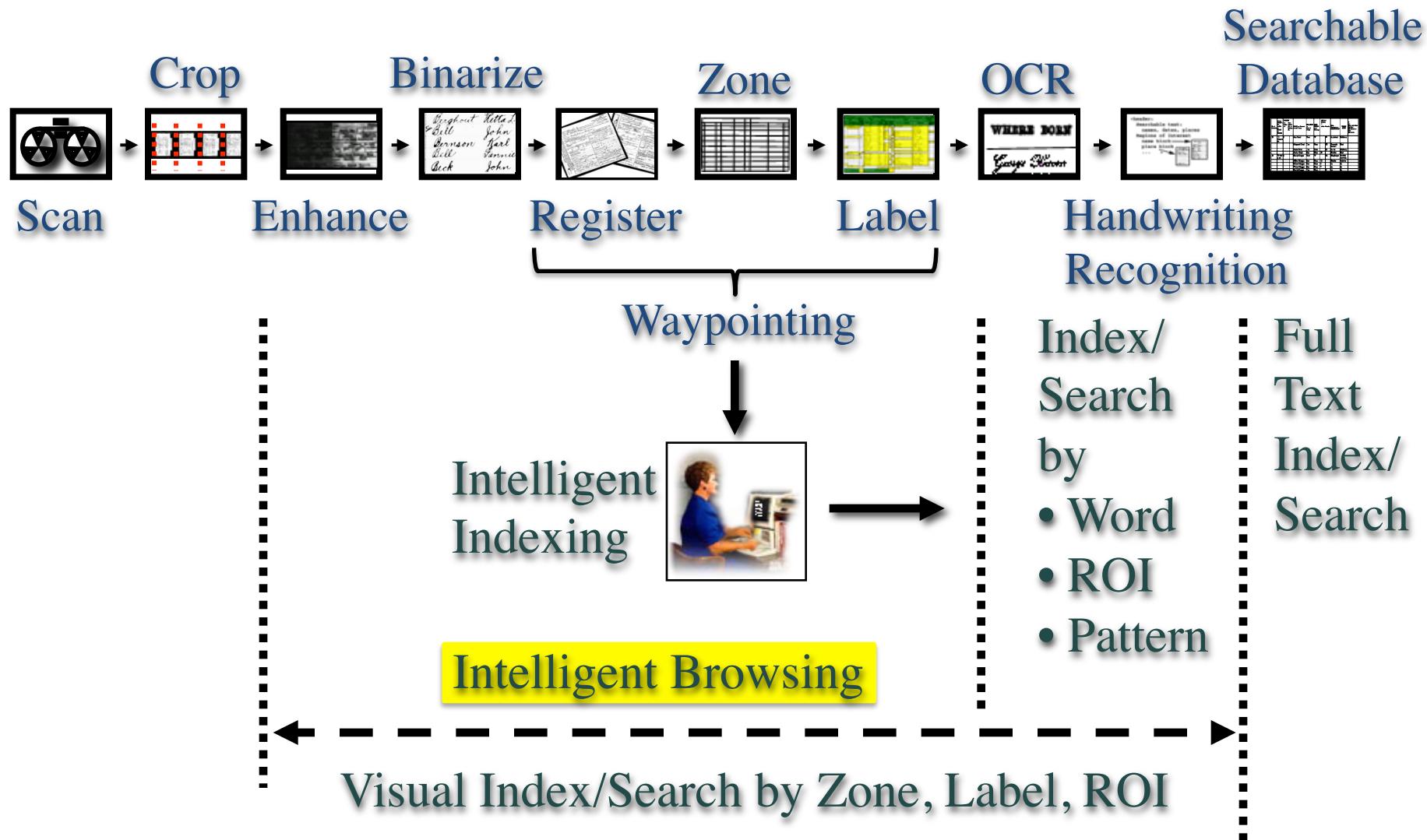
Leveraging the Label

Recommendation System



Leveraging the Label

The Digital Microfilm Pipeline





Search by Subwords/Patterns

- Davis 2018

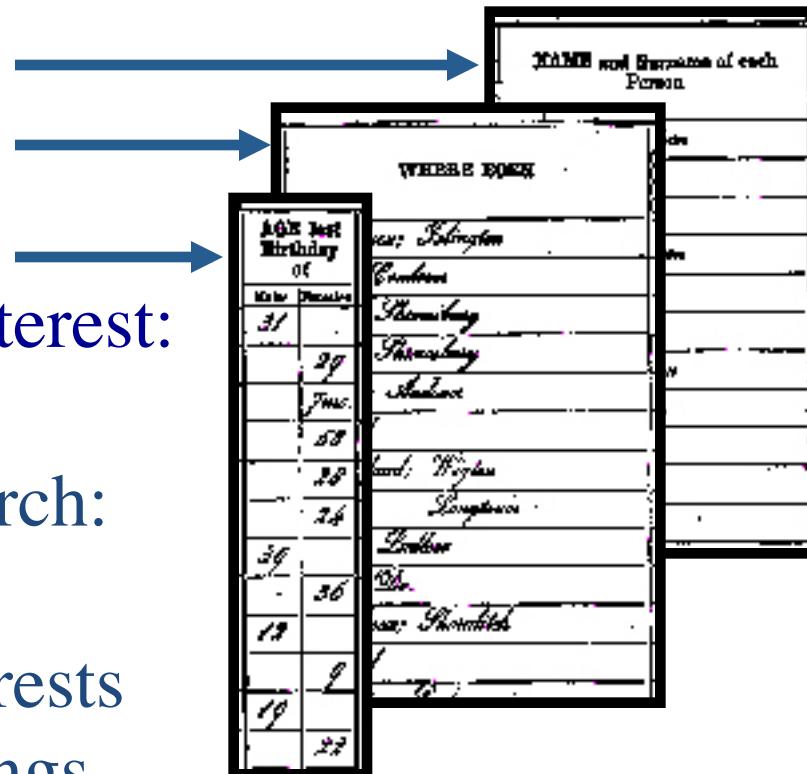
<header:
Name fields
Place fields
Date fields
Regions of I

Partial Word Search:

Subwords

Regions of Interests

Logical Groupings



N-Gram Word Spotting for

1-gram

n

Sum
new
mistrators
number
not
an
them
been
con.
an

2-gram

pr

prices
price
profit
proportionably
probable
process
provisions

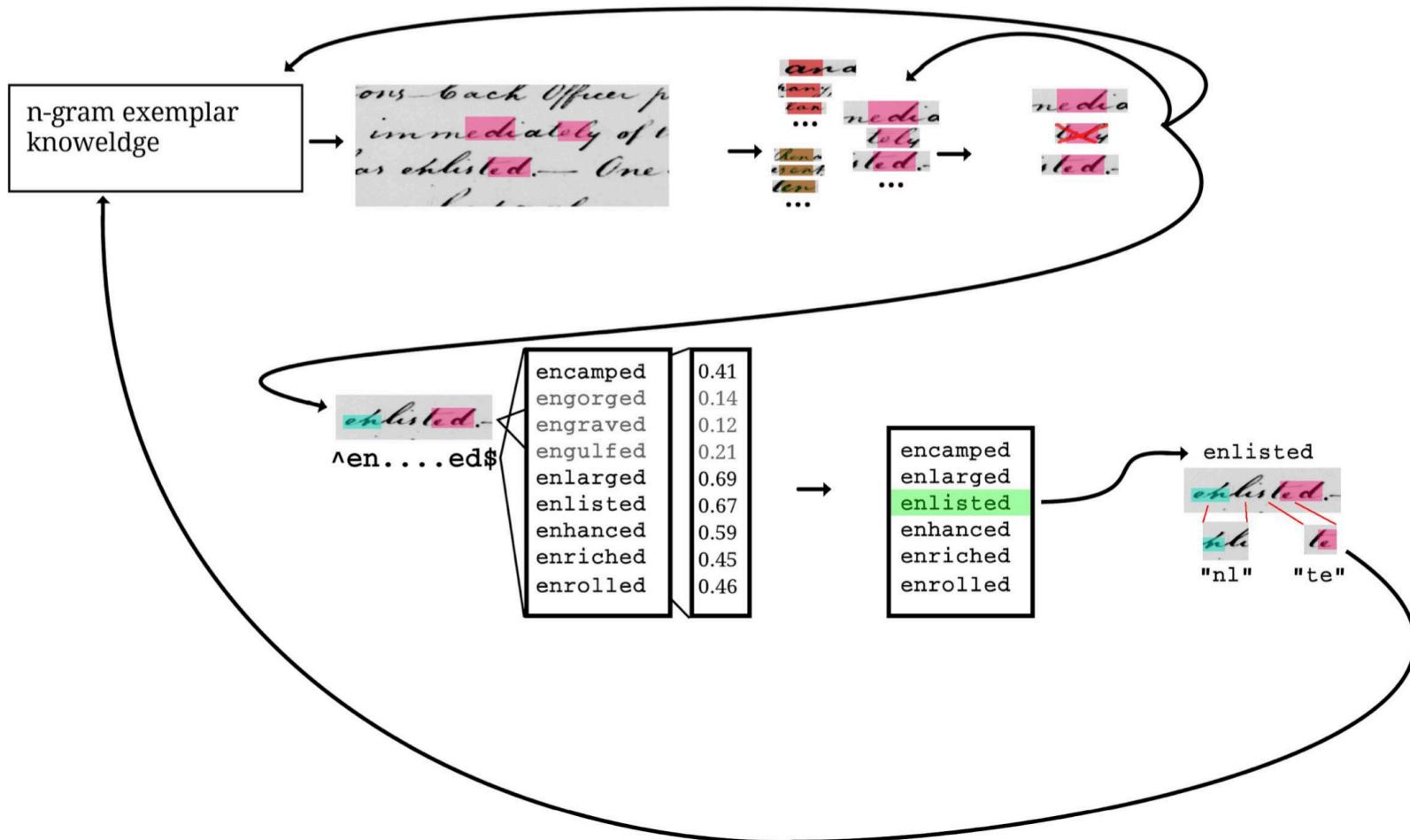
3-gram

sed

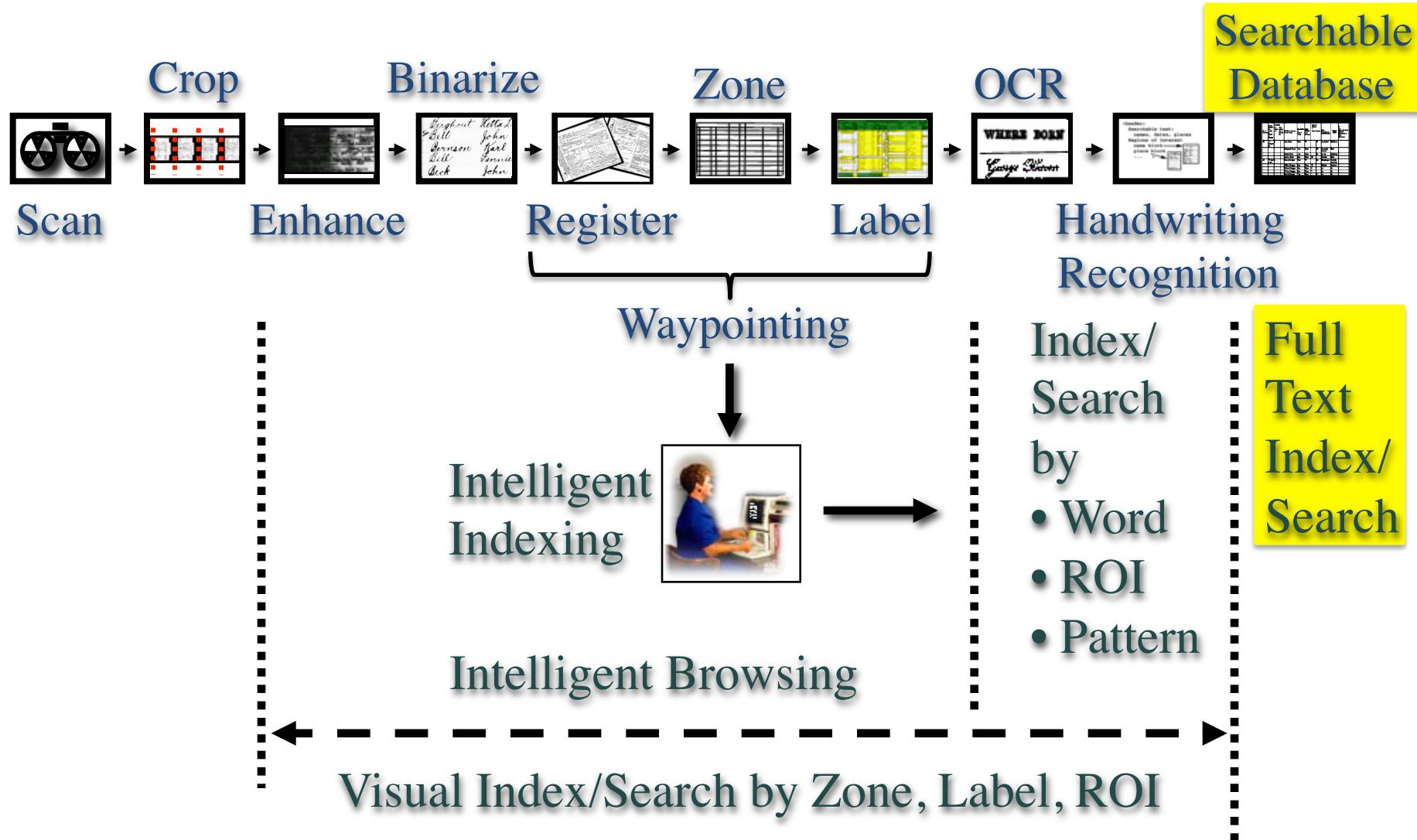
increased.
unpractised.
proposed
proposed
imposed
passed
discussed
Ideots

False positives

N-Gram Word Spotting for Computer-Assisted Transcription



The Digital Microfilm Pipeline





Full Text Database Search

No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES In-habit-ed (U.) or Building (B.)	NAME and Surname of each Person	RELATION to Head of Family	COND-ITION as to Marriage	AGE last Birthday of Males Females	Rank, Profession, or OCCUPATION	WHERE BORN	If (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic
28	15 Church St.	1	Ann Grant	Head	W		50	Anniutant	Salisbury
			Margaret Grant	Dau	Unm		22	Domestic Servant	Kent
			Annie Grant	Dau	Unm		18	do. do.	Salisbury
			Austin Grant	Son	Unm	20		Hawker	Kent
29	72 Salt Lane	1	Emily Morgan	Head	W		34	Shopkeeper	Norfolk, Norwich
			Alfred Morgan	Son	Umn	12		Scholar	Andover
			Hubert Morgan	Son	Umn	9		do.	do.
			Walter Morgan	Son	Umn	7		do.	do.
			Emily Morgan	Dau	Umn		4	do.	do.



Family History Technology Lab

Fhtl.byu.edu

Fan Chart



Elastic Paper

Elasticpaper.org

Fan Chart



Where are we going?

- Deploying handwriting recognition technology
 - FamilySearch
 - Joe Price
 - Mark Clement
 - Humanities, Library, Department of Religion
- Last Session this afternoon
- Sneak Peek

The Problem



Form V. S. No. 11-50M-8-20-17

**STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH**

PLACE OF DEATH.

County of Huron
Township of Norwalk Registration District No. 626 File No. 11899
or
Village of Primary Registration District No. 4934 Registered No. 26
or
City of (No., St., Ward)

[If death occurred in a hospital or institution, give its NAME instead of street and number.]

FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS						MEDICAL CERTIFICATE OF DEATH		
3 SEX	4 COLOR OR RACE	5 SINGLE MARRIED WIDOWED OR DIVORCED (Write the word)				16 DATE OF DEATH		
Male	White	Married				Feb 24	, 1919 (Month) (Day) (Year)	
6 DATE OF BIRTH						I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M.		
7 AGE						The CAUSE OF DEATH* was as follows:		
70 yrs. 9 mos. 22 ds. or min?						Right Lobar Pneumonia		
8 OCCUPATION (a) Trade, profession, or particular kind of work (b) General nature of industry, business, or establishment in which employed (or employer)						Contributory (Secondary)		
9 BIRTHPLACE (State or country)						Duration yrs. mos. 4 ds.		
10 NAME OF FATHER						Duration yrs. mos. ds.		
Patrick Martin						(Signed) 7-24, 1919 (Address) M.D. Norwalk Ohio		
11 BIRTHPLACE OF FATHER (State or country)						*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.		
Ireland						18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)		
12 MAIDEN NAME OF MOTHER						At place of death yrs. mos. ds. In the State yrs. mos. ds.		
Frances Weston						Where was disease contracted, If not at place of death? Former or usual residence		
13 BIRTHPLACE OF MOTHER (State or country)						19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL		
Ireland						20 UNDERTAKER ADDRESS		
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) Sarah Martin (Address) Norwalk, O						2-25, 1919 Raibed Brady Norwalk, O		
15 Filed 2/24, 1919 by Fred Martin Registrar								
11-3184								

The Problem



Many documents
are heterogeneous
consisting of

Form V. S. No. 11-50M-8-20-17

PLACE OF DEATH.		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH	
County of <i>Huron</i>	Registration District No. <i>626</i>	File No. <i>11899</i>	
Township of <i>Norwalk</i>	Primary Registration District No. <i>4934</i>	Registered No. <i>26</i>	
or Village of.....	(No.)	St., Ward	
or City of.....			[If death occurred in a hospital or institution, give its NAME instead of street and number.]
² FULL NAME <i>Frederick Martin</i>			
PERSONAL AND STATISTICAL PARTICULARS			
3 SEX <i>Male</i>	4 COLOR OR RACE <i>White</i>	5 SINGLE MARRIED WIDOWED OR DIVORCED (Write the word) <i>Married</i>	
6 DATE OF BIRTH <i>May 1st, 1878</i> (Month) (Day) (Year)			
7 AGE <i>70 yrs. 9 mos. 22 ds.</i>	If LESS than 1 day, _____ hrs. or _____ min.?		
8 OCCUPATION (a) Trade, profession, or particular kind of work <i>Retired laborer</i> (b) General nature of industry, business, or establishment in which employed (or employer) <i>92</i>			
9 BIRTHPLACE (State or country) <i>Ohio</i>			
10 NAME OF FATHER <i>Patrick Martin</i>			
11 BIRTHPLACE OF FATHER (State or country) <i>Ireland</i>			
12 MAIDEN NAME OF MOTHER <i>Frances Weston</i>			
13 BIRTHPLACE OF MOTHER (State or country) <i>Ireland</i>			
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <i>Sarah Martin</i> (Address) <i>Norwalk, Ohio</i>			
15 Filed <i>2/24, 1919</i> <i>Frank Martin</i> Registrar			
16 MEDICAL CERTIFICATE OF DEATH			
16 DATE OF DEATH <i>Feb 24, 1919</i> (Month) (Day) (Year)			
17 I HEREBY CERTIFY, That I attended deceased from <i>Feb 20, 1919</i> , to <i>24, 1919</i> , that I last saw him alive on <i>Feb 23, 1919</i> , and that death occurred, on the date stated above, at <i>1 A.M.</i> . The CAUSE OF DEATH* was as follows: <i>Right Lobar Pneumonia</i>			
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents) At place of death _____ yrs. _____ mos. _____ ds. Where was disease contracted, If not at place of death? Former or usual residence			
19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL <i>St. Mary's Cemetery</i> <i>2-25, 1919</i>			
20 UNDERTAKER ADDRESS <i>Raised Brady</i> <i>Norwalk, Ohio</i>			

The Problem



Many documents
are heterogeneous
consisting of
- handwriting

Form V. S. No. 11-50M-8-20-17

**STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH**

11899

PLACE OF DEATH.
County of **Huron** Township of **Morwalk** Registration District No. **626** File No. **.....**

Township of **Morwalk** or Village of **.....** Primary Registration District No. **4934** Registered No. **26**

City of **.....** (No. **.....**) St. **.....** Ward **.....** [If death occurred in a hospital or institution, give its NAME instead of street and number.]

FULL NAME **Frederick Martin**

PERSONAL AND STATISTICAL PARTICULARS

3 SEX **Male** **4 COLOR OR RACE** **White** **5 SINGLE MARRIED WIDOWED OR DIVORCED** **Married** (Write the word)

6 DATE OF BIRTH **May 1st**, **1878** (Month) (Day) (Year)

7 AGE **70** yrs. **9** mos. **22** ds. If LESS than 1 day, _____ hrs. or _____ min.?

8 OCCUPATION
(a) Trade, profession, or particular kind of work **Retired laborer**
(b) General nature of industry, business, or establishment in which employed (or employer) **.....**

9 BIRTHPLACE (State or country) **Ohio**

10 NAME OF FATHER **Patrick Martin**

11 BIRTHPLACE OF FATHER (State or country) **Ireland**

12 MAIDEN NAME OF MOTHER **Frances Weston**

13 BIRTHPLACE OF MOTHER (State or country) **Ireland**

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) **Sarah Martin**
(Address) **Morwalk, Ohio**

15 Filed **2/24**, **1919** **By** **Glost Martin**
Registrar

16 DATE OF DEATH **Feb 24**, **1919** (Month) (Day) (Year)

17 I HEREBY CERTIFY, That I attended deceased from **Feb 20, **1919**, to **24**, **1919**, that I last saw him alive on **Feb 23**, **1919**, and that death occurred, on the date stated above, at **1 A.M.** The CAUSE OF DEATH was as follows:**

Right Lobar Pneumonia

18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
At place of death _____ yrs. _____ mos. _____ ds. In the State _____ yrs. _____ mos. _____ ds.
Where was disease contracted, If not at place of death?
Former or usual residence

19 PLACE OF BURIAL OR REMOVAL **St. Mary's Cemetery** **DATE OF BURIAL** **2-25**, **1919**

20 UNDERTAKER **Raised Brady** **ADDRESS** **Morwalk, Ohio**

11-3184

The Problem



Many documents
are heterogeneous
consisting of
- handwriting
- machine print

Form V. S. No. 11-50M-8-20-17

STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH

11899

PLACE OF DEATH.
County of Huron
Township of Norwalk Registration District No. 626 File No.
or Village of Primary Registration District No. 4934 Registered No. 26
or City of (No. St., Ward)

[If death occurred in a hospital or institution, give its NAME instead of street and number.]

FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS				MEDICAL CERTIFICATE OF DEATH			
3 SEX	4 COLOR OR RACE	5 MARRIED	6 DATE OF BIRTH	16 DATE OF DEATH	17 I HEREBY CERTIFY, That I attended deceased from	18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)	
Male	White	Married	May 1st, 1878 (Month) (Day) (Year)	Feb 24, 1919 (Month) (Day) (Year)	Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M.	At place of death..... yrs. mos. ds. Where was disease contracted, If not at place of death? Former or usual residence	
7 AGE	8 OCCUPATION	9 If LESS than 1 day, hrs. or min.?	10 NAME OF FATHER	11 BIRTHPLACE OF FATHER	12 MAIDEN NAME OF MOTHER	13 BIRTHPLACE OF MOTHER	14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
70 yrs. 9 mos. 22 ds.	(a) Trade, profession, or particular kind of work Retired laborer	1 day, 0 hrs. or 0 min.?	Patrick Martin	Ohio	Frances Weston	Ireland	(Informant) Sarah Martin (Address) Norwalk, Ohio
15 Filed 2/24, 1919	16 PLACE OF BURIAL OR REMOVAL	17 UNDERTAKER	18 DATE OF BURIAL				
	St. Mary's Cemetery	Raised & Brady	2-25, 1919				
			ADDRESS				

11-3184

The Problem



Many documents
are heterogeneous
consisting of
 - handwriting
 - machine print
 - solid lines

Form V. S. No. 11-50M-8-20-17

PLACE OF DEATH		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH	
County of Huron		Registration District No. 626	File No. 11899
Township of Norwalk		Primary Registration District No. 4934	Registered No. 26
or Village of or City of		(No.)	St., Ward)
FULL NAME		Frederick Martin	
[If death occurred in a hospital or institution, give its NAME instead of street and number.]			
PERSONAL AND STATISTICAL PARTICULARS			
3 SEX	4 COLOR OR RACE	5 MARRIED	6 DATE OF BIRTH
Male	White	Married	May 1st, 1878
		(Write the word)	(Month) (Day) (Year)
7 AGE	8 OCCUPATION	9 If LESS than	10 NAME OF FATHER
70 yrs.	(a) Trade, profession, or particular kind of work	1 day, hrs. or min.?	Patrick Martin
	(b) General nature of industry, business, or establishment in which employed (or employer)		
11 BIRTHPLACE	12 MAIDEN NAME OF MOTHER	13 BIRTHPLACE OF MOTHER	14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(State or country)	(State or country)	(State or country)	(Informant) Sarah Martin
PARENTS	(Address)	(Address)	(Address) Norwalk, Ohio
15 Filed 2/24, 1919	16 DATE OF DEATH	17 I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M. The CAUSE OF DEATH was as follows:	18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
	Feb 24, 1919	Right Lobar Pneumonia	At place of death yrs. mos. ds. In the State yrs. mos. ds.
			Where was disease contracted, If not at place of death? Former or usual residence
			19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL
			20 UNDERTAKER ADDRESS
			Raised Brady Norwalk, Ohio

11-3184

The Problem



Many documents
are heterogeneous
consisting of

- handwriting
- machine print
- solid lines
- dashed lines

Form V. S. No. 11-50M-8-20-17

**STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH**

11899

PLACE OF DEATH.
County of Huron
Township of Norwalk Registration District No. 626 File No.
or Village of Primary Registration District No. 4934 Registered No. 26
or City of (No. St., Ward)

[If death occurred in a hospital or institution, give its NAME instead of street and number.]

FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS

SEX Male **COLOR OR RACE** White **MARITAL STATUS** Married

DATE OF BIRTH May 1st 1878
(Month) (Day) (Year)

AGE 70 yrs. 9 mos. 22 ds. If LESS than 1 day, hrs. or min.?

OCCUPATION
(a) Trade, profession, or particular kind of work Retired laborer
(b) General nature of industry, business, or establishment in which employed (or employer)

BIRTHPLACE (State or country) Ohio

PARENTS

NAME OF FATHER Patrick Martin

BIRTHPLACE OF FATHER (State or country) Ireland

MAIDEN NAME OF MOTHER Frances Weston

BIRTHPLACE OF MOTHER (State or country) Ireland

THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, O.
Filed 2/24, 1919 by Fred Martin
Registrar

MEDICAL CERTIFICATE OF DEATH

DATE OF DEATH Feb 24, 1919
(Month) (Day) (Year)

I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M. The CAUSE OF DEATH was as follows: Right Lobar Pneumonia

Contributory Hernia heart lesion

Length of Residence (Duration) yrs. mos. 4 ds.
(Signed) 7/24/1919 by Maynard M. D.
(Address) Norwalk, O.

*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.

Length of Residence At place of death yrs. mos. ds. In the State yrs. mos. ds.
Where was disease contracted, If not at place of death?
Former or usual residence

PLACE OF BURIAL OR REMOVAL DATE OF BURIAL
20 UNDERTAKER ADDRESS
Raibled Brady Norwalk, O. 2-25, 1919

11-3184

The Problem



Many documents
are heterogeneous
consisting of

- handwriting
- machine print
- solid lines
- dashed lines
- stamps

Form V. S. No. 11-50M-8-20-17

**STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH**

PLACE OF DEATH.
County of Huron
Township of Norwalk Registration District No. 626 File No. 11899
or
Village of _____ Primary Registration District No. 4934 Registered No. 26
or
City of _____ (No. _____, St., _____, Ward)

FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS

SEX Male **COLOR OR RACE** White **MARITAL STATUS** Married

DATE OF BIRTH May 1st 1878
(Month) (Day) (Year)

AGE 70 yrs. 9 mos. 22 ds. If LESS than 1 day, hrs. or min.?

OCCUPATION
(a) Trade, profession, or particular kind of work Retired laborer
(b) General nature of industry, business, or establishment in which employed (or employer)

BIRTHPLACE (State or country) Ohio

PARENTS
NAME OF FATHER Patrick Martin
BIRTHPLACE OF FATHER Ireland
MAIDEN NAME OF MOTHER Frances Weston
BIRTHPLACE OF MOTHER Ireland

THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, Ohio
Filed 2/24, 1919 by Fred Martin
Registrar

MEDICAL CERTIFICATE OF DEATH

DATE OF DEATH Feb 24, 1919
(Month) (Day) (Year)

I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M. The CAUSE OF DEATH was as follows: Right Lobar Pneumonia

Contributory Hernia heart lesion

LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
At place _____ yrs. _____ mos. _____ ds.
In the state _____ yrs. _____ mos. _____ ds.
Where was disease contracted, _____
If not at place of death, _____
Former or usual residence, _____

PLACE OF BURIAL OR REMOVAL St. Mary's Cemetery DATE OF BURIAL 2-25, 1919
UNDERTAKER Laible & Brady ADDRESS Norwalk, Ohio

11-3184

The Problem



Many documents
are heterogeneous
consisting of

- handwriting
- machine print
- solid lines
- dashed lines
- stamps

that are

- interleaved

Form V. S. No. 11-50M-8-20-17

PLACE OF DEATH		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH
County of Huron		Registration District No. 626 File No. 11899
Township of Norwalk		Primary Registration District No. 4934 Registered No. 26
Village of _____ or City of _____		(No. _____, St., _____, Ward)
[If death occurred in a hospital or institution, give its NAME instead of street and number.]		
2 FULL NAME Frederick Martin		
PERSONAL AND STATISTICAL PARTICULARS		
3 SEX	4 COLOR OR RACE	5 MARRIED
Male	White	Married
6 DATE OF BIRTH		
May 1st		1878
(Month) (Day)		(Year)
7 AGE	If LESS than 70 yrs. 9 mos. 22 ds. or min.?	
8 OCCUPATION (a) Trade, profession, or particular kind of work (b) General nature of industry, business, or establishment in which employed (or employer)		
Retired laborer		
9 BIRTHPLACE (State or country)		
Ohio		
10 NAME OF FATHER		
Patrick Martin		
11 BIRTHPLACE OF FATHER (State or country)		
Ireland		
12 MAIDEN NAME OF MOTHER		
Frances Weston		
13 BIRTHPLACE OF MOTHER (State or country)		
Ireland		
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE		
(Informant) Sarah Martin		
(Address) Norwalk, Ohio		
15 Filed	1919	
Glost Martin		Registrar
16 MEDICAL CERTIFICATE OF DEATH		
16 DATE OF DEATH		
Feb 24		1919
(Month) (Day)		(Year)
17 I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M.		
The CAUSE OF DEATH was as follows:		
Right Lobar Pneumonia		
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)		
At place of death yrs. mos. ds. In the State yrs. mos. ds.		
Where was disease contracted, If not at place of death? Former or usual residence		
19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL		
20 UNDERTAKER		ADDRESS
Raibled Brady		Norwalk, Ohio

The Problem



Many documents
are heterogeneous
consisting of

- handwriting
- machine print
- solid lines
- dashed lines
- stamps

that are

- interleaved
- and frequently
- overlap

Form V. S. No. 11-50M-8-20-17

**STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH**

PLACE OF DEATH.
County of Huron
Township of Norwalk Registration District No. 626 File No. 11899
or
Village of _____ Primary Registration District No. 4934 Registered No. 26
or
City of _____ (No. _____, St., _____, Ward)
[If death occurred in a hospital or institution, give its NAME instead of street and number.]

FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS

5 SEX Male 6 COLOR OR RACE White 7 MARRIED Single Married Widowed Divorced
(Write the word) married

8 DATE OF BIRTH May 1st 1878
(Month) (Day) (Year)

9 AGE 70 yrs. 1 mos. 22 ds. If LESS than 1 day, hrs.
or min.?

10 OCCUPATION Retired laborer

(a) Trade, profession, or particular kind of work
(b) General nature of industry, business, or establishment in which employed (or employer)

11 BIRTHPLACE Ohio
(State or country)

12 NAME OF FATHER Patrick Martin

13 BIRTHPLACE OF FATHER Ireland
(State or country)

14 MAIDEN NAME OF MOTHER Frances Lester

15 BIRTHPLACE OF MOTHER Ireland
(State or country)

16 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, Ohio
Filed 2/24/1919, 1919 by Frank Martin, Registrar

MEDICAL CERTIFICATE OF DEATH

17 DATE OF DEATH Feb 24, 1919
(Month) (Day) (Year)

I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M. The CAUSE OF DEATH was as follows: Right Lobar Pneumonia

18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
At place of death _____ yrs. _____ mos. _____ ds.
Where was disease contracted, If not at place of death?
Former or usual residence

19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL
St. Mary's Cemetery 2-25, 1919
20 UNDERTAKER ADDRESS
Raibled Brady Norwalk, Ohio

11-3184

1919 Ohio Death Record

The Problem



The interleave and overlap, greatly compounds the difficulty of OCR and handwriting recognition.

Form V. S. No. 11-50M-8-20-17

PLACE OF DEATH		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH
County of Huron		Registration District No. 626 File No. 11899
Township of Norwalk		Primary Registration District No. 4934 Registered No. 26
Village of _____ or City of _____		(No.) St., Ward)
[If death occurred in a hospital or institution, give its NAME instead of street and number.]		
2 FULL NAME <u>Frederick Martin</u>		
PERSONAL AND STATISTICAL PARTICULARS		
3 SEX <u>Male</u>	4 COLOR OR RACE <u>White</u>	5 MARRIED <u>Married</u> SINGLE MARRIED WIDOWED OR DIVORCED (Write the word)
6 DATE OF BIRTH <u>May 1st 1878</u> (Month) (Day) (Year)		
7 AGE <u>70 yrs.</u>	8 OCCUPATION <u>Retired laborer</u>	If LESS than 1 day, hrs. or min.?
(a) Trade, profession, or particular kind of work <u>Retired laborer</u>		
(b) General nature of industry, business, or establishment in which employed (or employer) <u>None</u>		
9 BIRTHPLACE <u>Ohio</u> (State or country)		
10 NAME OF FATHER <u>Patrick Martin</u>		
11 BIRTHPLACE OF FATHER <u>Ireland</u> (State or country)		
12 MAIDEN NAME OF MOTHER <u>Frances Weston</u>		
13 BIRTHPLACE OF MOTHER <u>Ireland</u>		
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <u>Sarah Martin</u> (Address) <u>Norwalk, Ohio</u>		
15 Filed <u>2/24/1919</u>	16 DATE OF DEATH <u>Feb 24, 1919</u> (Month) (Day) (Year)	MEDICAL CERTIFICATE OF DEATH
17 I HEREBY CERTIFY, That I attended deceased from <u>Feb 20, 1919</u> , to <u>24</u> , 1919, that I last saw him alive on <u>Feb 23, 1919</u> , and that death occurred, on the date stated above, at <u>10</u> a.m. The CAUSE OF DEATH was as follows: <u>Right Lobar Pneumonia</u>		
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents) At place _____ yrs. mos. ds. In the state _____ yrs. mos. ds. Where was disease contracted, If not at place of death? Former or usual residence		
19 PLACE OF BURIAL OR REMOVAL <u>St. Mary's Cemetery</u> DATE OF BURIAL <u>2-25-1919</u>		
20 UNDERTAKER <u>Raised Brady</u> ADDRESS <u>Norwalk, Ohio</u>		

11-3184

The Problem



The **interleave** and **overlap**, greatly compounds the difficulty of **OCR** and **handwriting** recognition.

- Machine print and lines compete with handwriting

Form V. S. No. 11-50M-8-20-17

PLACE OF DEATH		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH
County of Huron		File No. 11899
Township of Norwalk		Registration District No. 626
or Village of _____		Primary Registration District No. 4934
or City of _____		Registered No. 26
(No. _____, St., _____, Ward)		[If death occurred in a hospital or institution, give its NAME instead of street and number.]
2 FULL NAME <i>Frederick Martin</i>		
PERSONAL AND STATISTICAL PARTICULARS		
3 SEX <i>Male</i>	4 COLOR OR RACE <i>White</i>	5 MARRITAL STATUS <i>Married</i>
6 DATE OF BIRTH <i>May 1st 1878</i>		If LESS than 1 day, _____ hrs. or _____ min.?
(Month) <i>May</i>	(Day) <i>1st</i>	(Year) <i>1878</i>
7 AGE <i>70 yrs.</i>	8 OCCUPATION <i>Retired laborer</i>	9 BIRTHPLACE <i>Ohio</i>
(a) Trade, profession, or particular kind of work.		(b) General nature of industry, business, or establishment in which employed (or employer)
10 NAME OF FATHER <i>Patrick Martin</i>		
11 BIRTHPLACE OF FATHER (State or country) <i>Ireland</i>		
12 MAIDEN NAME OF MOTHER <i>Frances Weston</i>		
13 BIRTHPLACE OF MOTHER (State or country) <i>Ireland</i>		
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE		
(Informant) <i>Sarah Martin</i>		(Address) <i>Norwalk, Ohio</i>
15 Filed <i>2/24/1919</i>	16 DATE OF DEATH <i>Feb 24, 1919</i>	MEDICAL CERTIFICATE OF DEATH
		17 I HEREBY CERTIFY, That I attended deceased from <i>Feb 20, 1919</i> , to <i>24, 1919</i> , that I last saw him alive on <i>Feb 23, 1919</i> , and that death occurred, on the date stated above, at <i>1 A.M.</i> The CAUSE OF DEATH was as follows: <i>Right Lobar Pneumonia</i>
		18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents) At place _____ of death _____ yrs. _____ mos. _____ ds. Where was disease contracted, If not at place of death? Former or usual residence
		19 PLACE OF BURIAL OR REMOVAL <i>St. Mary's Cemetery</i>
		DATE OF BURIAL <i>2-25-1919</i>
	20 UNDERTAKER <i>Raised Brady Norwalk, Ohio</i>	ADDRESS

11-3184

The Problem



The interleave and overlap, greatly compounds the difficulty of OCR and handwriting recognition.

- Machine print and lines compete with handwriting
- Machine print and handwriting are best recognized separately

Form V. S. No. 11-50M-8-20-17

PLACE OF DEATH		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH
County of Huron Township of Norwalk		Registration District No. 626 File No. 11899
Village of _____ or City of _____		Primary Registration District No. 4934 Registered No. 26
(No. _____, St., _____, Ward)		[If death occurred in a hospital or institution, give its NAME instead of street and number.]
2 FULL NAME <u>Frederick Martin</u>		
PERSONAL AND STATISTICAL PARTICULARS		
3 SEX	4 COLOR OR RACE	5 MARRITAL STATUS
Male	White	Married
6 DATE OF BIRTH		
May 1st 1878 (Month) (Day) (Year)		
7 AGE	8 OCCUPATION	9 If LESS than 1 day, hrs. or min.?
70 yrs. 9 mos. 22 ds.	Retired laborer	1 day, 0 hrs. or 0 min.?
10 NAME OF FATHER		
Patrick Martin		
11 BIRTHPLACE OF FATHER (State or country)		
Ireland		
12 MAIDEN NAME OF MOTHER		
Frances Weston		
13 BIRTHPLACE OF MOTHER (State or country)		
Ireland		
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE		
(Informant) Sarah Martin (Address) Norwalk, Ohio		
15 Filed 2/24, 1919	Gloss Martin	Registrar
16 MEDICAL CERTIFICATE OF DEATH		
16 DATE OF DEATH Feb 24, 1919 (Month) (Day) (Year)		
17 I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M. The CAUSE OF DEATH was as follows:		
Right Lobar Pneumonia		
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents) At place of death _____ yrs. _____ mos. _____ ds. Where was disease contracted, If not at place of death? Former or usual residence		
19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL		
20 UNDERTAKER ADDRESS		

The Pairing Problem



- Learn the pairings
- 1-to-1
- 1-to-many
- many-to-1
- Improved contextual recognition

Form V. S. No. 11-50M-8-20-17

STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH

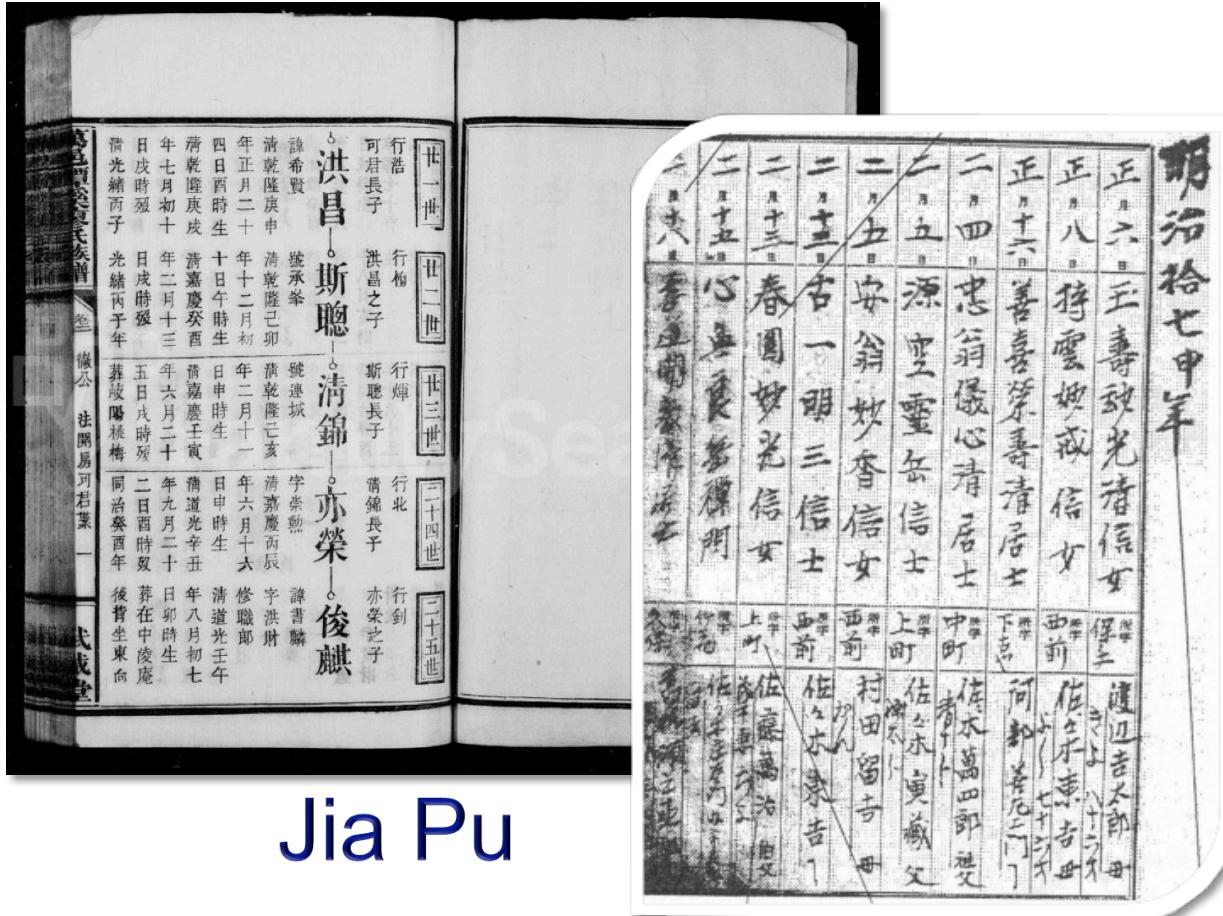
PLACE OF DEATH
County of Huron
Township of Norwalk Registration District No. 626 File No. 11899
or Village of Primary Registration District No. 4934 Registered No. 26
or City of (No.) St., Ward) [If death occurred in a hospital or institution, give its NAME instead of street and number.]

FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS				MEDICAL CERTIFICATE OF DEATH			
3 SEX Male	4 COLOR OR RACE White	5 MARRIED Single	6 WIDOWED Married	16 DATE OF DEATH Feb 24, 1919			
7 DATE OF BIRTH May 1st 1878		(Month)	(Day)	(Year)			
8 AGE 70 yrs.	9 mos.	22 ds.	10 If LESS than 1 day, hrs. or min.?				
11 OCCUPATION Retired laborer				17 I HEREBY CERTIFY, That I attended deceased from Feb 20, 1919, to 24, 1919, that I last saw him alive on Feb 23, 1919, and that death occurred, on the date stated above, at 1 A.M. The CAUSE OF DEATH* was as follows:			
(a) Trade, profession, or particular kind of work				Right Lobar Pneumonia			
(b) General nature of industry, business, or establishment in which employed (or employer)							
12 BIRTHPLACE (State or country) Ohio				18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)			
13 NAME OF FATHER Patrick Martin				At place of death yrs. mos. ds. In the State yrs. mos. ds.			
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) Sarah Martin				Where was disease contracted, If not at place of death? Former or usual residence			
(Address) Norwalk, O				19 PLACE OF BURIAL OR REMOVAL DATE OF BURIAL			
15 Filed 2/24, 1919, by G. H. Martin, Registrar				20 UNDERTAKER ADDRESS			

11-3184

What about different alphabets?





ARTS

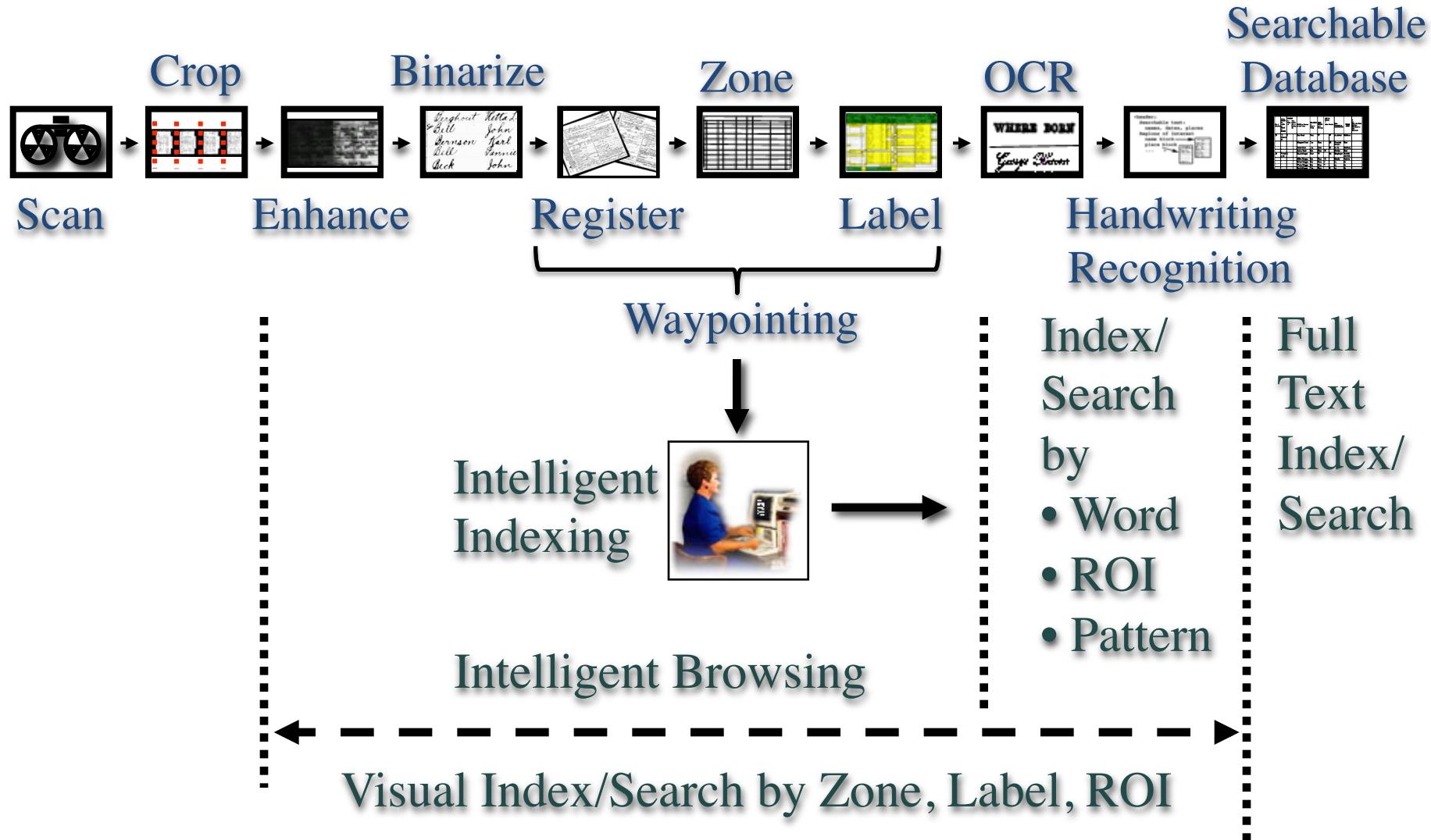
- Asian Records Transcription System

The screenshot displays a transcription system interface for Asian records. On the left, there is a vertical scrollable document containing handwritten Chinese characters and some printed text. The main area shows a pedigree chart with three generations. The first generation consists of 漣璫 (female) and 潘氏 (male). They have four children in the second generation: 斯彬 (male), 李氏 (female), 斯形 (male), and 姚氏 (female). The third generation includes several individuals such as 漣璫's son 斯彬, his wife 潘氏, their son 斯璫, and their daughter 潘氏. Other branches of the family tree are also visible, including a branch for 可君 and another for 日昌.

My big footer. (0 < 0 > 100) slider =>(24.67 < 24.67 > 200) scale

[bin] rb-prototype-Google... [backend] [backend] [doc_15160.doc-Lib... [Manual - Acknowled...]

The Digital Microfilm Pipeline



Credits: The Students



Kelly Lent – 1991

Xiaoxi Tan – 1991

Mai Zhuang – 1992

Timothy Heaton – 1993

Ralf Lehmann – 1995

Yoonkoo Cho – 1995

Kevin Christiansen – 1995

Alex Tang – 1996

Kai Wing Tam – 1998

Glen Sawyer – 1999

Kirk Duffin – 1999

Jack Reese – 1999

Eric Mortensen – 2000

Kevin Petersen – 2001

Alan Cheney – 2001

Luke Hutchison – 2003

Doug Kennard – 2003

Kenric White – 2003

Heath Nielson – 2003

Sam Pinson – 2005

Chris Nelson - 2006

Brian Price - 2006

Chris Armstrong - 2007

Seth Holladay - 2007

Oliver Nina - 2010

Cameron Christiansen - 2012

Doug Kennard - 2013

Robert Clawson - 2014

Kevin Bauer - 2016

Brian Davis - 2018

Seth Stewart - 2018

Curtis Wigington – 2018

Chris Tensmeyer

Iain Lee

Allen Liao

Zhi-han Tsai



Intelligent Pen

Kevin's Slides



Intelligent Pen

A handwritten signature "Edgar" written in black ink on a light gray background with a faint grid pattern. The signature is fluid and cursive, with a vertical line extending upwards from the top of the 'E'.

The same handwritten signature "Edgar" as the one above, but with stroke outlines highlighted in magenta. Two small green dots are placed on the outline of the 'E' and the 'g' to indicate specific points of interest or analysis.



Intelligent Pen

Brooks. Martin S.

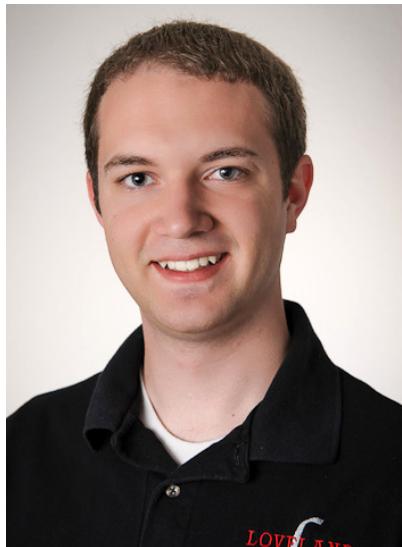
A photograph of a handwritten signature in black ink on a light-colored, textured background. The signature reads "Brooks. Martin S." in a cursive script.

Brooks. Martin S.

A photograph of the same handwritten signature as above, but with magenta stroke outlines. A small green dot marks the starting point of the first stroke on the left.



The Students



Ryan Cheatham
Sophomore
Loveland, Colorado

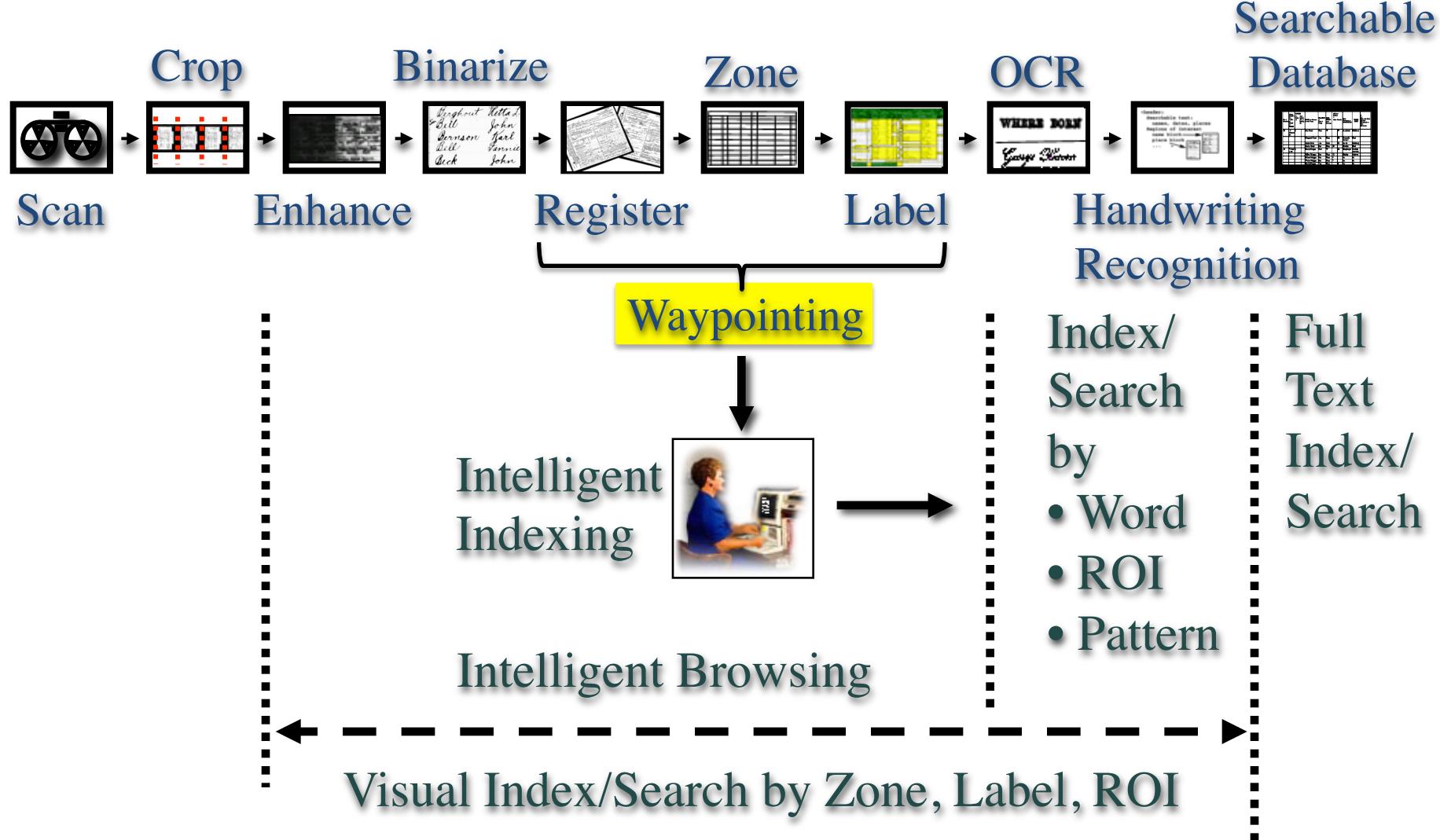


Robert Clawson
MS Student
San Diego, California



Doug Kennard
PhD Student
Austin, Texas

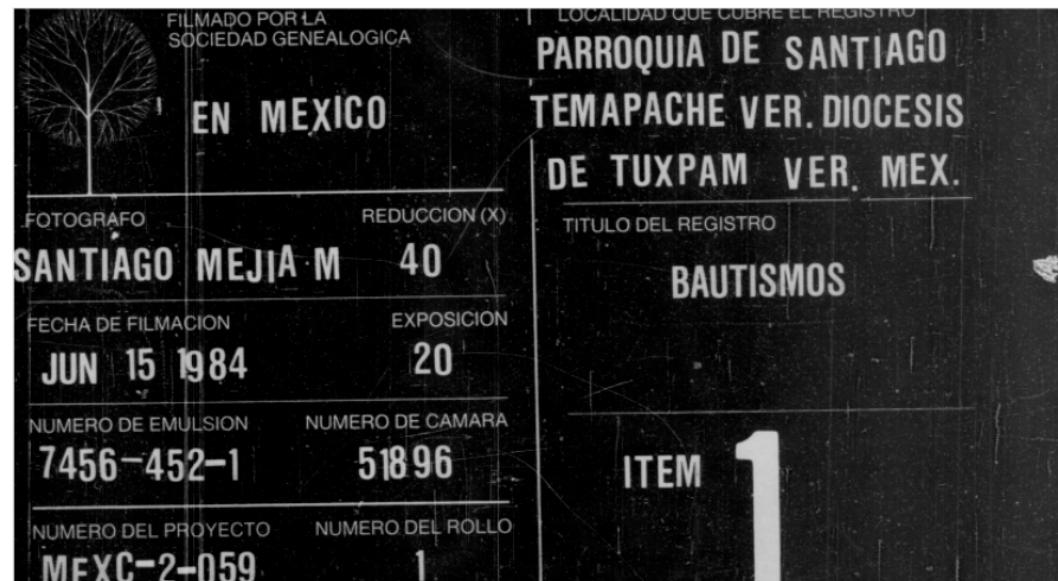
The Digital Microfilm Pipeline



Automated Waypointing: Extract content from titleboards*

Focus on key fields:

- Record type
- Place
- Date range
- Repository
- Film number



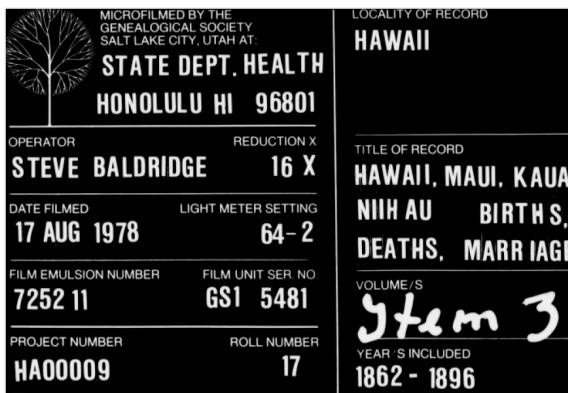
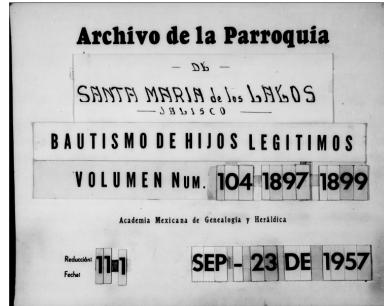
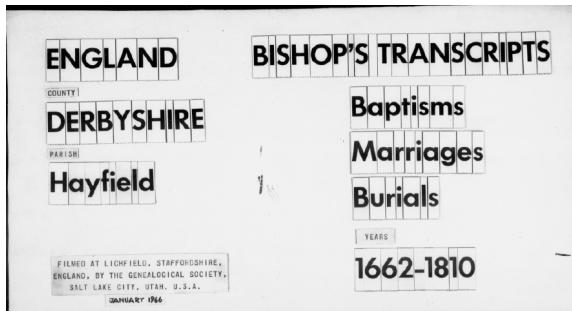
OCR/Index the text

* David Ouimette, Jake

Gehring



Automated Waypointing: Group titleboards uniformly

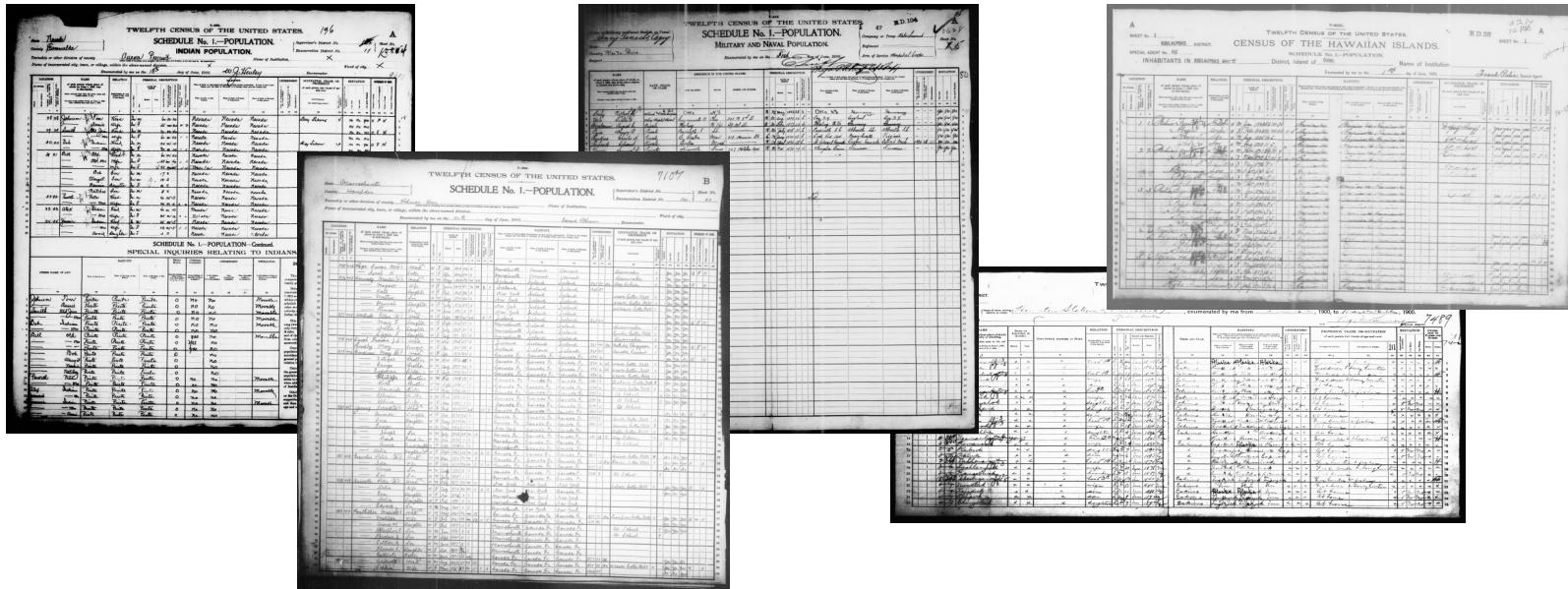


To facilitate collection analysis and indexing prep



Automated Waypointing: Classify images that look alike

Example 1: There are five distinct forms used in the 1900 census. They need to be treated differently when indexing



Tool for index batching and automated waypointing



Automated Waypointing:

Classify images that look alike

- Tool for waypointing and index batching
- Example 2: Create brief indexes (e.g., Rev War pensions)
 - Isolate images to send to indexers (e.g., a single card image which precedes dozens of pages in a pension packet)
 - Waypoint entire set of images, including indexed images

