

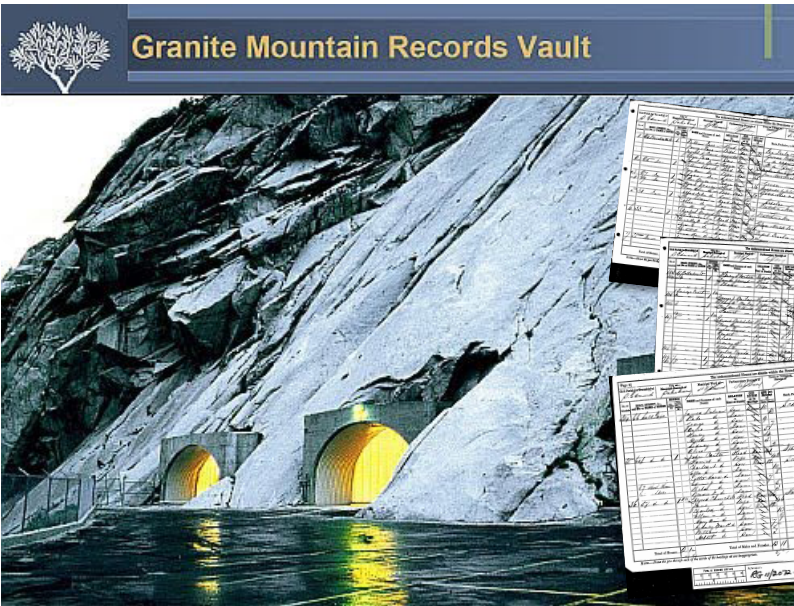


**25 Years of Research in
Family History Technologies at BYU:**
Where we have been
Where we are going

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

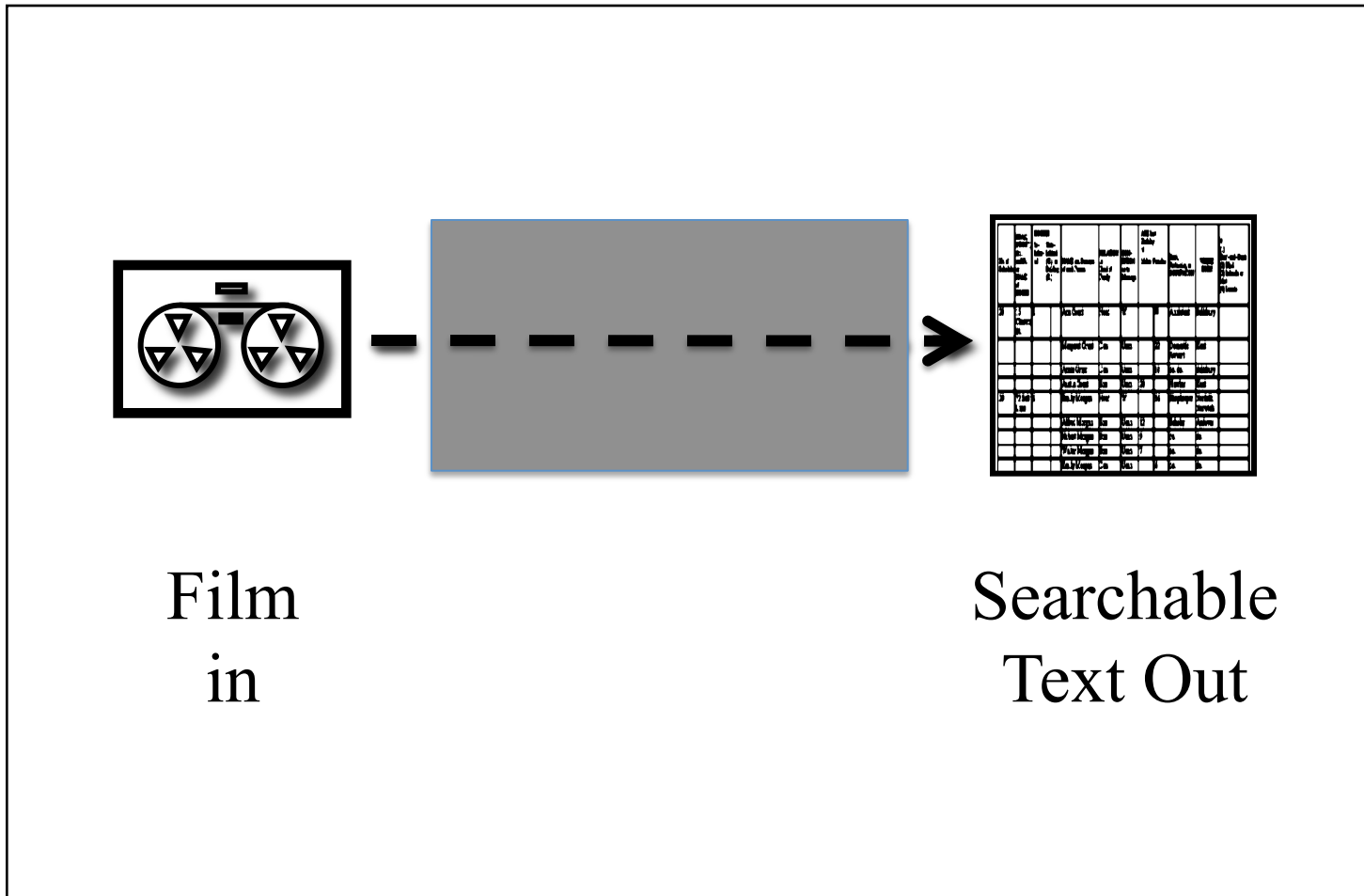
Up to 25 Megabytes per image

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

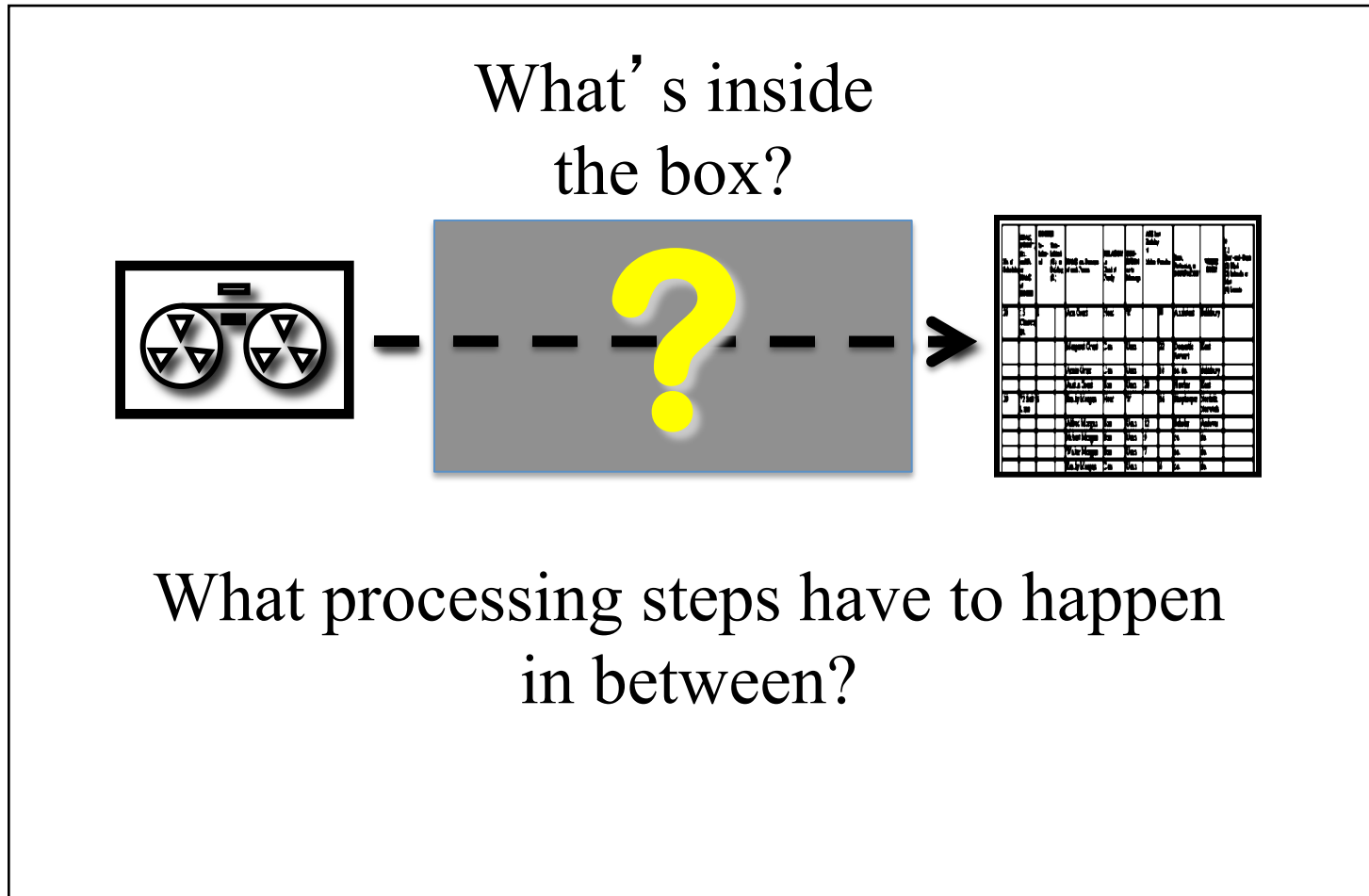
= 81.25 Petabytes



The Digital Microfilm Pipeline*

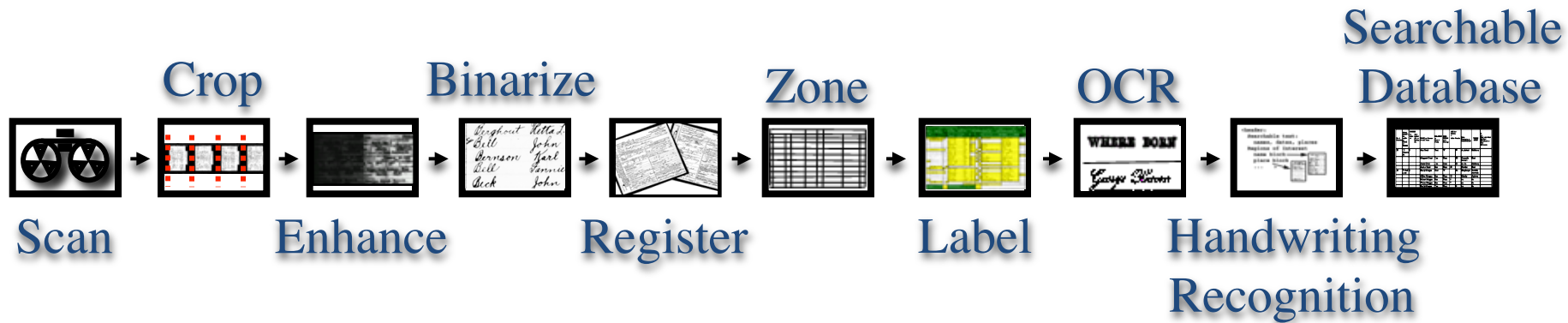


The Digital Microfilm Pipeline



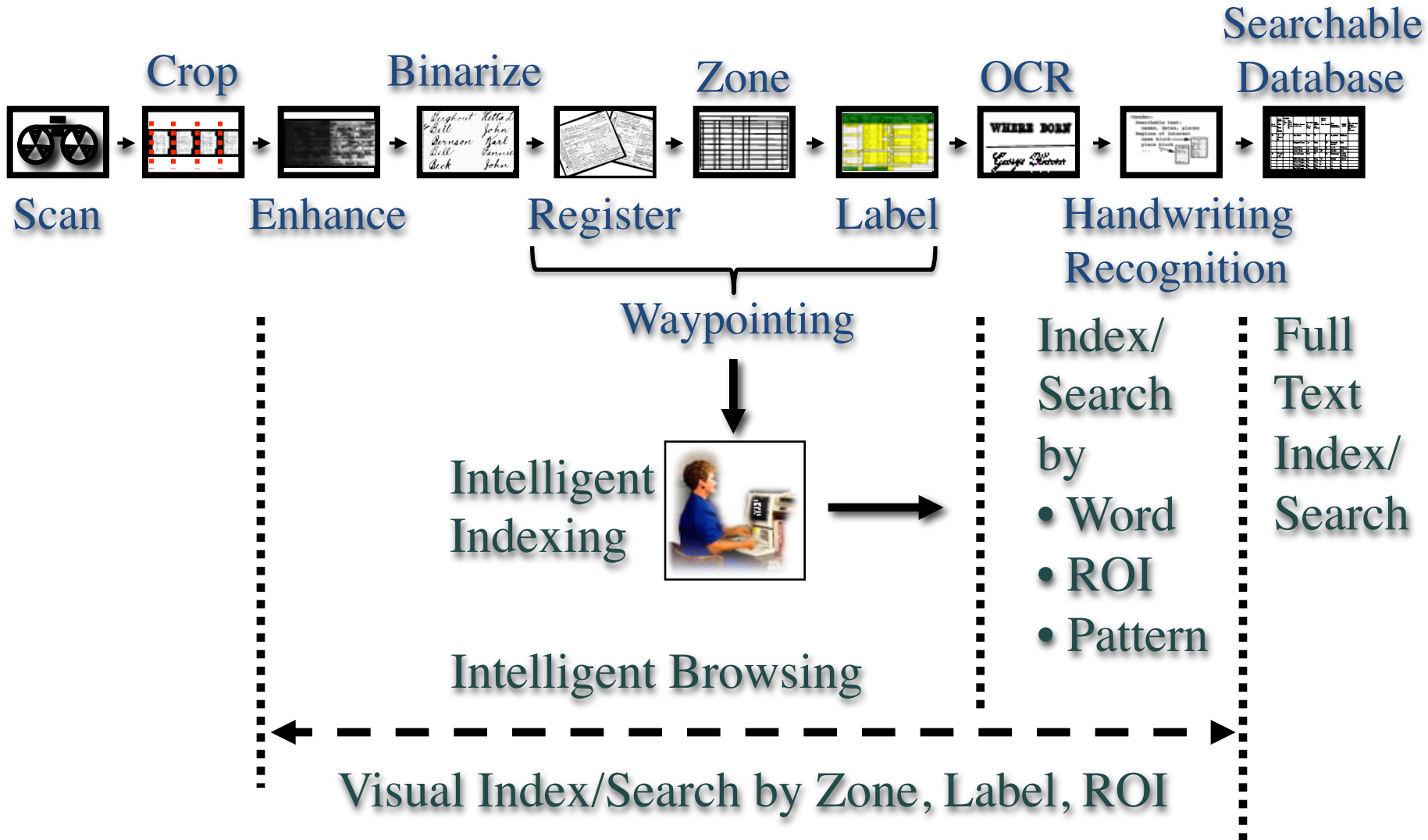


The Digital Microfilm Pipeline

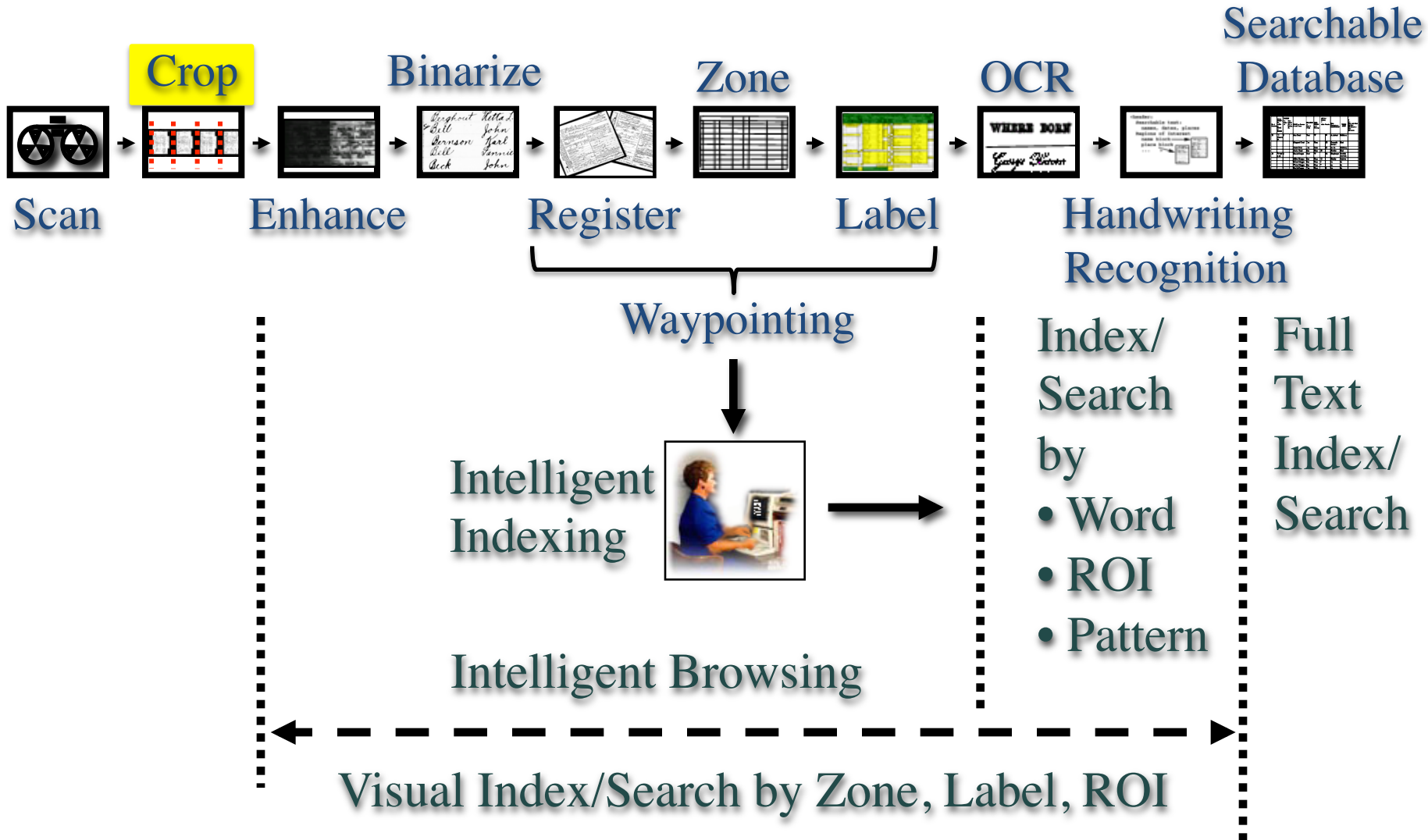


We envisioned a sequence of processing steps that needed to happen
– like the Henry Ford Assembly line

The Digital Microfilm Pipeline



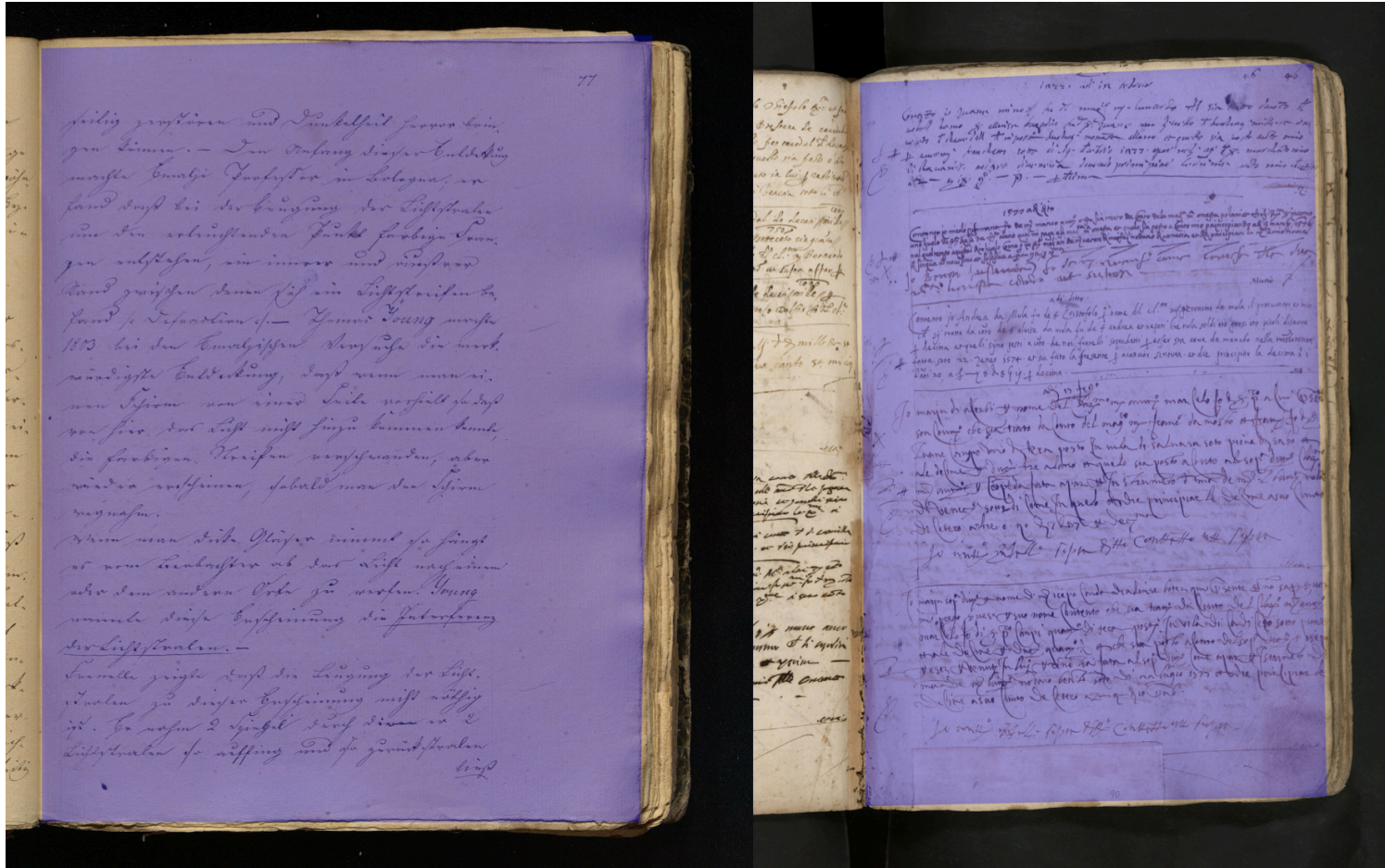
The Digital Microfilm Pipeline



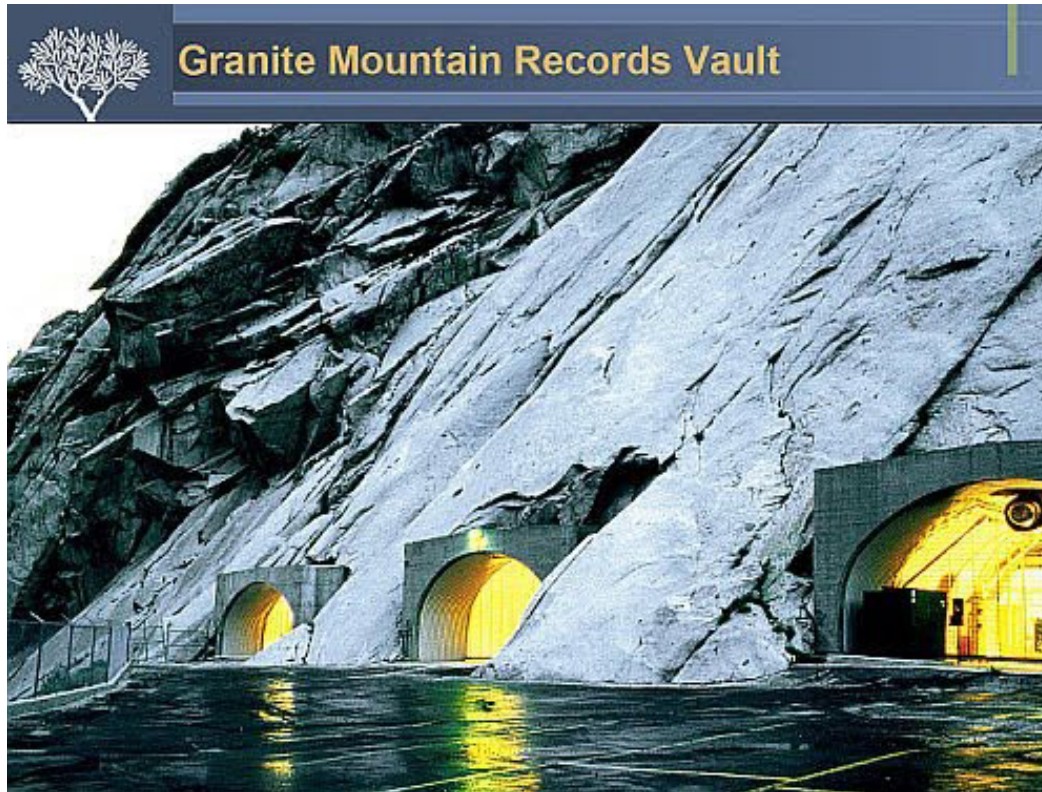
Cropping – 2017 Tensmeyer et al



PageNet – Fully Convolutional Network

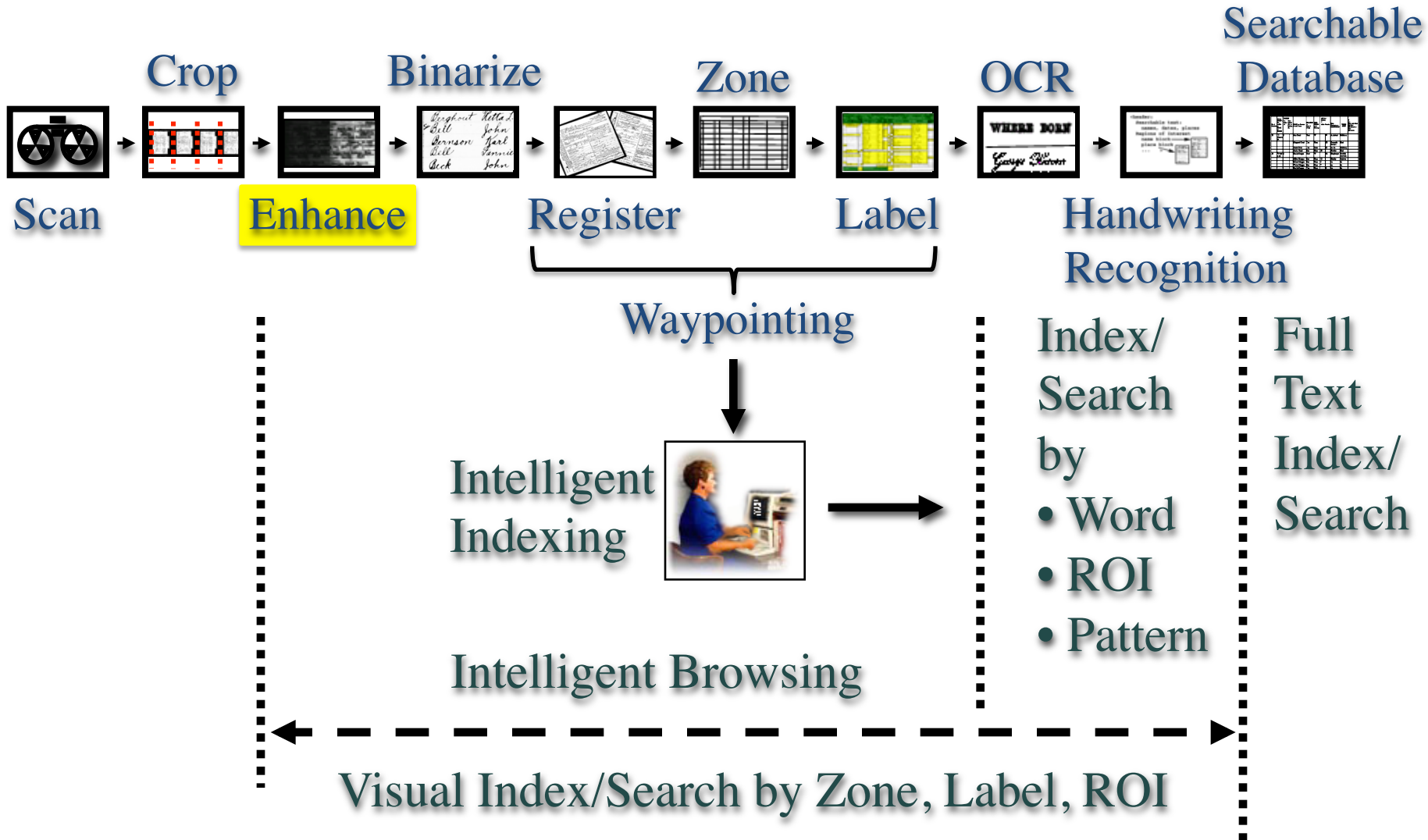


The Digital Microfilm Pipeline

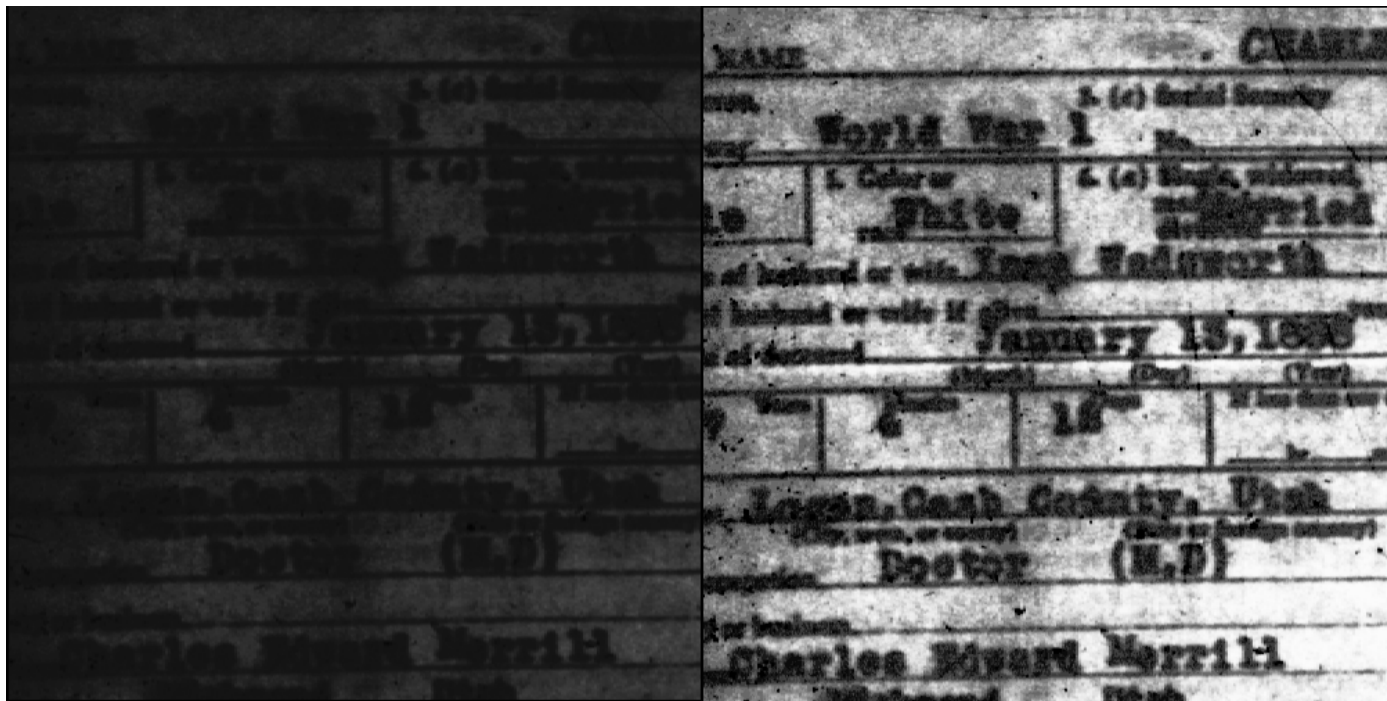


10 Terabytes/day

The Digital Microfilm Pipeline

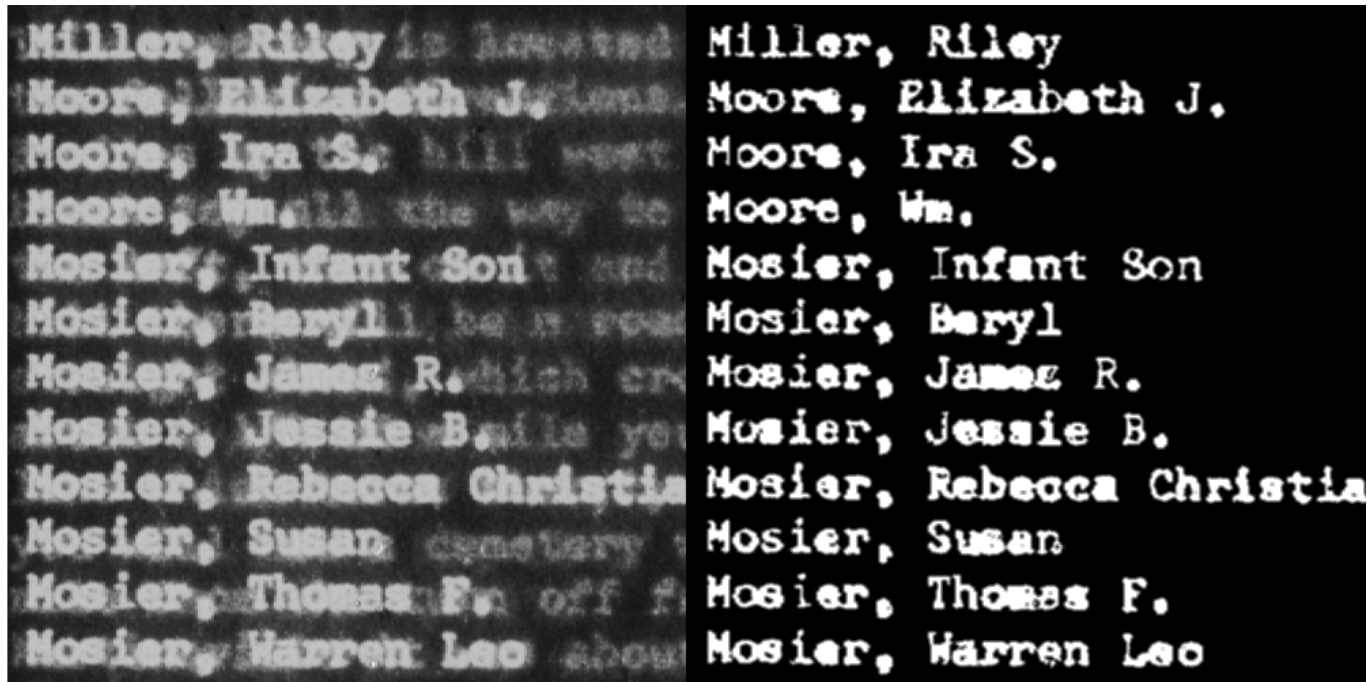


Histogram Enhancement



(Left) Original scanned record (Right) After Enhancement

Threshold Enhancement



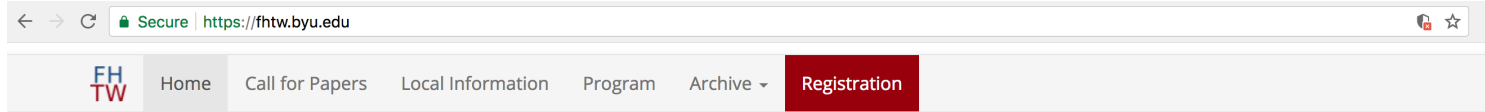
(Left) Original scanned record (Right) After thresholding for Bleed-through removal, with Enhancement, and Antialiasing

Stroke Enhancement



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

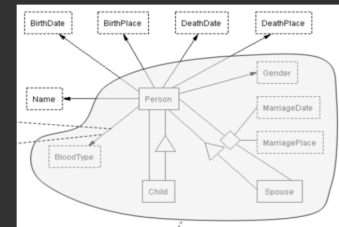
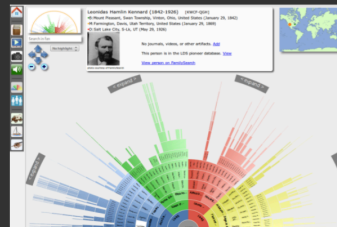
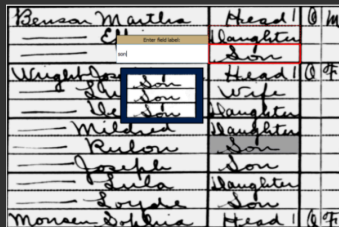
Family History Technology Workshop (fhtw.byu.edu) -2001



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.



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) - 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

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

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.

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



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.

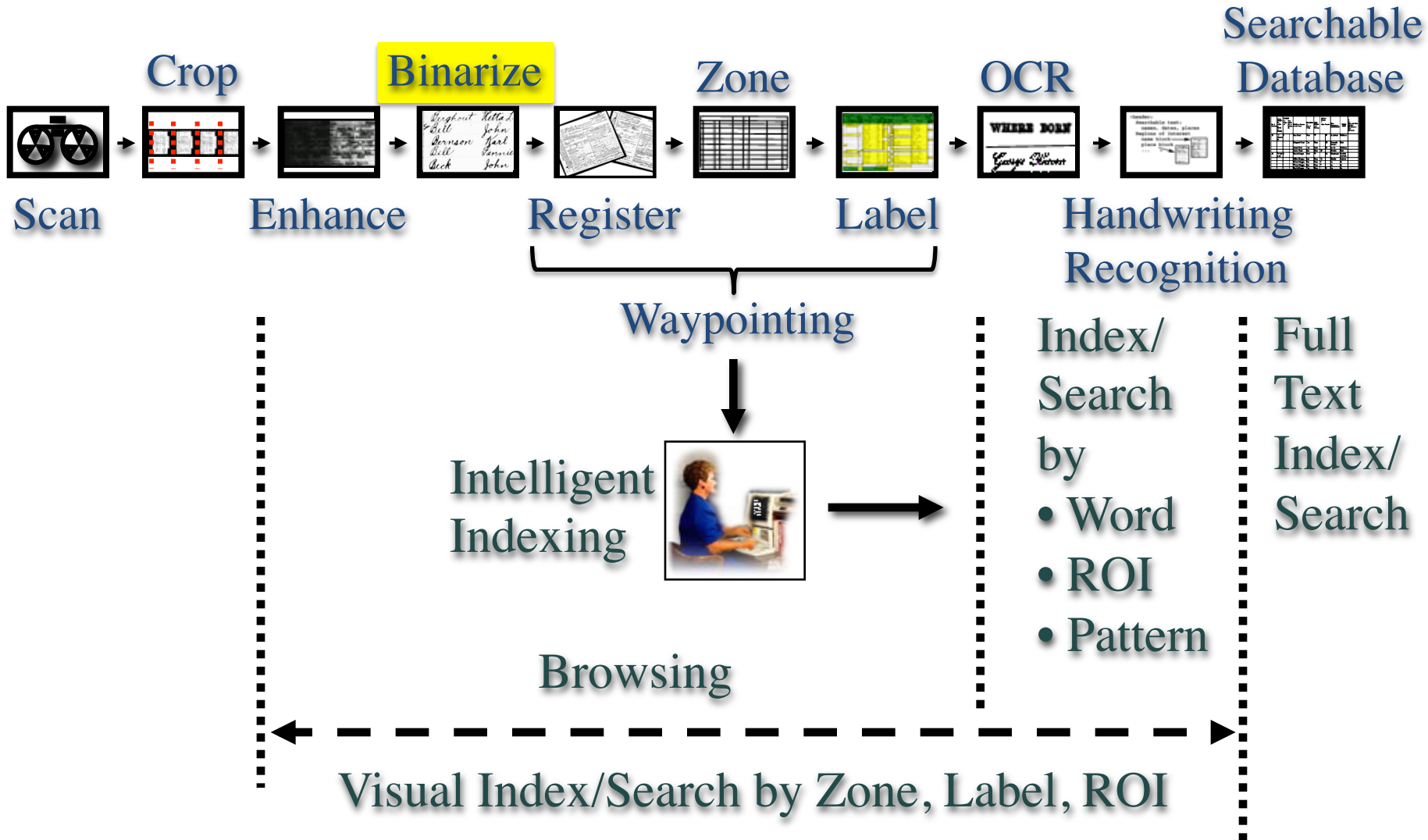
[Watch a Training Video](#)



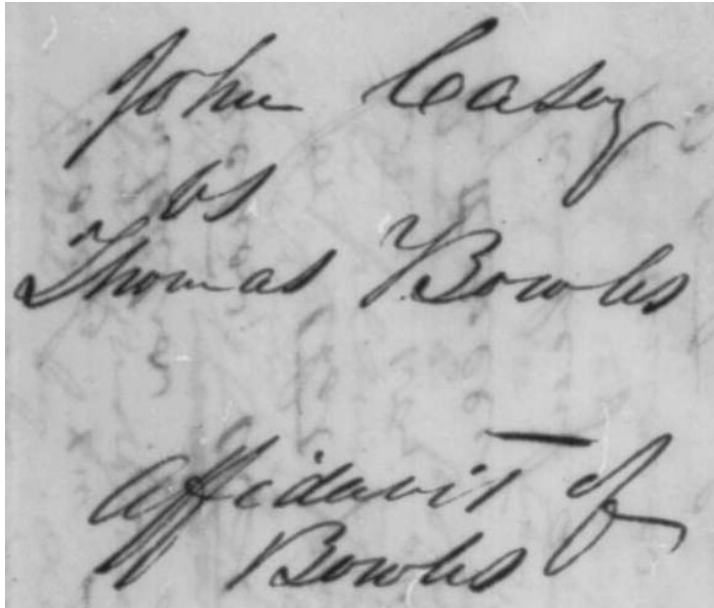
OPG



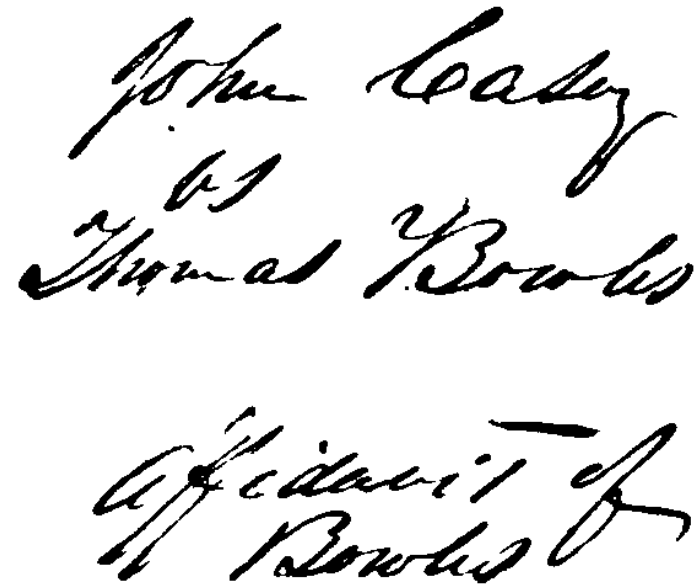
The Digital Microfilm Pipeline



Binarization



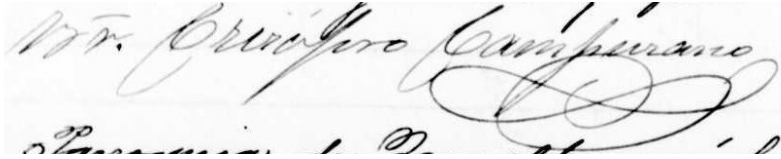
(a) Original Image



(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)

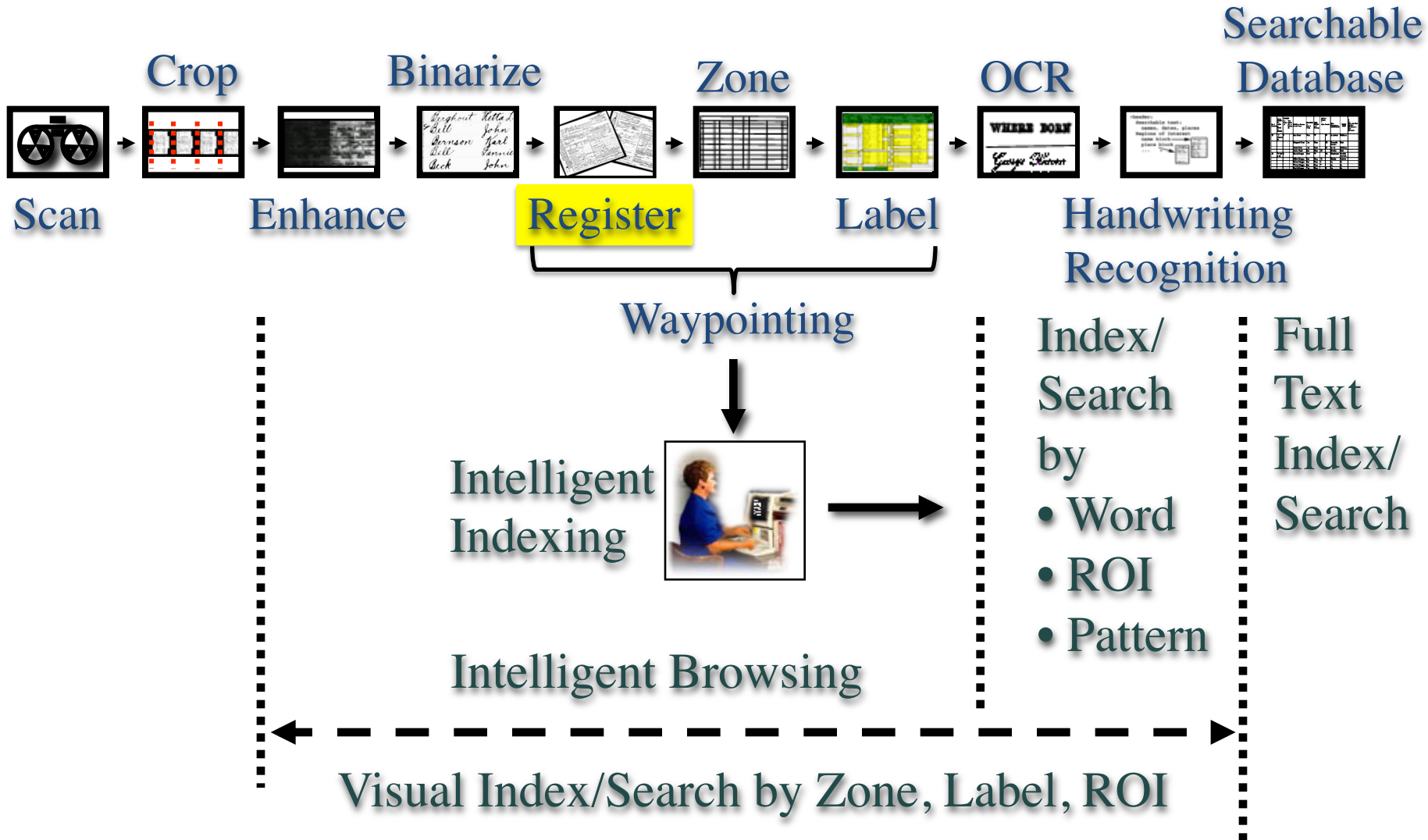
·
·



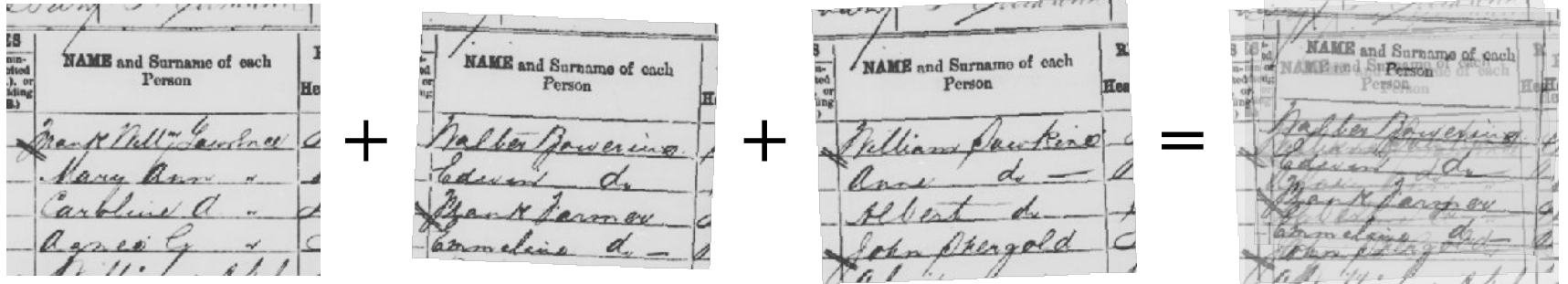
(f) Result ($S_1 + S_2 + S_3 + S_4$)

Recursive Otsu Intensity Parsing and Accumulation

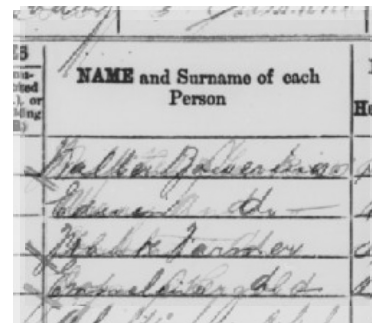
The Digital Microfilm Pipeline



Registration



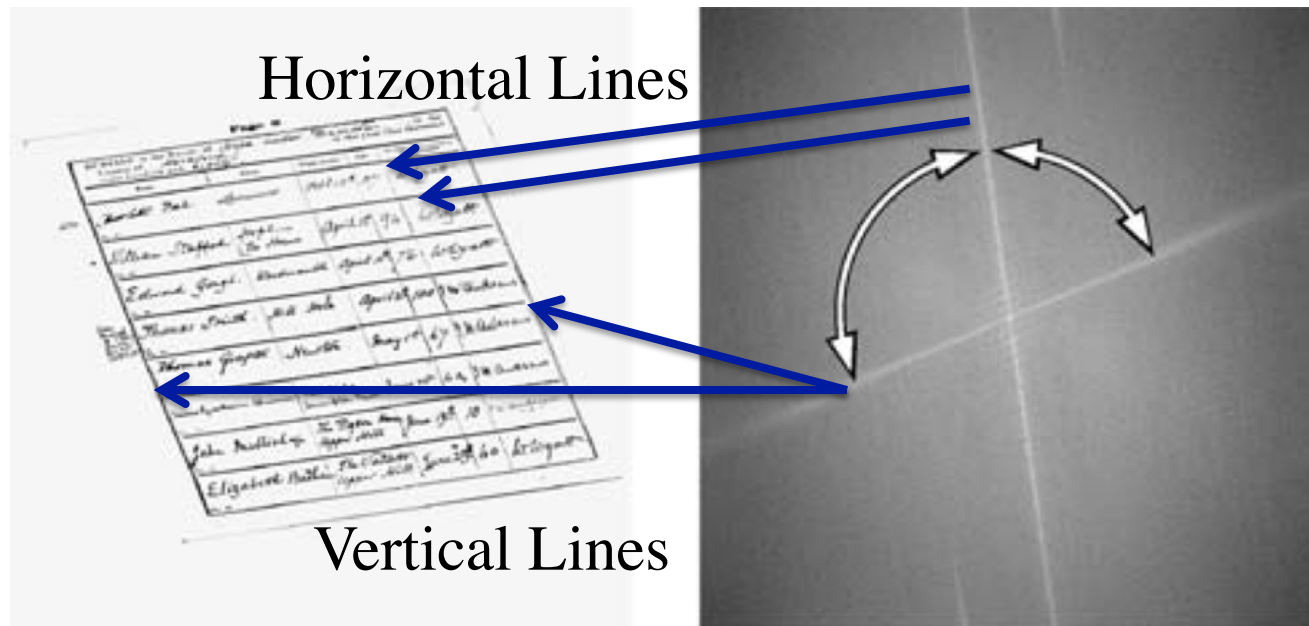
Not registered



Registered

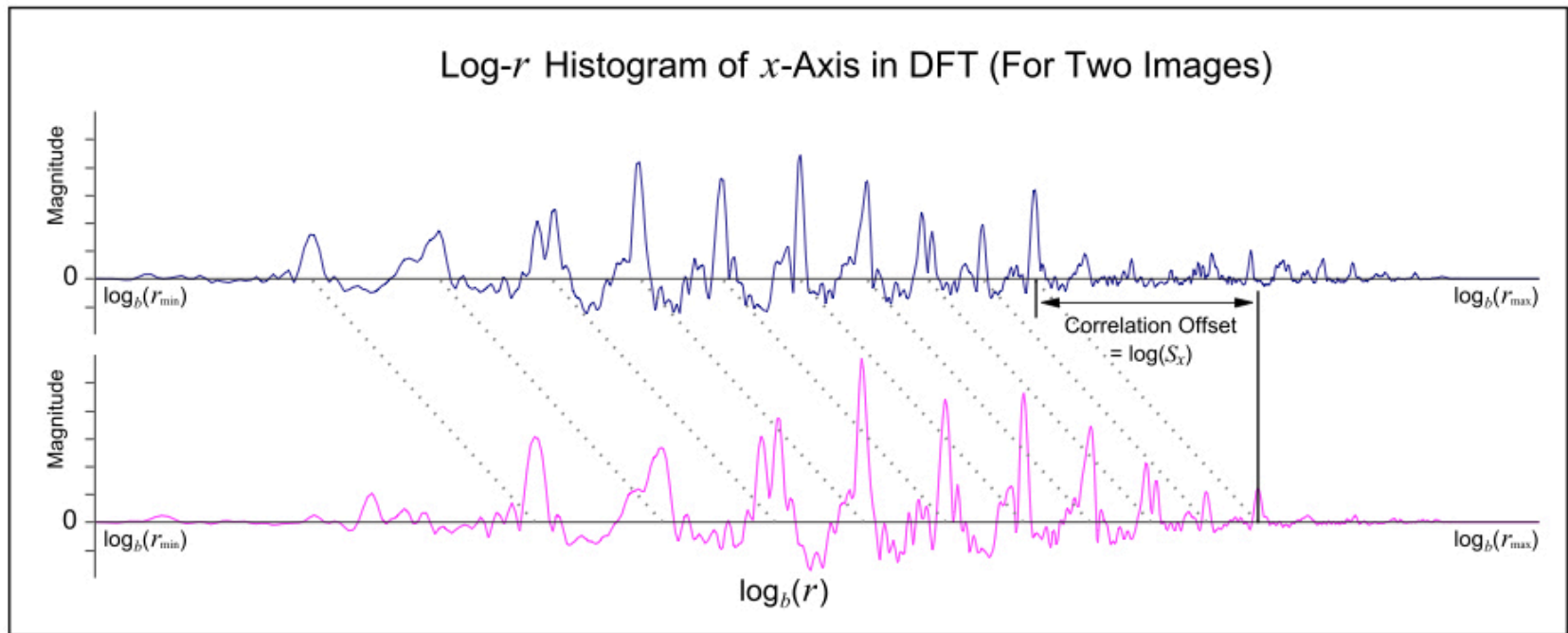
Documents must be aligned, rotated and scaled to register them

Registration – Hutchison 2003



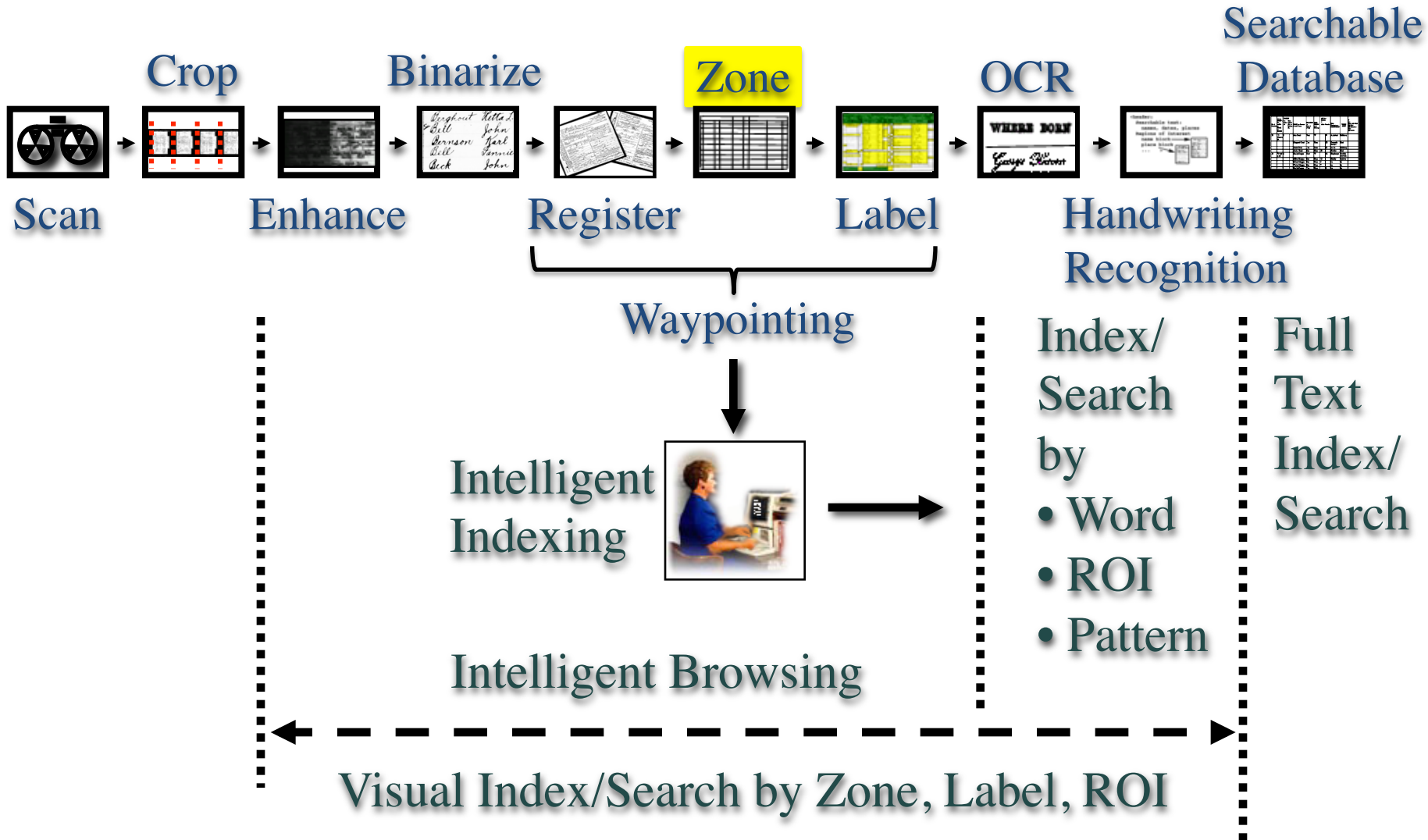
Fourier-Mellin Transform for detection of rotation and shear

Registration – Hutchison 2003



Fourier-Mellin Log Correlation, Scale Correction

The Digital Microfilm Pipeline



Segmentation & Zoning

- Nielson 2003



Page 2] The undermentioned Houses are situate within the Boundaries of the

Civil Parish [or Township] of	City or Municipal Borough of	Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of	Urban Sanitary District of	Rural Sanitary District of	Eccelesiastical Parish or District of		
<i>Fuggleslowe St Peter</i>			<i>Wotton</i>	<i>Wotton</i>					
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES Inhabited (A) or Building (B)	NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of	Rank, Profession, or OCCUPATION	WHERE BORN	If (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic
4	Southwood		John Parr	Servant	Mar	51	Müller	London Middlesex	
5	Witney Cottages	1	Edna A. Moore	Serv	Mar	28	Footman	Newcastle	
11	do	1	James Lake	Head	Mar	51	Dr. Buchanan	Greenwich Kent	
			Edna Humphard	Head	Mar	16	Dr. Gardner	Widhampton Wilt	
			Fanny Humphard	Wife	Mar	14		Widhampton Wilt	
			Emily Lane	Serv	Mar	14	Popul teacher	Widhampton Wilt	
			James	do	do	7	Scholar	Widhampton Wilt	
7	do	1	Edwin Humphard	Serv	Mar	10	Wid. Gardner	Fuggleslowe St Peter	
2	Witney Cottages	1	Alfred H. Smith	Head	Mar	31	Labourer	Widhampton Wilt	
			John	Wife	Mar	31	do	Widhampton Wilt	
			Harriet	Serv	Mar	13	Labourer	Widhampton Wilt	
			James H. do	Daughter	do	10	Scholar	Widhampton Wilt	
			Frank	do	do	2	do	do	
			Edna	do	do	11	do	Widhampton Wilt	
			Miss	Daughter	do	5	do	do	
			Fanny Lane	do	do	1	do	Widhampton Wilt	
6	do	1	John King	Head	Mar	30	Grass	Widhampton Wilt	
			Elizabeth	Wife	Mar	32		Widhampton Wilt	
			Caroline	Daughter	do	7	Domestic Serv. Manuf.	do	
			John	Serv	do	12	Aggrav. Serv.	do	
			Edwin	Serv	do	12	Scholar	Widhampton Wilt	
10	Church Lane	1	Maria Swickman	Servant	W	10	Book Bopute	Widhampton Wilt	
			Marian Staples	Serv	do	7	Servant	Widhampton Wilt	
h	Total of Houses...	17		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



Page 2

The undermentioned Houses are situate within the Boundaries of the

Civil Parish [or Township] of	City or 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			
<i>Fuggleslow & Peter</i>			<i>Wotton</i>	<i>Wotton</i>						
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES		NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of	Rank, Profession, or OCCUPATION	WHERE BORN	If (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic
		In habit-ed (A)	Un-habit-ed (B)							
4	Southwood			John Parr	Servant	Mar	51	Müller	London Middlesex	
				Eda A. Moore	Serv	Mar	28	Footman	Newell Kent	
5	<i>Witney Cottages</i>	1		<i>James Lake</i>	<i>Head</i>	<i>Mar</i>	<i>51</i>	<i>Dr. Buchanan</i>	<i>Greenwich Kent</i>	
11	<i>do</i>	1		<i>Edw. Humphard</i>	<i>Head</i>	<i>Mar</i>	<i>60</i>	<i>Dr. Gardner</i>	<i>Widhampton Wilt</i>	
				<i>Fanny Humphard</i>	<i>Wife</i>	<i>Mar</i>	<i>60</i>		<i>Widhampton Wilt</i>	
				<i>Emily Ann do</i>	<i>Daugh</i>	<i>Mar</i>	<i>14</i>	<i>Popul teacher</i>	<i>Widhampton Wilt</i>	
				<i>Ann do</i>	<i>do</i>	<i>do</i>	<i>7</i>	<i>Scholar</i>	<i>Widhampton Wilt</i>	
7	<i>do</i>	1		<i>Edw. Humphard</i>	<i>Head</i>	<i>Mar</i>	<i>60</i>	<i>Wid. Gardner</i>	<i>Fuggleslow & Peter W</i>	
2	<i>Wain Cottages</i>	1		<i>Alfred H. Smith</i>	<i>Head</i>	<i>Mar</i>	<i>34</i>	<i>Laborer</i>	<i>Wid. Gardner Wilt</i>	
				<i>Ann do</i>	<i>Wife</i>	<i>Mar</i>	<i>34</i>	<i>do wife</i>	<i>Widhampton Wilt</i>	
				<i>Harriet do</i>	<i>Serv</i>	<i>Mar</i>	<i>13</i>	<i>Laborer</i>	<i>Widhampton Wilt</i>	
				<i>Edw. H. do</i>	<i>Daughter</i>	<i>do</i>	<i>10</i>	<i>Scholar</i>	<i>Widhampton Wilt</i>	
				<i>Edw. do</i>	<i>Serv</i>	<i>do</i>	<i>2</i>	<i>do</i>	<i>do do</i>	
				<i>Ann do</i>	<i>do</i>	<i>do</i>	<i>11</i>	<i>do</i>	<i>Widhampton Wilt</i>	
				<i>Mrs. do</i>	<i>Daughter</i>	<i>do</i>	<i>5</i>	<i>do</i>	<i>do do</i>	
				<i>Fanny Ann do</i>	<i>do</i>	<i>do</i>	<i>1</i>	<i>do</i>	<i>Widhampton Wilt</i>	
6	<i>do</i>	1		<i>John King</i>	<i>Head</i>	<i>Mar</i>	<i>30</i>	<i>& Grocer</i>	<i>Widhampton Wilt</i>	
				<i>Elizabeth do</i>	<i>Wife</i>	<i>Mar</i>	<i>30</i>		<i>Widhampton Wilt</i>	
				<i>Ann do</i>	<i>Serv</i>	<i>do</i>	<i>12</i>	<i>Domestic Serv. Wimp.</i>	<i>do do</i>	
				<i>Ann do</i>	<i>Serv</i>	<i>do</i>	<i>12</i>	<i>Aggrav. Serv.</i>	<i>do do</i>	
				<i>Edw. do</i>	<i>Serv</i>	<i>do</i>	<i>12</i>	<i>Scholar</i>	<i>Widhampton Wilt</i>	
10	<i>Church Lane</i>	1		<i>Maria Swickman</i>	<i>Servant</i>	<i>W</i>	<i>60</i>	<i>Wid. Gardner</i>	<i>Wain Wilt</i>	
				<i>Marian Staples</i>	<i>Serv</i>	<i>do</i>	<i>4</i>	<i>Servants</i>	<i>Wotton Wilt</i>	
h	Total of Houses...	h		Total of Males and Females...		h	h			

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

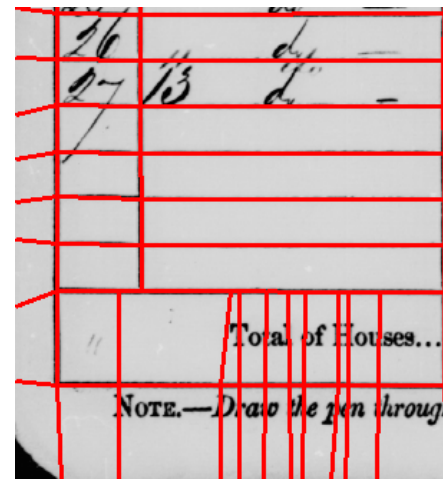
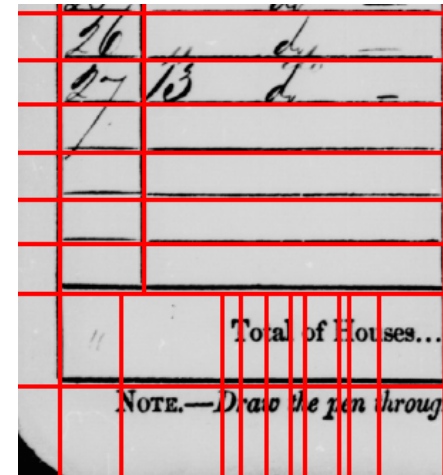
Local “Snapping”

- 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

$$s(x) = e^{\frac{-(p_g - x)^2}{2s^2}} lp_f(x)$$

- “Snap” to the location maximizing $s(x)$



Zoning



Page 2

The undermentioned Houses are situate within the Boundaries of the

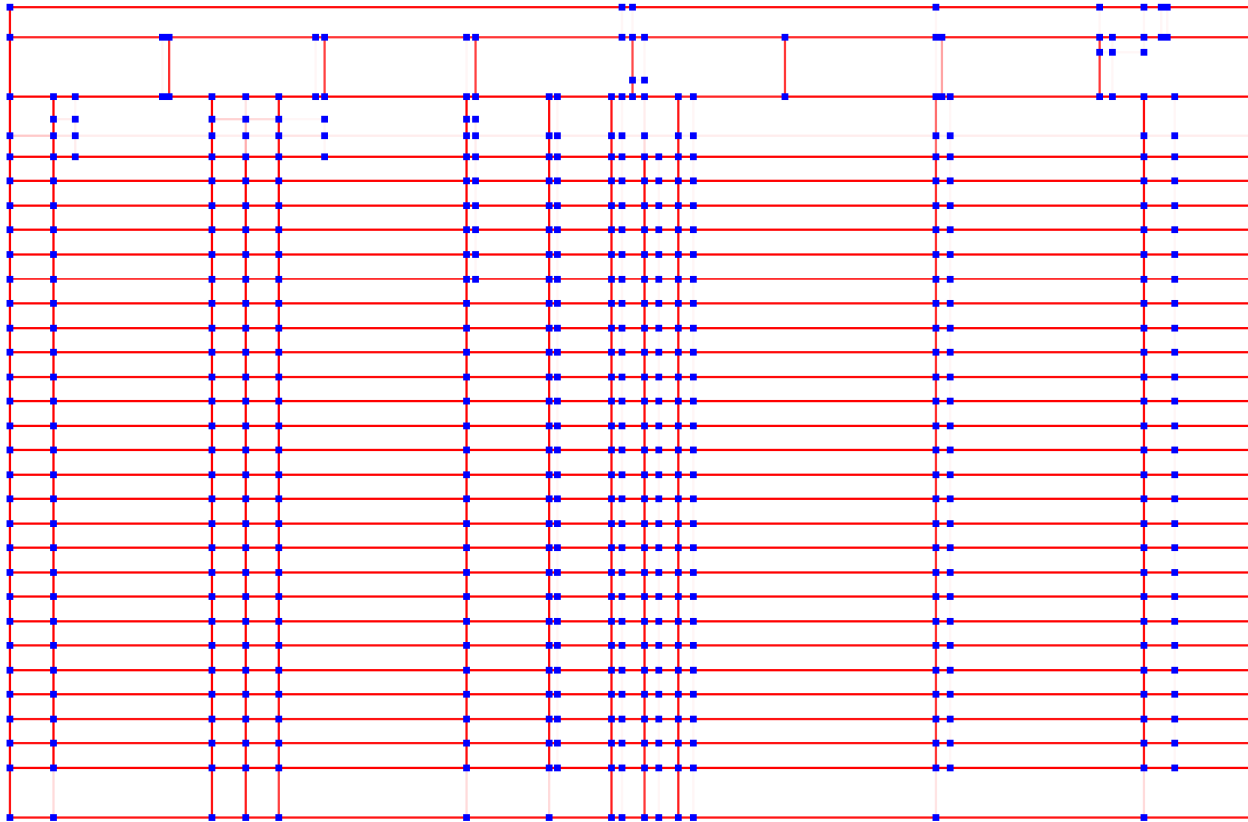
Civil Parish [or Township] of	City or Municipal Borough of	Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of	Urban Sanitary District of	Rural Sanitary District of	Religious Parish or District of	If	
R. Edmund.	Dahsburg	M. Edmund	S. S. S. S.	St. S. S.	S. S. S.			(1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic	
No. of Sched. and No. or NAME of HOUSE	ROAD, STREET, &c. or NAME of HOUSE	HOUSES Inhabited (1) or (2)	NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of Males Females	Rank, Profession, or OCCUPATION	WHERE BORN	
8	10 St.	1	Charles Holloway	Head	Mar	40	Jan Keeper	St. S. S.	
9	10 St.	1	Edward H. d.	Wife	Mar	40	Assist. S. S.	St. S. S.	
			Edna d.	Daughter	Mar	18	d. d.	St. S. S.	
			William d.	Son	Mar	17	d. d.	St. S. S.	
			Frances d.	Daughter	Mar	16	d. d.	St. S. S.	
			William Cook	Son	Mar	16	d. d.	St. S. S.	
10	41 St.	1	Robert John	Head	Mar	40	W. S. S.	St. S. S.	
			Elizabeth	Wife	Mar	35	d. d.	St. S. S.	
			William	Son	Mar	12	d. d.	St. S. S.	
			Margaret	Daughter	Mar	11	d. d.	St. S. S.	
11	43 St.	1	John Hart	Head	Mar	40	W. S. S.	St. S. S.	
			Elizabeth	Wife	Mar	35	d. d.	St. S. S.	
			William	Son	Mar	20	W. S. S.	St. S. S.	
12	45 St.	1	Charles Foreman	Head	Mar	40	Schoolmaster	St. S. S.	
			Laura	Wife	Mar	30	d. d.	St. S. S.	
			Hannah	Daughter	Mar	12	d. d.	St. S. S.	
			John	Son	Mar	10	d. d.	St. S. S.	
			Robert	Son	Mar	8	d. d.	St. S. S.	
			Georgina	Daughter	Mar	7	d. d.	St. S. S.	
			Walter	Son	Mar	7	d. d.	St. S. S.	
			Laura	Daughter	Mar	6	d. d.	St. S. S.	
Total of Houses...		4	Total of Males and Females...		8/15				

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

Line detection and local refinement for document zoning

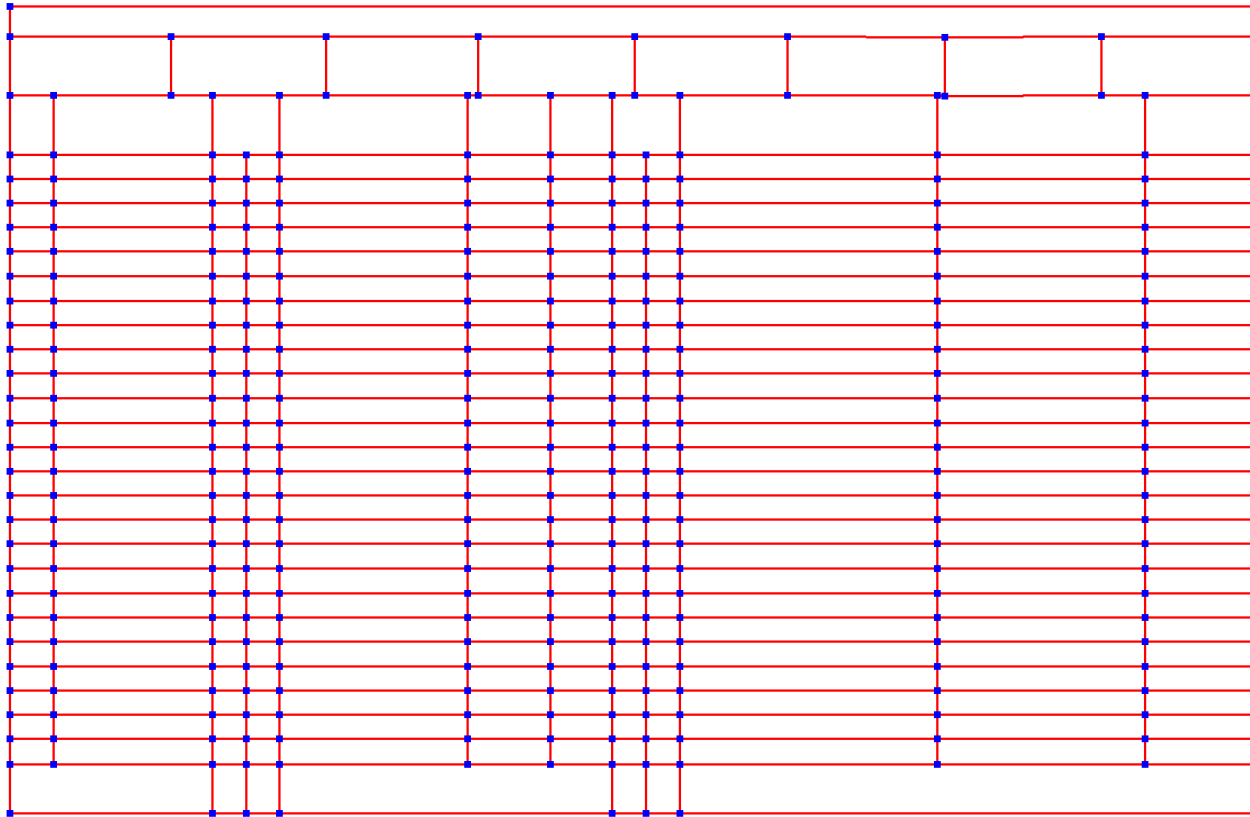


Zoning by Consensus



Consensus from 28 frames

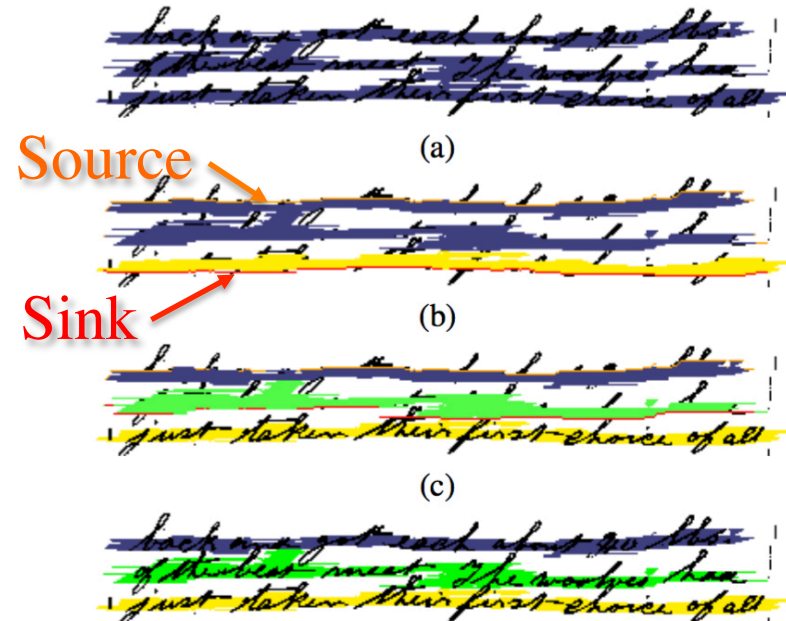
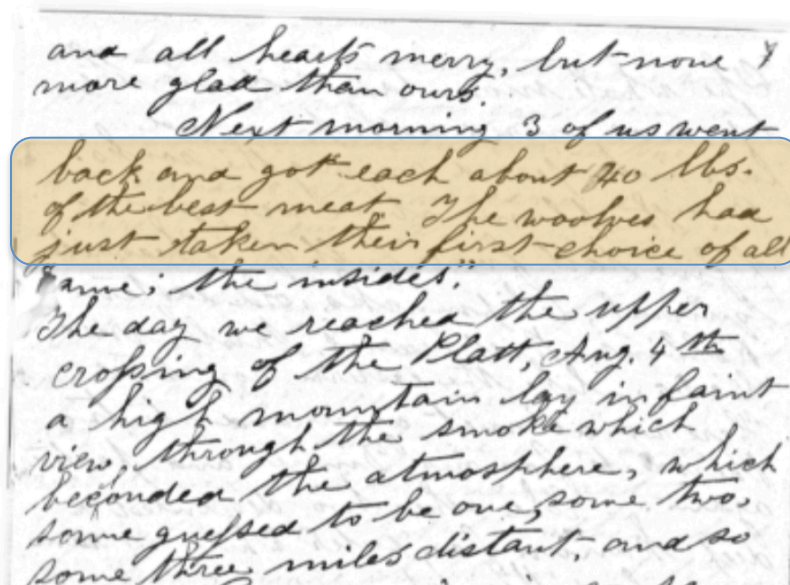
Zoning by Consensus



Final Zone Template

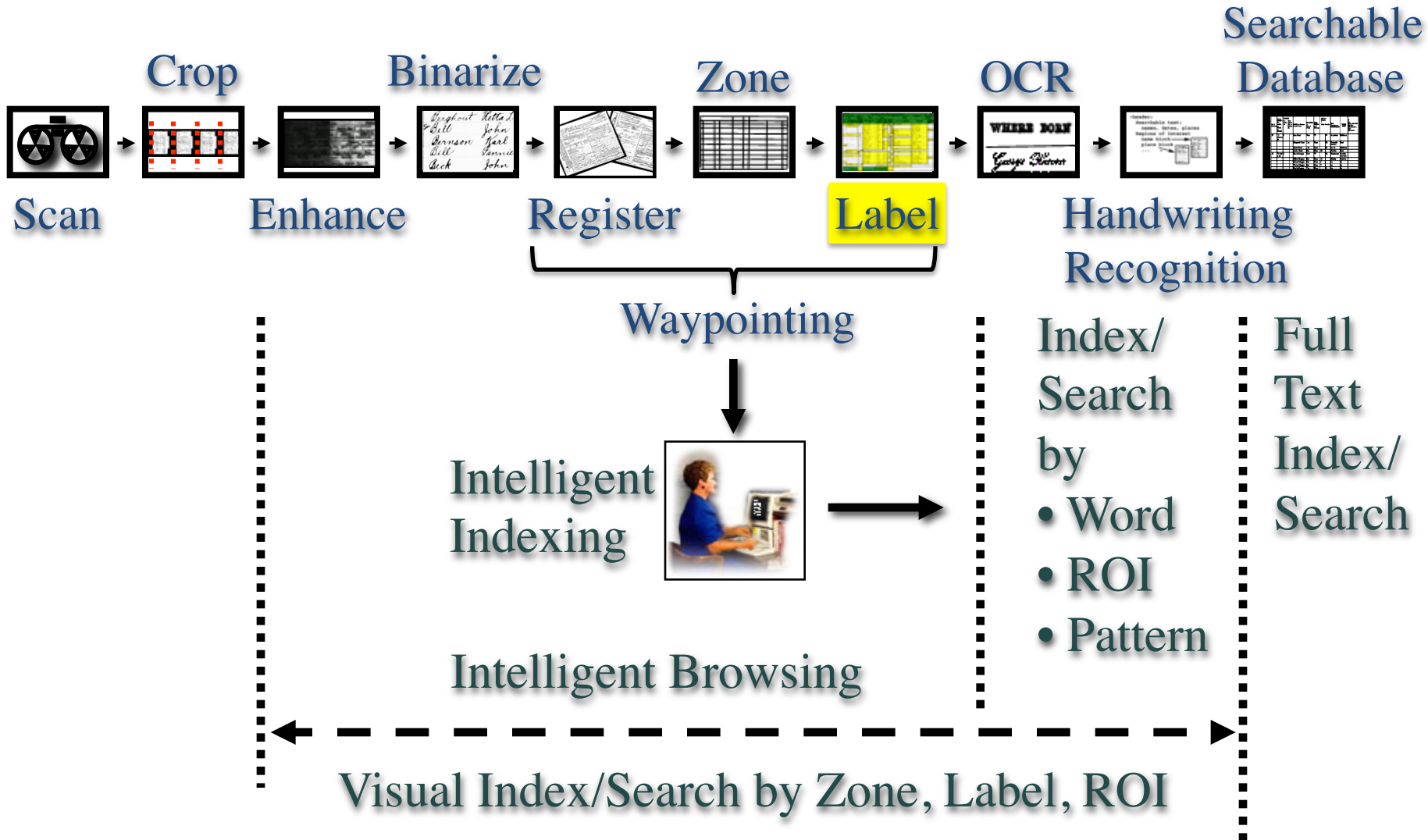
Free-form Line Detection

- Kennard, 2006



Recursive Graph Cut Line Segmentation

The Digital Microfilm Pipeline



Label Machine Print



The image shows three overlapping pages of a 1930 US Census form, specifically Page 1, titled "The Underenumerated House on whom within the Enumeration of the". Each page has a blue circle around the "WHERE BORN" field. Blue arrows point from these circles to the corresponding labels on the right.

Serial	NAME AND SURNAME of each Person	SEX	AGE	DATE OF BIRTH	Place of Birth	Color or Race	Marital Status	Occupation, Profession, or Occupation	Value of Real Estate	Value of Personal Estate
1	John Thomas	M	30	1898	Utah	W	M			
2	John Thomas	F	28	1902	Utah	W	M			
3	John Thomas	M	26	1904	Utah	W	M			
4	John Thomas	F	24	1906	Utah	W	M			
5	John Thomas	M	22	1908	Utah	W	M			
6	John Thomas	F	20	1910	Utah	W	M			
7	John Thomas	M	18	1912	Utah	W	M			
8	John Thomas	F	16	1914	Utah	W	M			
9	John Thomas	M	14	1916	Utah	W	M			
10	John Thomas	F	12	1918	Utah	W	M			
11	John Thomas	M	10	1920	Utah	W	M			
12	John Thomas	F	8	1922	Utah	W	M			
13	John Thomas	M	6	1924	Utah	W	M			
14	John Thomas	F	4	1926	Utah	W	M			
15	John Thomas	M	2	1928	Utah	W	M			

Three rectangular labels, each containing the text "WHERE BORN" in a bold, black, sans-serif font. Each label is enclosed in a light blue oval. Below each label is a red waveform pattern, resembling a barcode or a signal trace, which is used for machine reading.

Label Handwriting



The image shows three overlapping forms, likely from a historical document. The top form is partially obscured by the middle one, which is in turn obscured by the bottom one. Blue arrows point from the forms to the calligraphic examples on the right. The bottom form has a circled entry in the 'Name' column that reads 'Nitto Sakibara'.

Serial Number	Name	Rank	Position	Notes
1	Nitto Sakibara	Private	1st	
2	Nitto Sakibara	Private	2nd	
3	Nitto Sakibara	Private	3rd	
4	Nitto Sakibara	Private	4th	
5	Nitto Sakibara	Private	5th	
6	Nitto Sakibara	Private	6th	
7	Nitto Sakibara	Private	7th	
8	Nitto Sakibara	Private	8th	
9	Nitto Sakibara	Private	9th	
10	Nitto Sakibara	Private	10th	

Martin Nitto

[Illegible]

Nitto Sakibara

Labeled Image



[Example]

iii.

The undermentioned Houses are situate within the Boundaries of the												18			
Civil Parish [or Township] of <i>St. Mary.</i>		City or Municipal Borough of <i>Shrewsbury.</i>		Municipal Ward of <i>Wobb.</i>		Parliamentary Borough of <i>Shrewsbury.</i>		Town or Village or Hamlet of		Urban Sanitary District of <i>Shrewsbury.</i>		Rural Sanitary District of		Ecclesiastical Parish or District of <i>St. Michael.</i>	
No. of Schedule	ROAD, STREET, &c. and No. or NAME of HOUSE	HOUSES		NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of		Rank, Profession, or OCCUPATION	WHERE BORN	If				
		Inhabited (1, 2, or 3.)	Uninhabited (4.)				Male	Female			(1) Deaf & Dumb	(2) Blind	(3) Idiot	(4) Lunatic or Insane	
4	7, Charlotte St. (“Queen’s Arms”)	1		Michael Morrison	Head	Mar.	37		Licensed Victualler	Middlesex; Islington					
				Mary J. Do.	Wife	Mar.		29		Salep; Convent					
				Ellen Do.	Daus.			7mo		Salep; Shrewsbury					
				Elizabeth Morrison	Mother	W.		58	Annuitant	Salep; Shrewsbury		Lunatic			
				Ann Fox	Son.	Unm.		28	General Sev.	Leam; Andover					
				Catherine Doyle	Serv.	Unm.		24	Barmaid	Ireland					
5	8, Charlotte St.	1		Lambert Newton	Head	Mar.	39		Greens; (master, employing 2 men)	Cumberland; Weyton					
				Emma Do.	Wife	Mar.		36		Do.; Longlevon					
				William Do.	Son			12	Scholar	Salep; Ludlow					
				Leviella Do.	Daus.			9	Do.	Do.; Do.		Deaf and Dumb			
				George Dixon	Shopman	Unm.		19	Greens & Shopman	Middlesex; Norwiche					
				Jane Cook	Serv.	Unm.		22	General Sev.	Scotland					
6				James F. Phillips	Head	Mar.	41		Banker’s Clerk	Yorkshire; Leeds					
				Harriet Do.	Wife	Mar.		29		Do.; Bradford					
				Sophia White	Serv.	Unm.		16	General Sev.	Salep; Bridgenorth					
	9, Do.	1													
7	1, Bird Lane	1		William Hampton	Head	Mar.	72		Coach Driver	Staffordsh.; Bilston					
				Anno Do.	Wife	Mar.		69		Do.; Tamworth					
8	2, Bird Lane	1		Thomas Johnson	Head	Wid.	68		Botical Grees	Devon; Coniton					
				Henry Johnson	Son	Unm.		39	Organist	Salep; Shrewsbury		Blind			
				Emma Do.	Daus.	Unm.		41	Corset Maker	Middlesex; St. Pancras					
				Jane Palmer	Apprentice	Unm.		18	Corset Maker (Apprentice)	Salep; Ludlow					
9				Walter Campbell	Ledger	Unm.		23	Ship Carpenter, (out of employ)	Durham; Sunderland					
	Do.	2													
End of St. Michael Ecclesiastical District															
Total of Houses...		1	2	Total of Males and Females...		9	13								

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.	

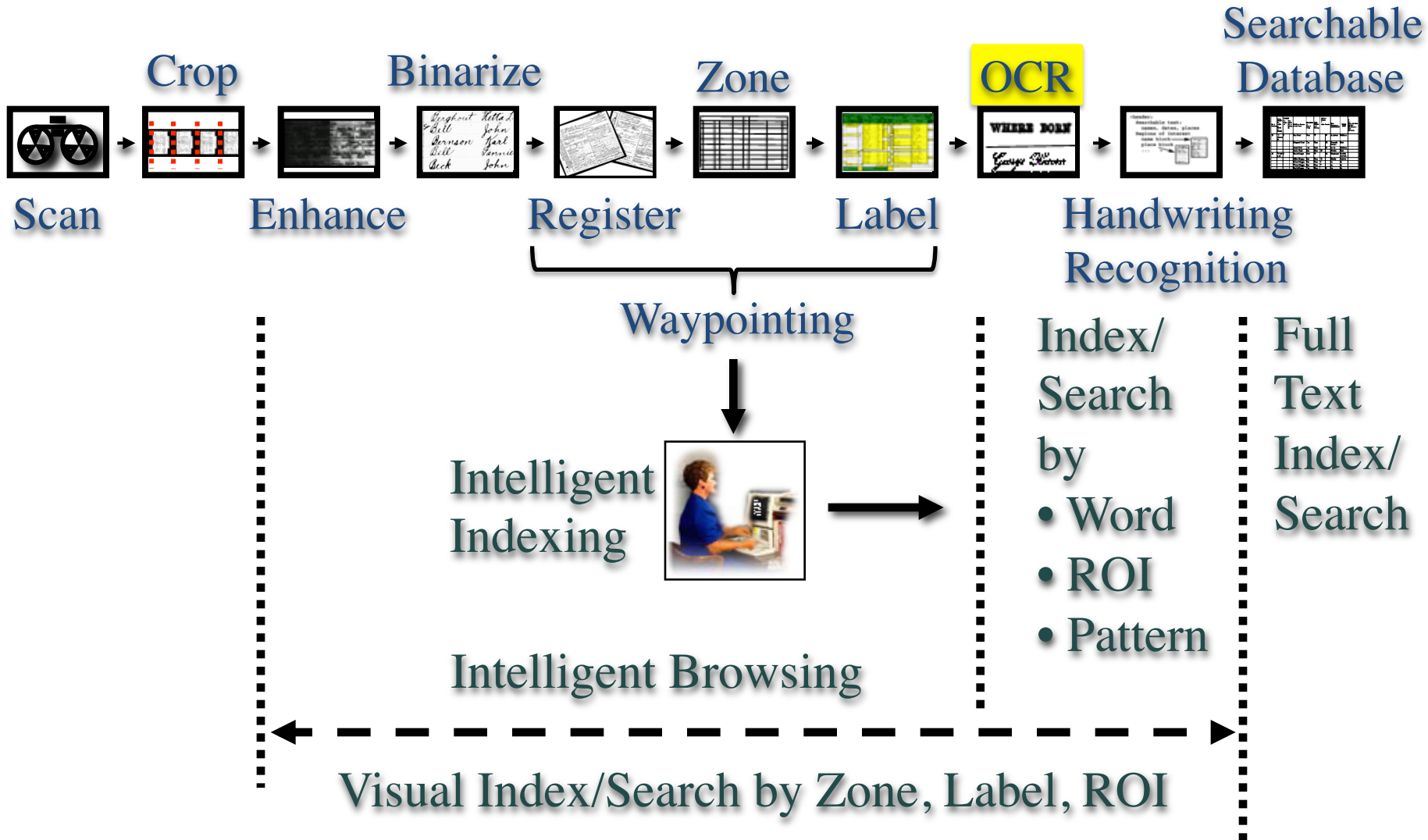
Zoned and Labeled Image



The undermentioned Houses are situate within the Boundaries of the [Page 1]

Civil Parish [or Township] of <i>Stagginton & Peter</i>		City or Municipal Borough of		Municipal Ward of		Parliamentary Borough of <i>Wilton</i>		Town or Village or Hamlet of <i>Wimston</i>		Urban Sanitary District of		Statistical Parish or District of <i>4</i>	
No. of Schedule	ROAD, STREET, &c. and No. or NAME of HOUSE	HOUSES Inhabited (A) or Building (B)	NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of Males Females	Rank, Profession, or OCCUPATION	WHERE BORN	(1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic				
1	<i>Lower Road</i>		<i>Miss Helen Gwenton</i>	<i>Wife</i>	<i>Married</i>	<i>26</i>	<i>Housewife</i>	<i>Wimston Wilts</i>					
			<i>Charles Ann Gwenton</i>	<i>Wife</i>	<i>Married</i>	<i>26</i>		<i>Wimston Wilts</i>					
			<i>John Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>5</i>		<i>Wimston Wilts</i>					
			<i>Edith Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>5</i>		<i>Wimston Wilts</i>					
			<i>Edith Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>3</i>		<i>Wimston Wilts</i>					
			<i>Edith Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>3</i>		<i>Wimston Wilts</i>					
			<i>Elizabeth Gwenton</i>	<i>Servant</i>	<i>Single</i>	<i>20</i>	<i>Wife</i>	<i>Wimston Wilts</i>					
			<i>Elizabeth Gwenton</i>	<i>Servant</i>	<i>Single</i>	<i>18</i>	<i>Housemaid</i>	<i>Wimston Wilts</i>					
			<i>Elizabeth Ann Moody</i>	<i>Servant</i>	<i>Single</i>	<i>16</i>	<i>Nurse</i>	<i>Wimston Wilts</i>					
2	<i>Hermitage</i>	1	<i>William Gwenton</i>	<i>Head</i>	<i>Married</i>	<i>50</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>George Gwenton</i>	<i>Servant</i>	<i>Single</i>	<i>70</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Wife</i>	<i>Married</i>	<i>42</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
3	<i>Lower Road</i>	1	<i>Richard Gwenton</i>	<i>Head</i>	<i>Married</i>	<i>40</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Wife</i>	<i>Married</i>	<i>30</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
3			<i>William Gwenton</i>	<i>Servant</i>	<i>Single</i>	<i>21</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
14	<i>Wilton House</i>	1	<i>William Gwenton</i>	<i>Head</i>	<i>Married</i>	<i>47</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Wife</i>	<i>Married</i>	<i>40</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>27</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>25</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>19</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Edith Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>10</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>Ann Gwenton</i>	<i>Daughter</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					
			<i>William Gwenton</i>	<i>Son</i>	<i>Single</i>	<i>7</i>	<i>Wine Merchant</i>	<i>Wimston Wilts</i>					

The Digital Microfilm Pipeline



Machine-Printed OCR

A solved problem?

Archivo de la Parroquia

— de —

SANTA MARIA de los BAÑOS

— JUBILICO —

BAUTISMO DE HIJOS LEGITIMOS

VOLUMEN NUM. 104 1897 1899

Academia Mexicana de Genealogía y Heráldica

NAME and Surname of each Person

PLACE OF ABODE

Street, avenue, road, etc.	House number (in cities or towns)	Number of dwelling house in order of visitation	Number of family in order of visitation
1	2		

LOCATION.				NAME	RELATION.
IN CITIES.		Number of dwelling house, in the order of visitation.	Number of family, in the order of visitation.		
Street.	House Number.			1	2
				of each person whose place of abode on June 1, 1900, was in this family. Enter surname first, then the given name and middle initial, if any. INCLUDE every person living on June 1, 1900. Omit children born since June 1, 1900.	Relationship of each person to the head of the family.
		1	2	<i>Carlos Lucas family</i> <i>1900 over from sheet 8</i>	4

FILMADO POR LA SOCIEDAD GENEALOGICA		LOCALIDAD QUE CUBRE EL REGISTRO	
EN MEXICO			
IAFO	REDUCCION (X)		
LAGO MEJIA M	40		
FE FILMACION	EXPOSICION		
15 1984	20		
NO DE EMULSION	NUMERO DE CAMARA		
6-452-1	51896		
NO DEL PROYECTO	NUMERO DEL ROLLO		
(C-2-059	1		

PARROQUIA DE SANTIAGO
TEMAPACHE VER. DIOCESIS
DE TUXPAM VER. MEX.

TITULO DEL REGISTRO

BAUTISMOS

ITEM 1

Must detect Machine Print before recognizing it

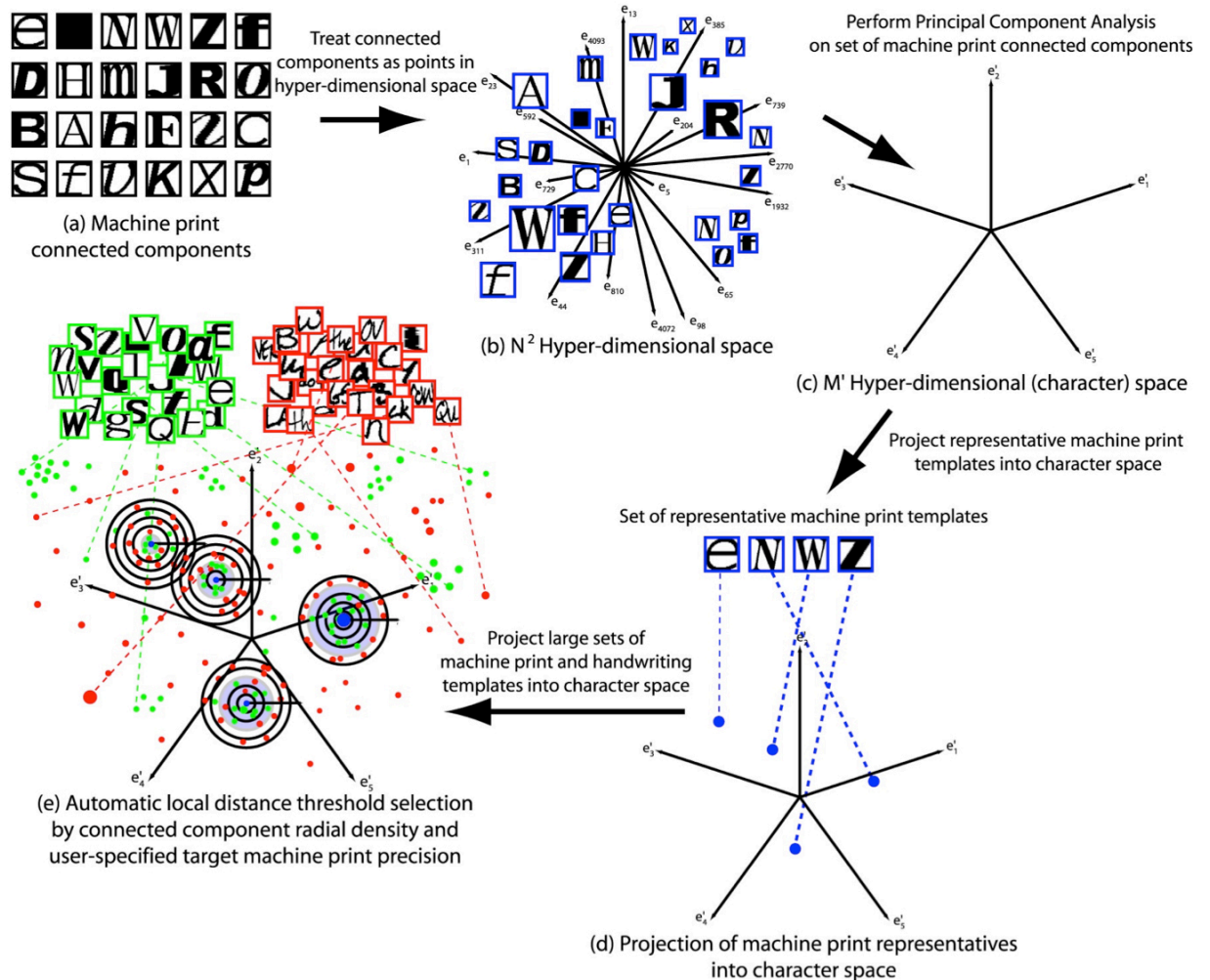
PCA Classification

Machine Print

- 26 fonts
- 273K CCs
- 6000 Templates

Hand Print

- 2100 writers
- ~600K CCs





Precision

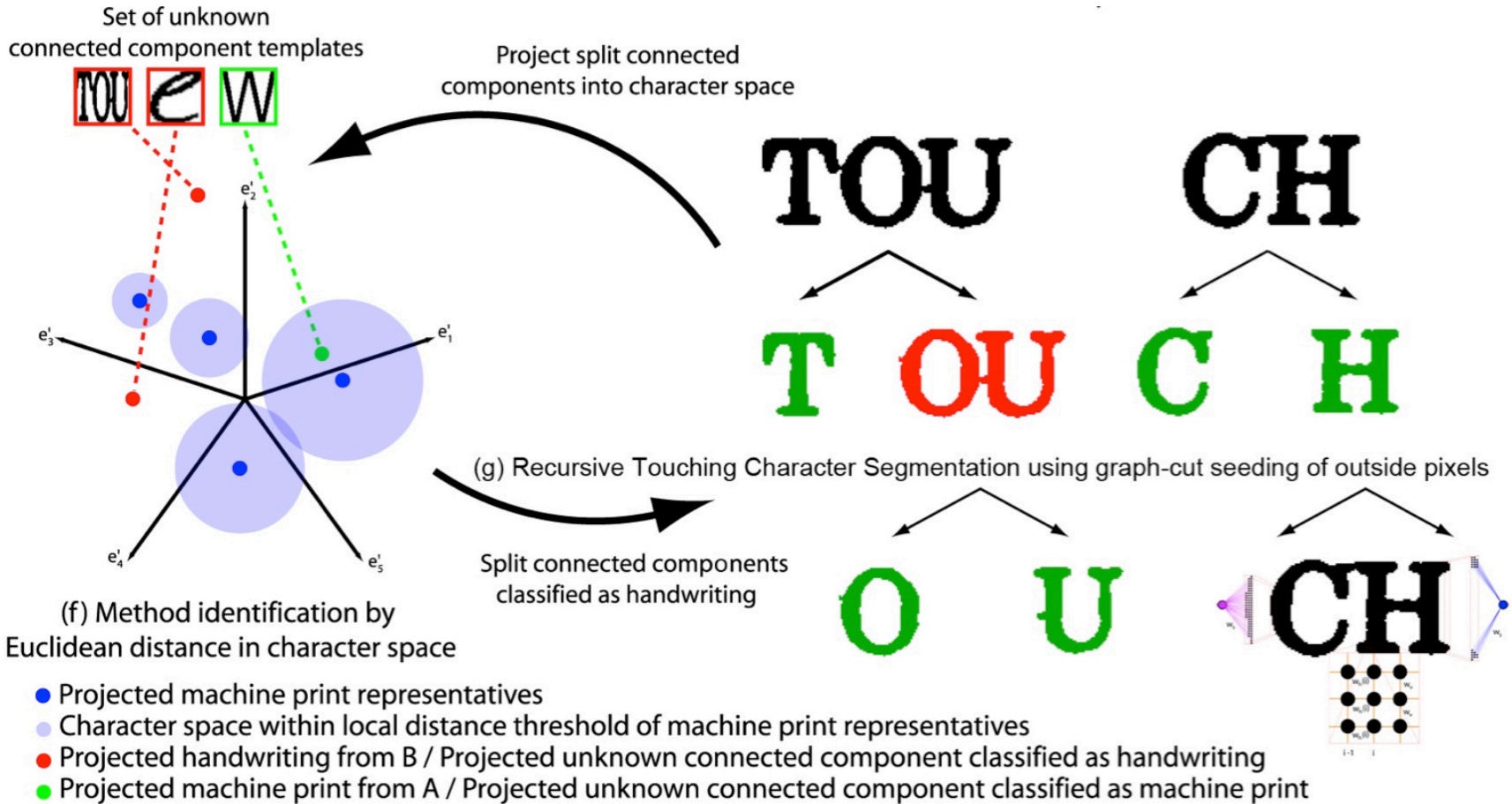
Machine Print – 94%

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

||

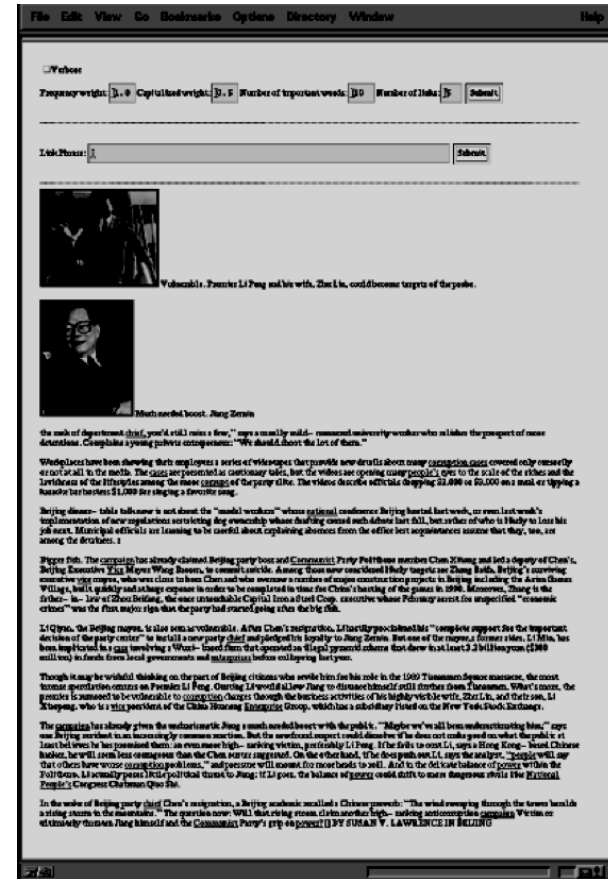
99.8% compared to 99.94% (human)

Machine-Printed OCR – Barzee '96

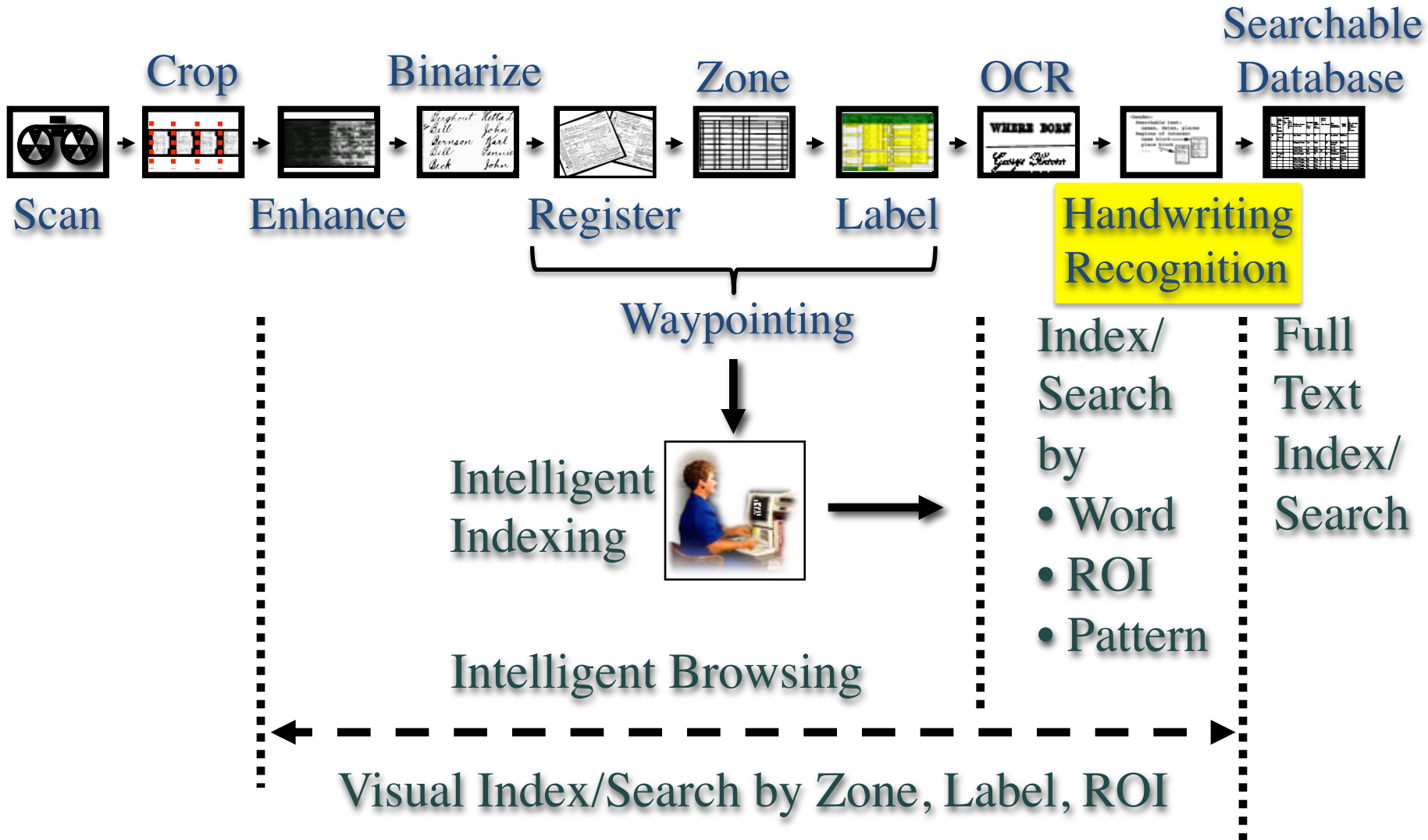
- Posting Paper on the Web



- 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



Segmentation in tabular documents – Davis 2015



← Pattern Thomas →

—	Anna
—	Pattern Thomas
—	Robert Campbell
—	Elizabeth
—	Ralph
—	Robinson
—	Robert Mariah
—	William Leece
—	Edith
—	Just
—	Edith

Segmentation in tabular documents



← Patton Thomas →

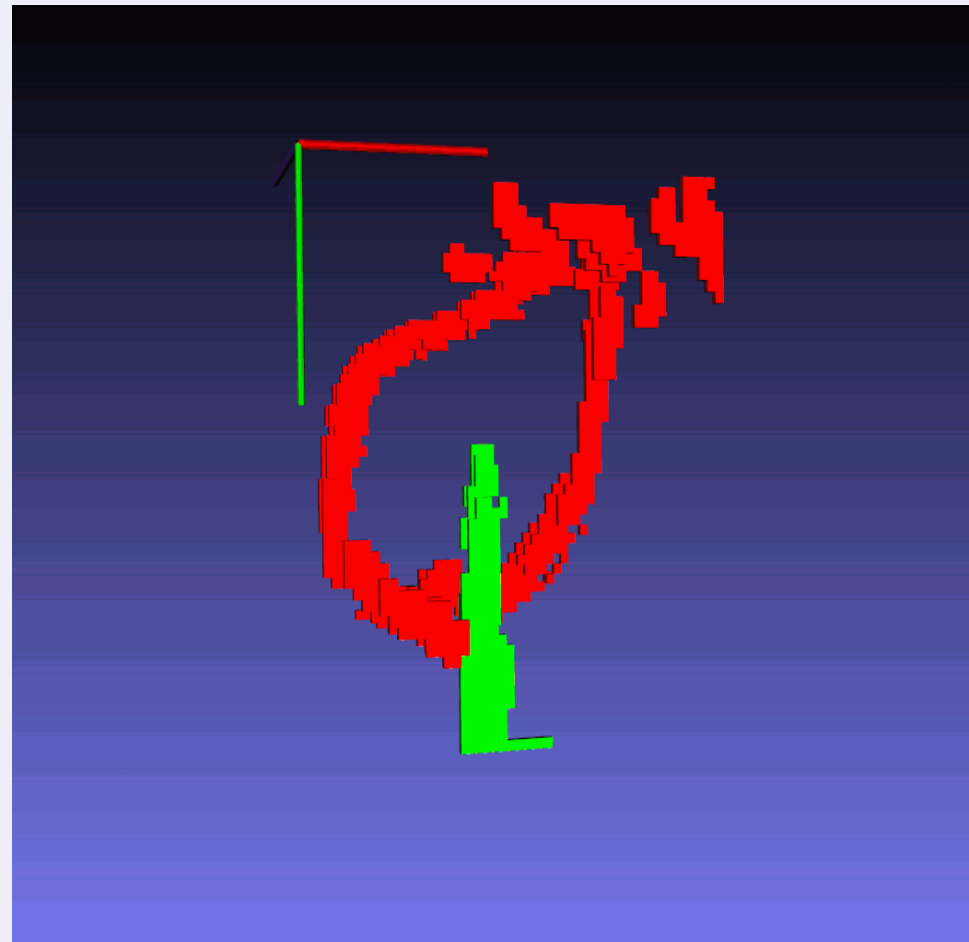
Patton Thomas
Patton George
— Higley
20917

Patton Thomas
Patton George
— Higley
20917

— Anna
Patton Thomas
Patton George
— Higley
— High
— Johnson
Patton Maria
Williams Lee
— Gill
— Jack
— Edith

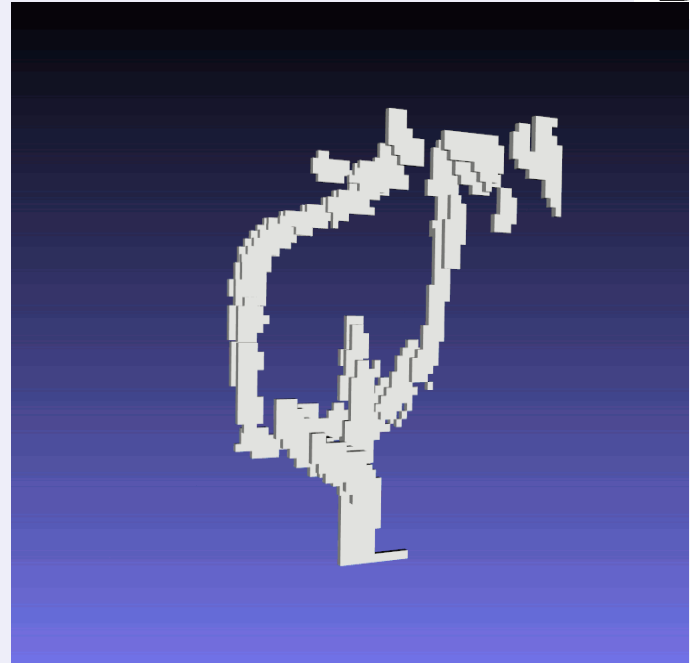
3D cut overview

- 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



chief

Monday

Handwriting Recognition



- *using Word Warping*

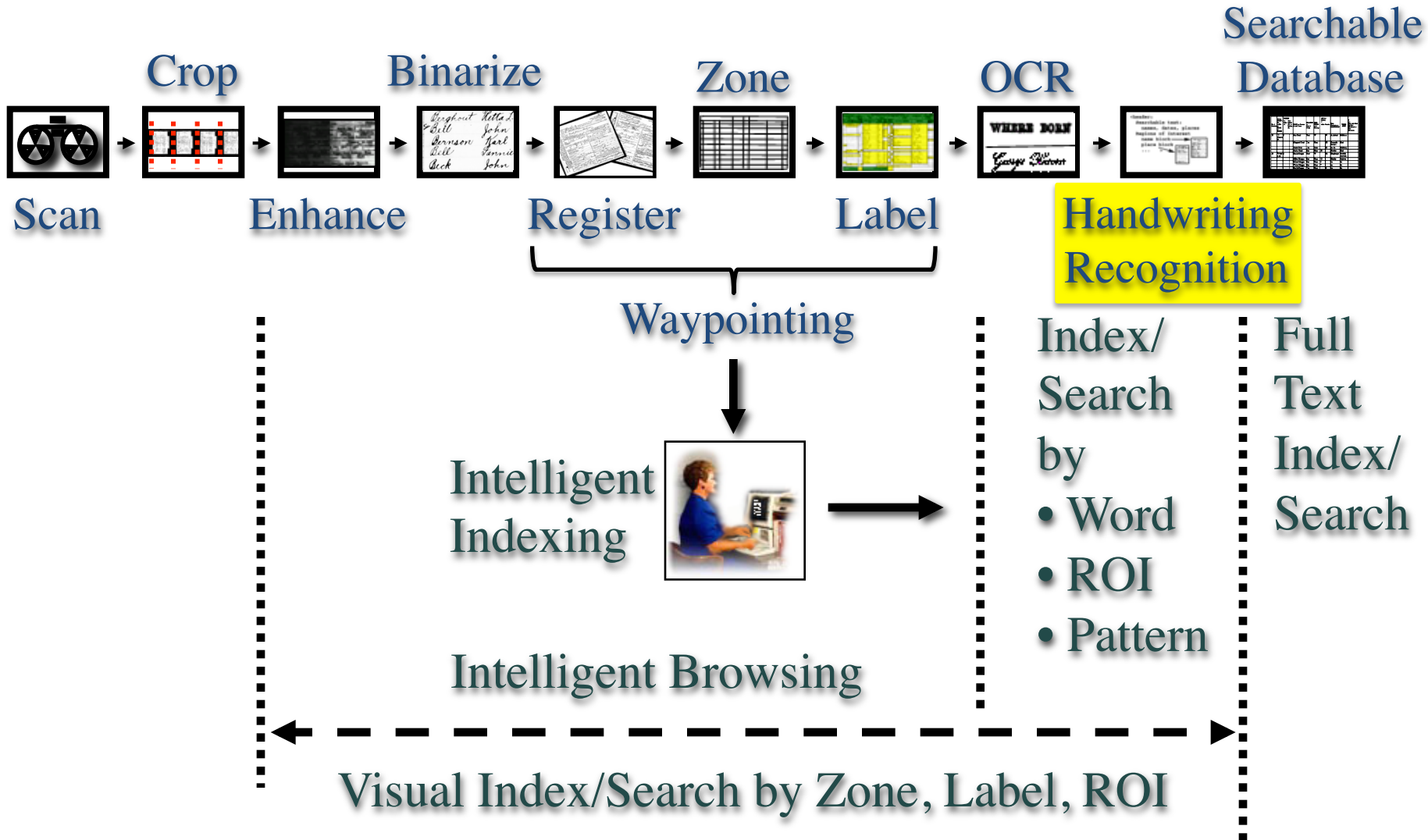
Single author
in vocabulary – 91-92%

Single author
in vocabulary and
out of vocabulary – 69-72%

Multi-author *in vocabulary* – 66%

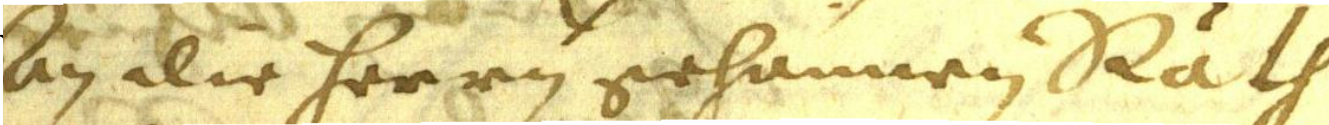
Multi-author *in+out* of vocab – 56%

The Digital Microfilm Pipeline



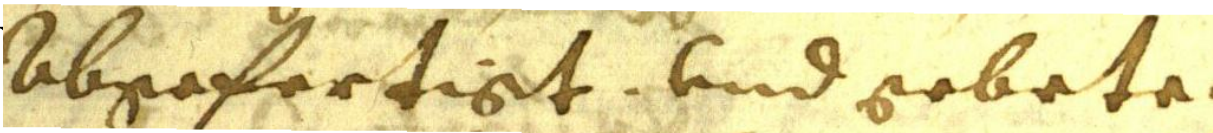
Handwriting Recognition with Deep Learning Networks – '16

Wigington, Davis, Stewart

- A horizontal strip of a handwritten document in cursive script, showing the words "An die herrn geheimen Rāth".
 - Expert Transcription:
 - An die herrn geheimen Rāth
 - Our system:
 - An die herrn geheimen Rāth

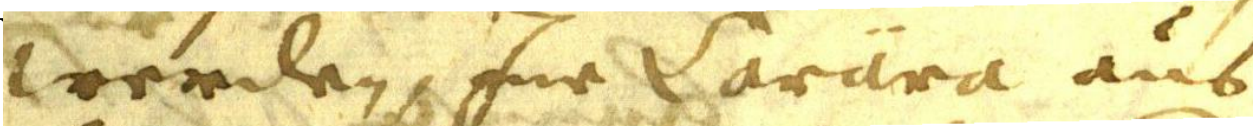
Handwriting Recognition with Deep Learning Networks – '16

Wigington, Davis, Stewart

- A horizontal strip of a handwritten document in a cursive script. The text is written in brown ink on aged, yellowish paper. The words are "Abgefertigt. vnd gebeten."
- Expert Transcription:
- Abgefertigt. vnd gebeten
- Our system:
- Abgefertigt. vnd gebete ■

Handwriting Recognition with Deep Learning Networks – '16

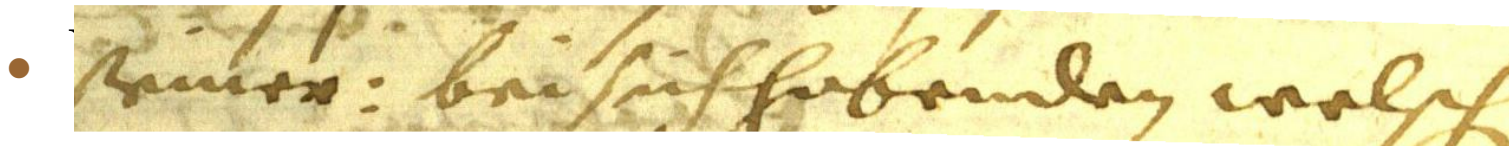
Wigington, Davis, Stewart

- A horizontal strip of aged, yellowed paper with handwritten text in a cursive script. The text is difficult to read due to the handwriting and the age of the paper.

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

Handwriting Recognition with Deep Learning Networks – '16

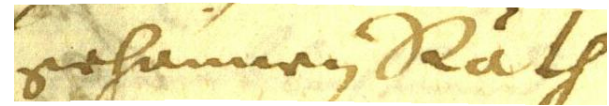
Wigington, Davis, Stewart



- Expert Transcription:
- seiner: bei sich habenden welsch
- Our system:
- seiner. bei sich habenden welsch

Handwriting Datasets


- German (15th-18th Century)

A sample of German handwriting from the 15th-18th century, showing a cursive script on aged paper. The text is partially obscured by a yellow highlight.

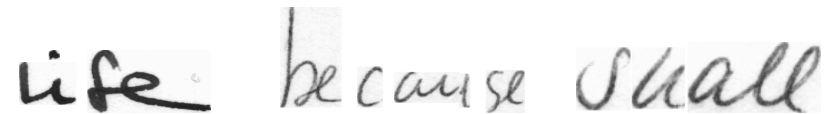
- George Washington

A sample of George Washington's handwriting, showing the words "properly company" in a cursive script.

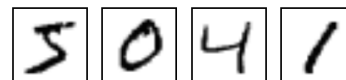
- Smith Journals

A sample of handwriting from the Smith Journals, showing the words "retired practised" in a cursive script.

- IAM (Multi-Author)

A sample of handwriting from the IAM dataset, showing the words "life because shall" in a cursive script.

- MNIST (Numbers)

A sample of handwriting from the MNIST dataset, showing the numbers "5041" in a cursive script, with each digit enclosed in a small box.

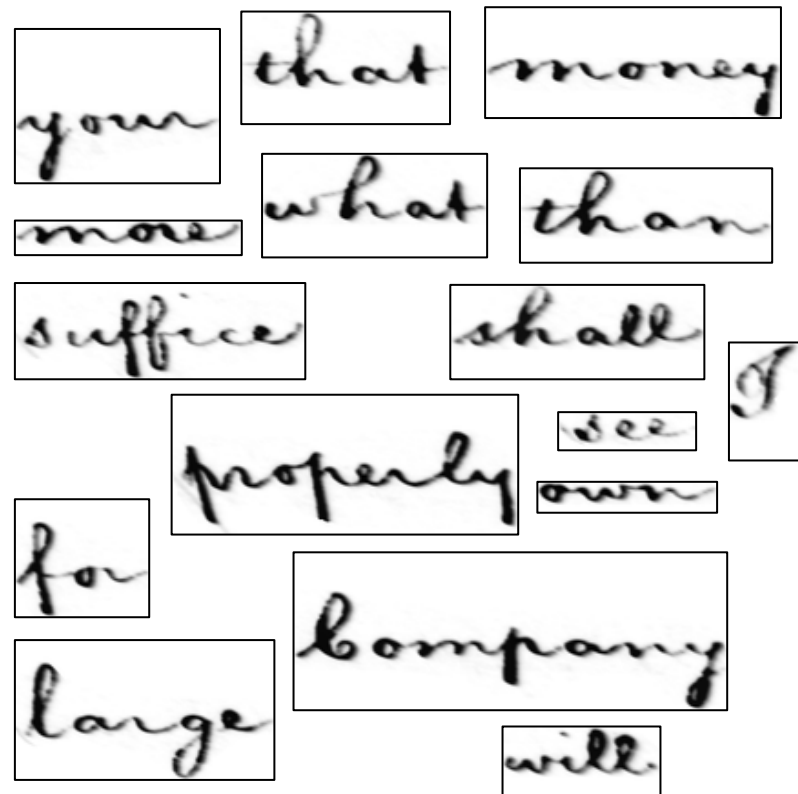
George Washington



Single Writer

- 99.3% Character Accuracy
- 98.4% Word Accuracy

Word Examples



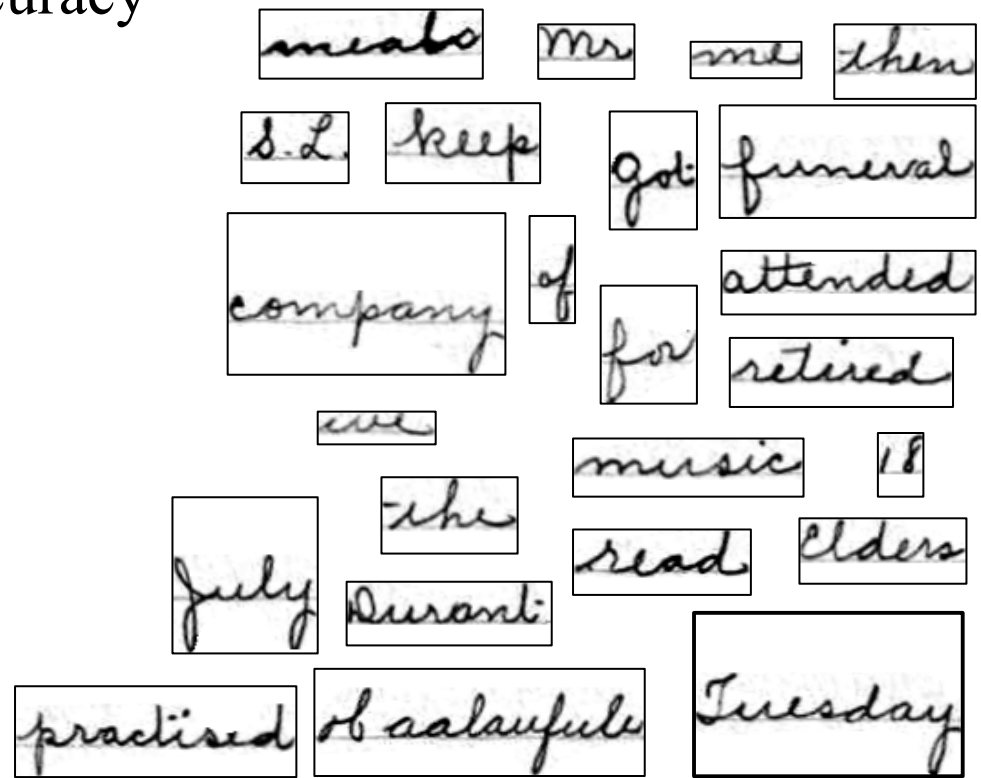
Smith Journals



Single Writer

- 98.3% Character Accuracy
- 97% Word Accuracy

Word Examples





IAM Handwriting Dataset

657 Writers
- 97.2% Character
Accuracy
- 95.03% Word
Accuracy

Word Examples





MNIST Handwritten Digits

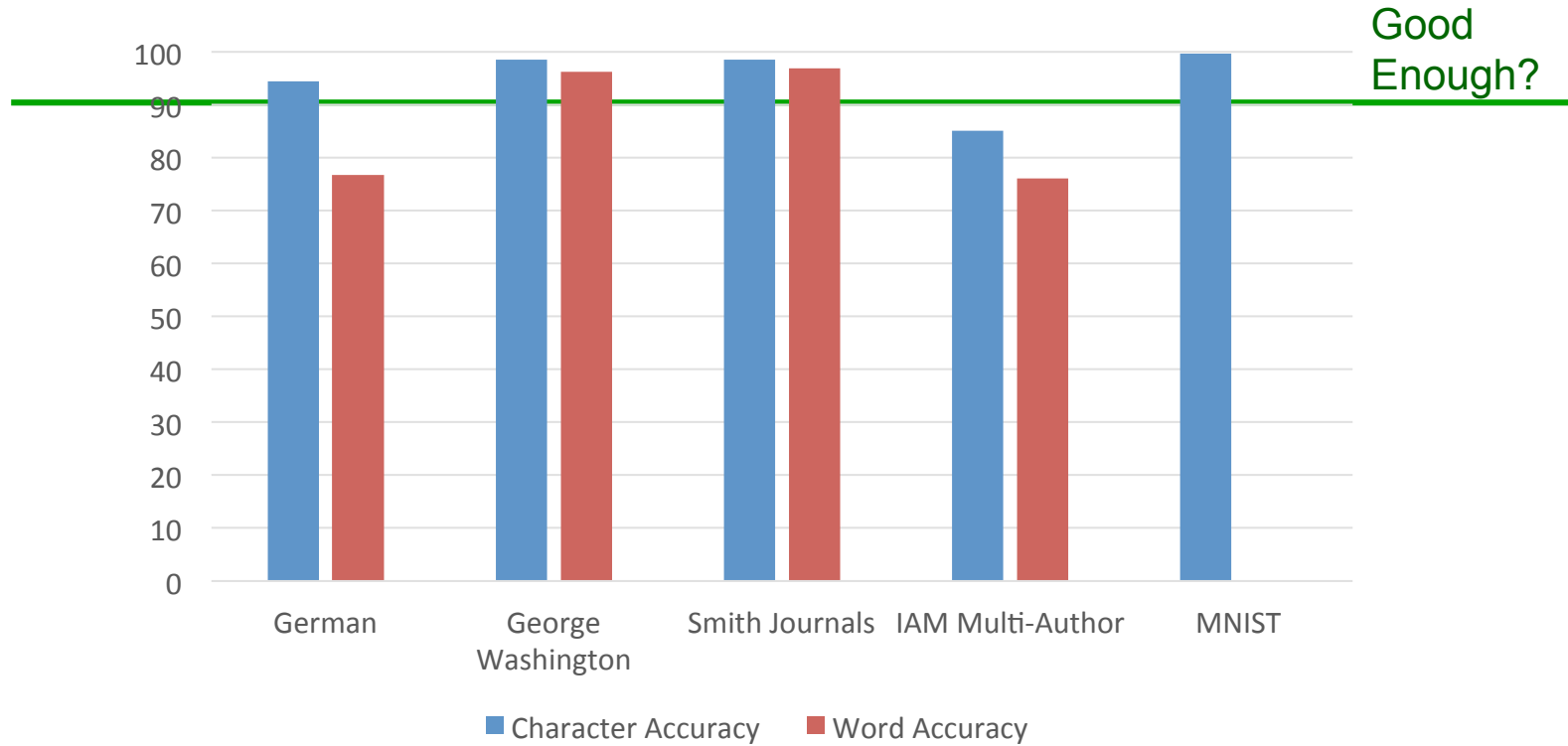
500 Writers
- 99.62% Character
Accuracy

Examples





Handwriting Datasets



But what about



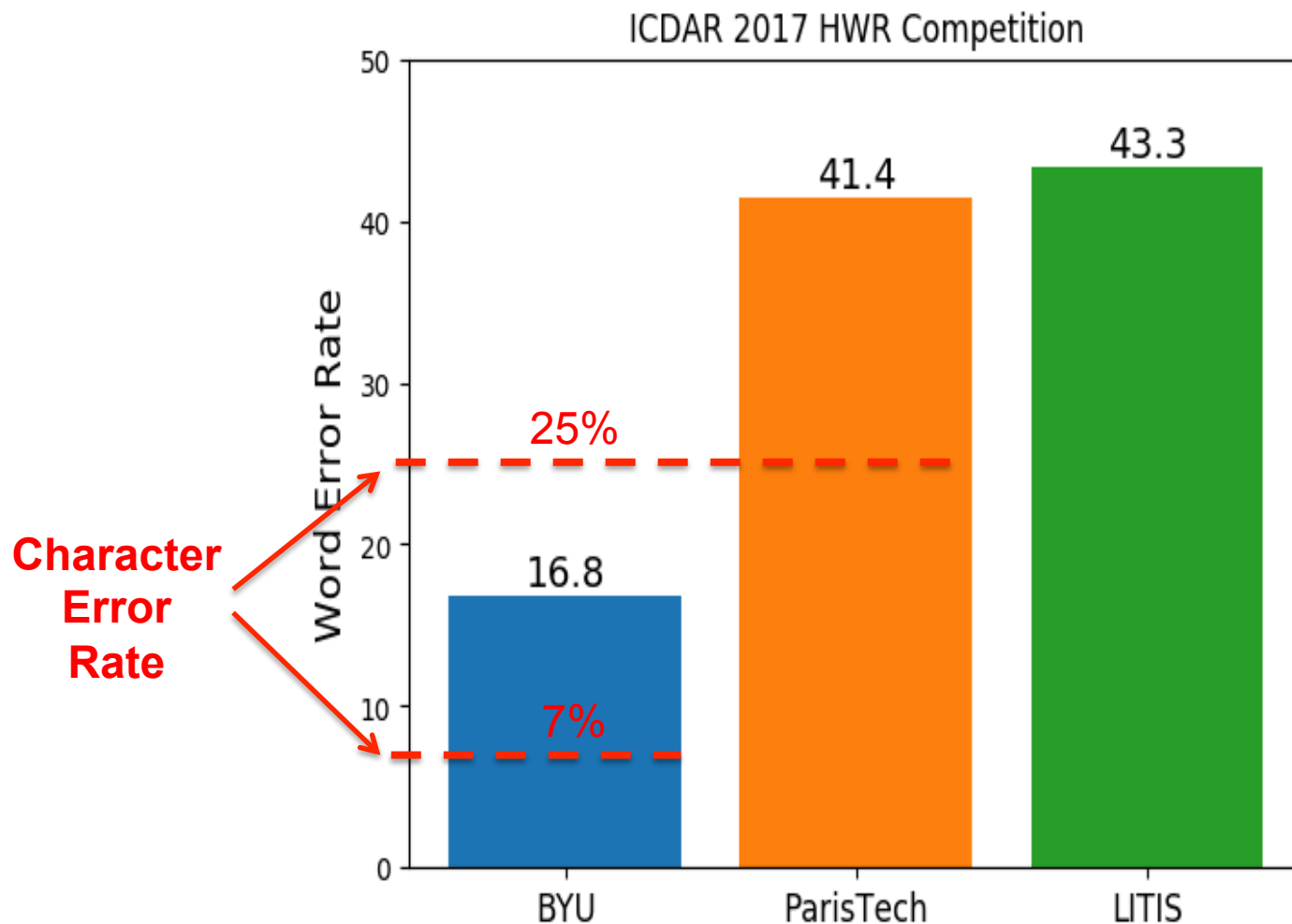


The READ Dataset

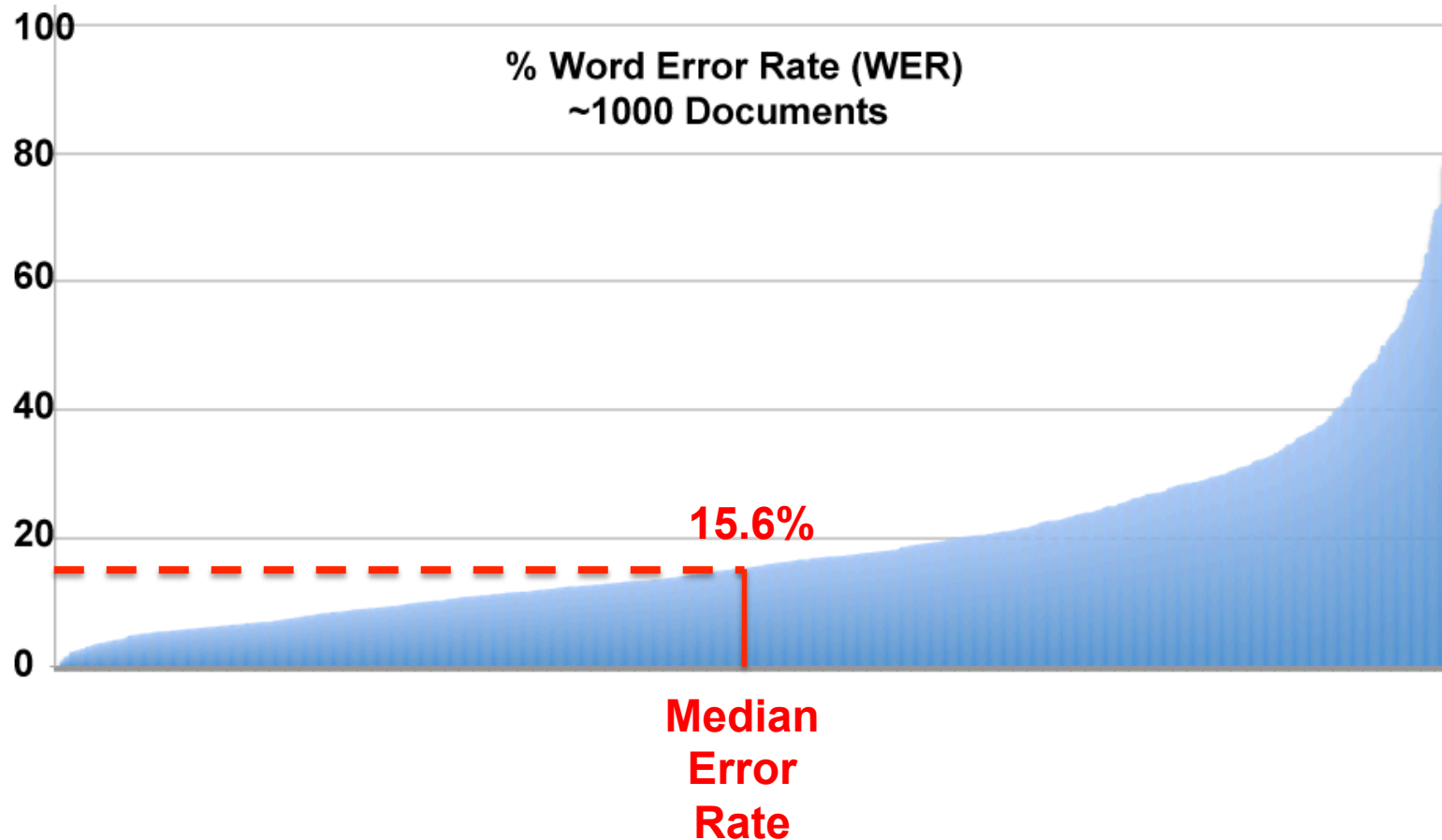
- ICDAR 2017 Competition

- 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

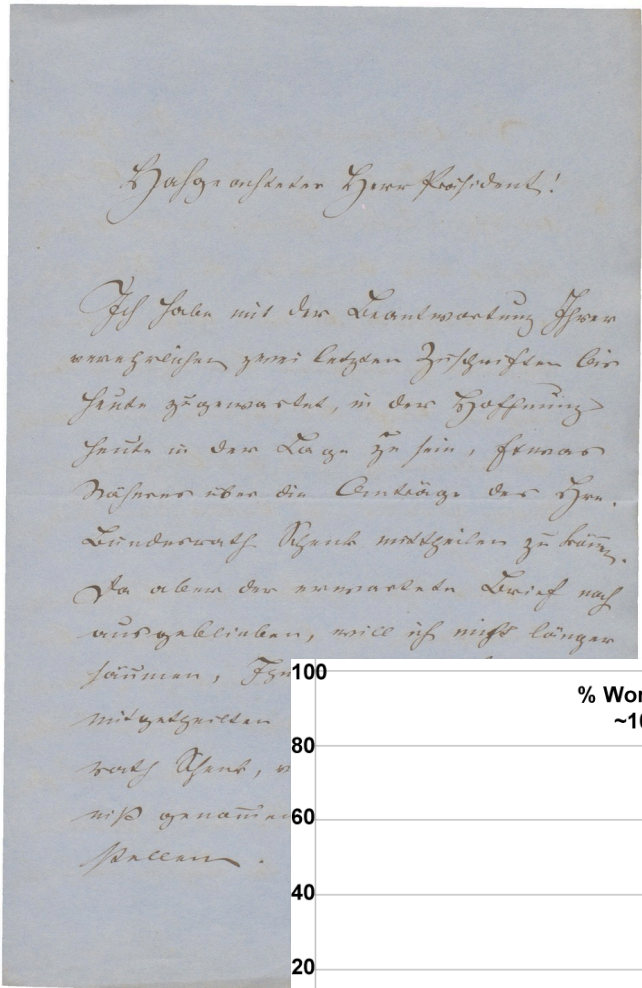
Handwriting Recognition:



Handwriting Recognition:



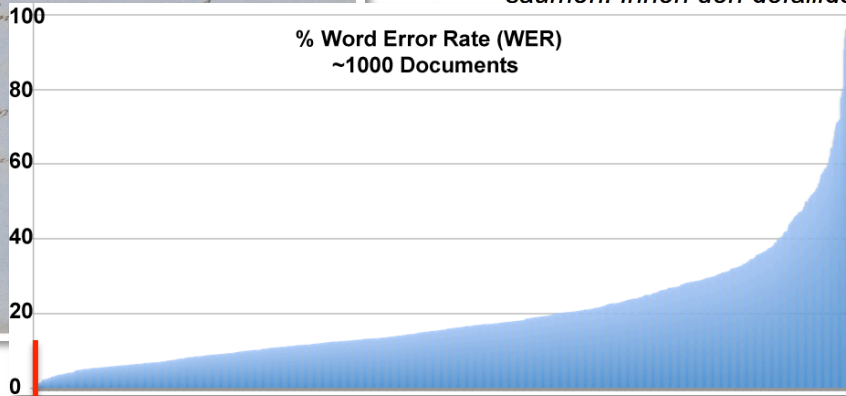
0 Errors German Handwriting Recognition



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 aefälligst

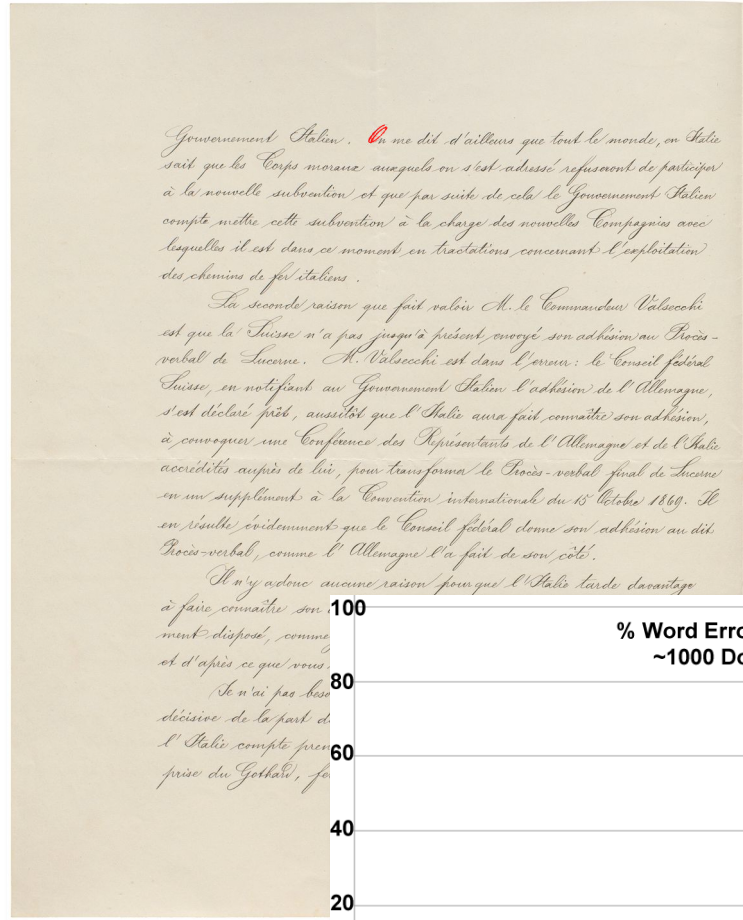


0%

Bundes-
rath Schenk
Kanzlei

French Handwriting Recognition

0.1% CER .77% WER



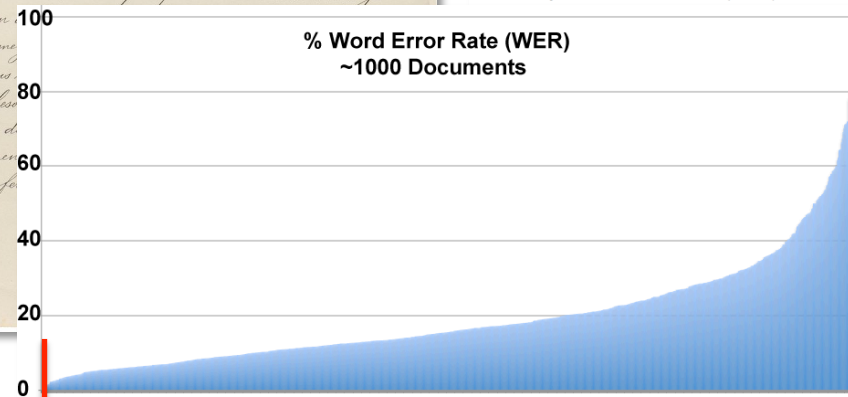
Automated Transcription

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

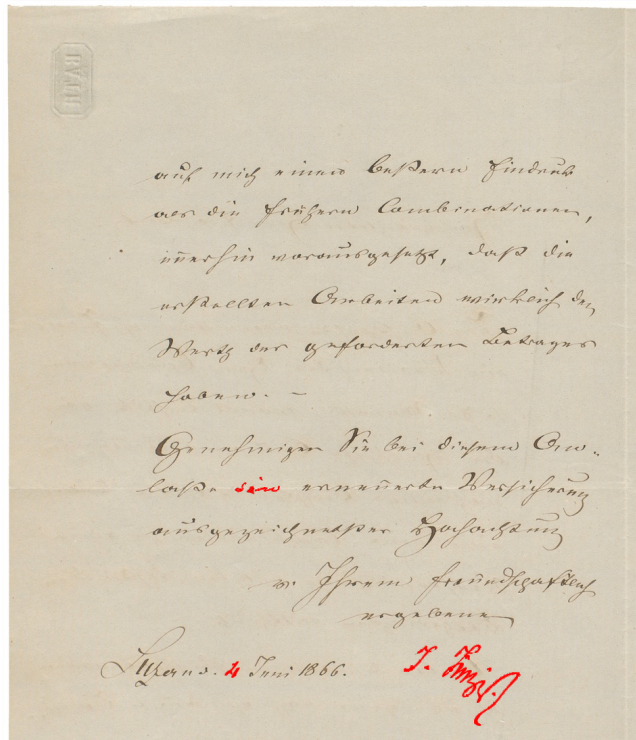
à faire connaître son
ministère est favorable.
à tout ce qui s'est passé
maintes et maintes fois.
déclaration prompte et
concernant l'attitude que
réorganisation de l'entre-
la meilleure impression.



.77%

German Handwriting Recognition

5.8% CER 15.6% WER

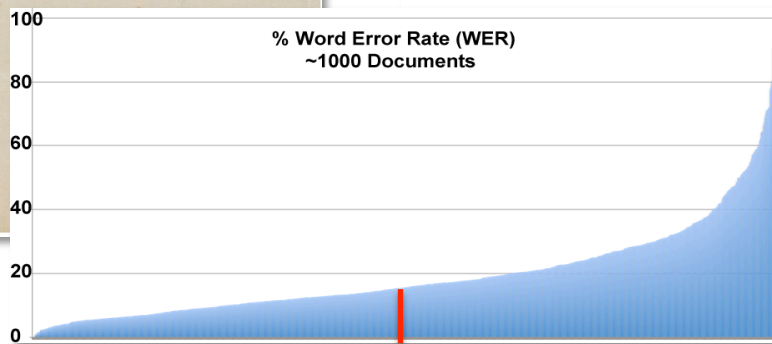


Automated Transcription

auf mich einen beßern Eindruck
als die frühern 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

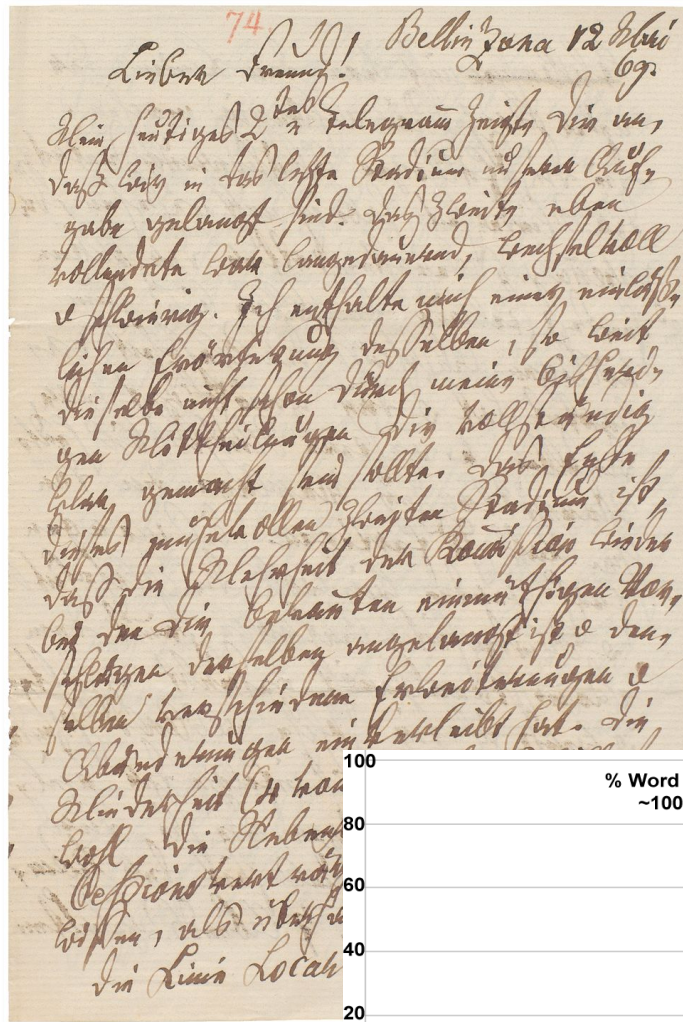
Luzern d. 20 Juni 1866. J. Zingg.



15.6%

German Handwriting Recognition

49.25% CER 71.93% WER



Automated Transcription

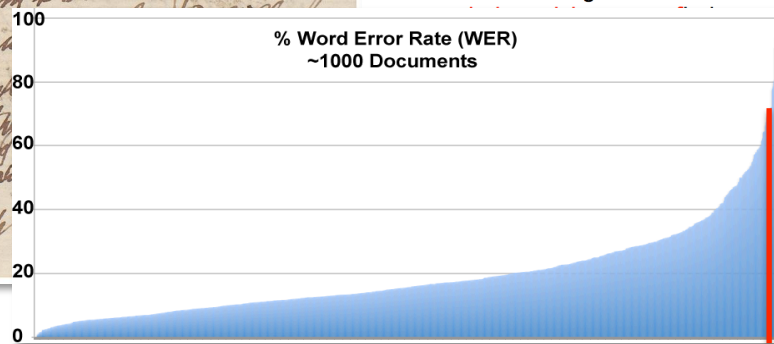
G. Bellinzona 12 Mai 69.

Lieber Freund!

A. Stelle mit Rhein heutiges Telegramm zeigte Dir an bei diesem Anlass unser Stichdaß wir in das letzte Stadium unserer Aufgabe gelangt sind. Das zweite eben vollendete war langedauemd, wechselvoll & schwierig. Ich enthalte mich einer einläßlichen Erörterungen in Bern willf. nicht schon durch meine Aufe. der unterblieben bin vollständig Mittheilungen zeigen zu besten gemacht sein sollte. Dann für klar ist den zweiten Stadium, ist daß unser allen zweiten Teil in wieder selbst der Sonnelles der & die Mehrheit der Sie sich von den die Erwarten einmüthigen Vorschlägen denselben angelangt ist & den bleiben. Erweiterungen & selben verschiedene einverleibt hat. Die Abänderungen ein kann man die

so befinden in den

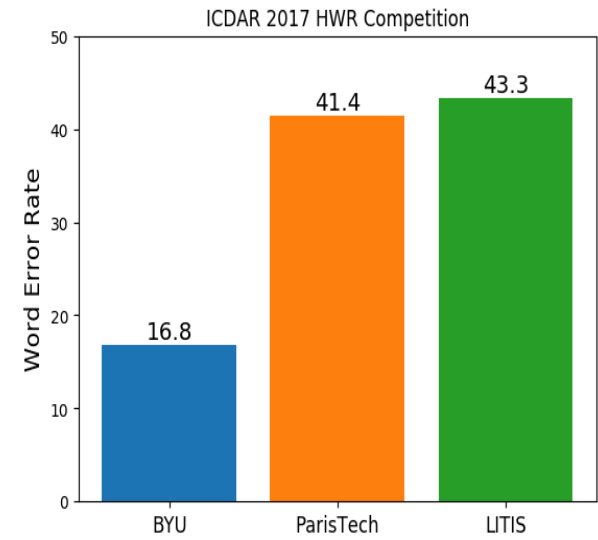
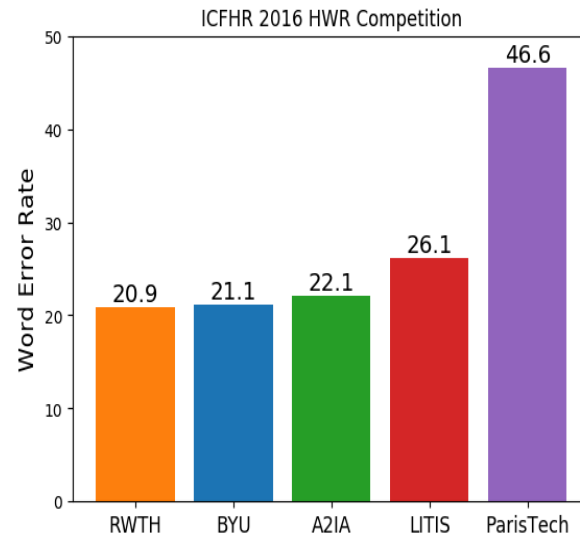
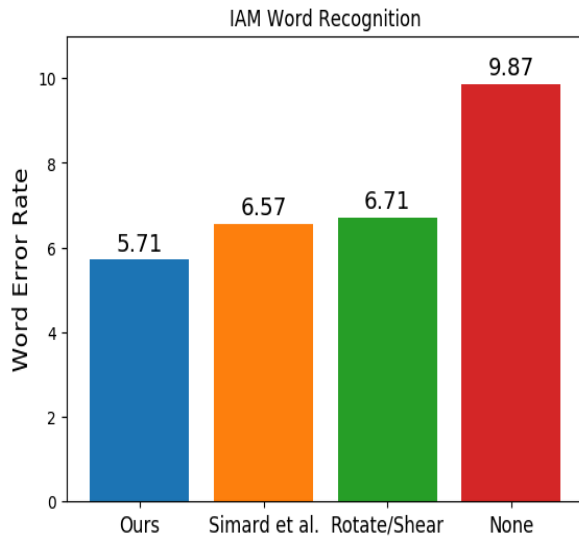
ahnen insbesondere



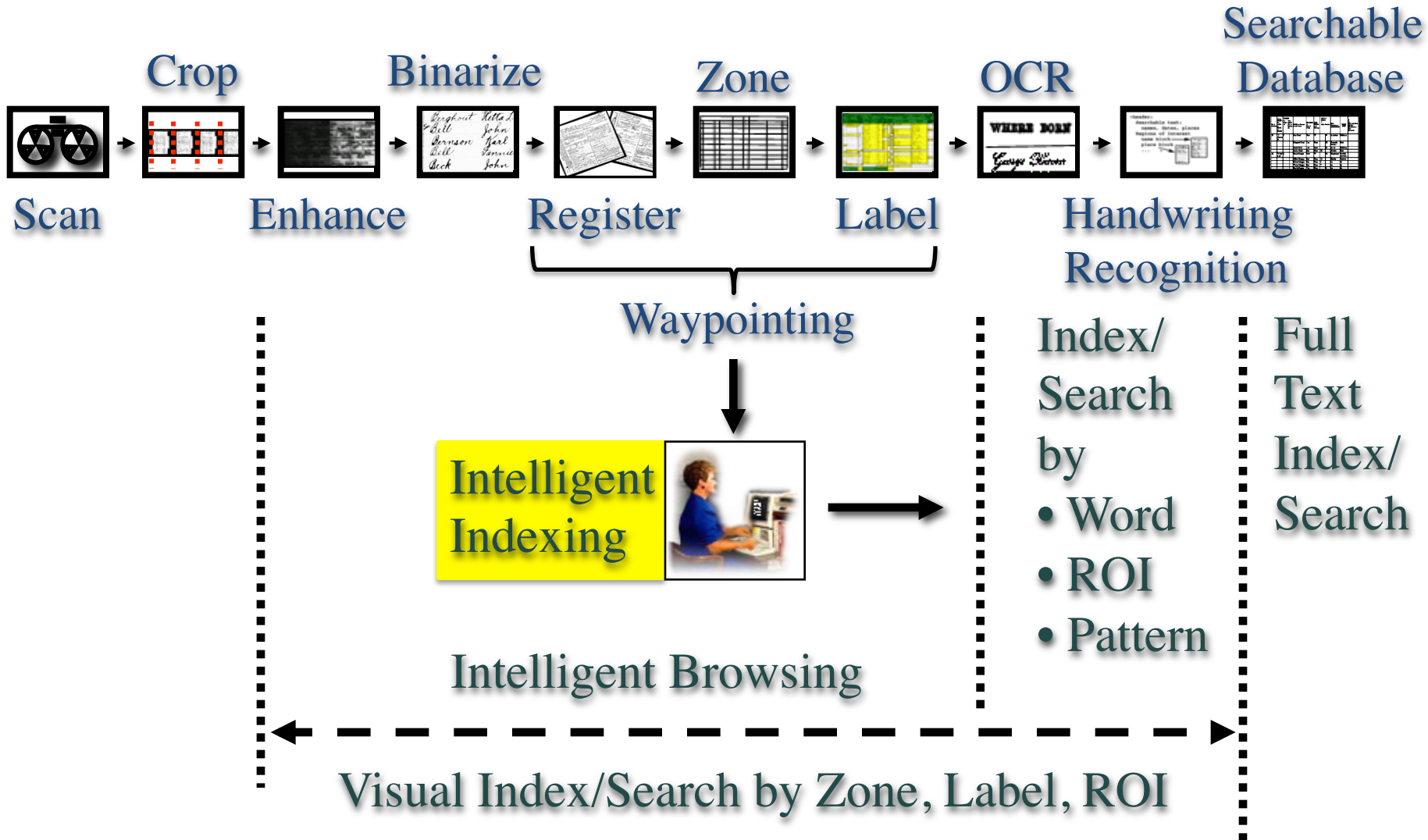
71.93%

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

Press <Enter> to label:

Click on any that don't match to remove.

Wife	Wife	Son
Son	Son	Son
	Son	Son
	Son	Son
	Son	Son

Handwritten text on the right side of the interface includes: Son, Son, Son, Daughters, Son, Head 1, Wife, Son, Daughters.

Intelligent Indexing

- Robert Clawson 2014



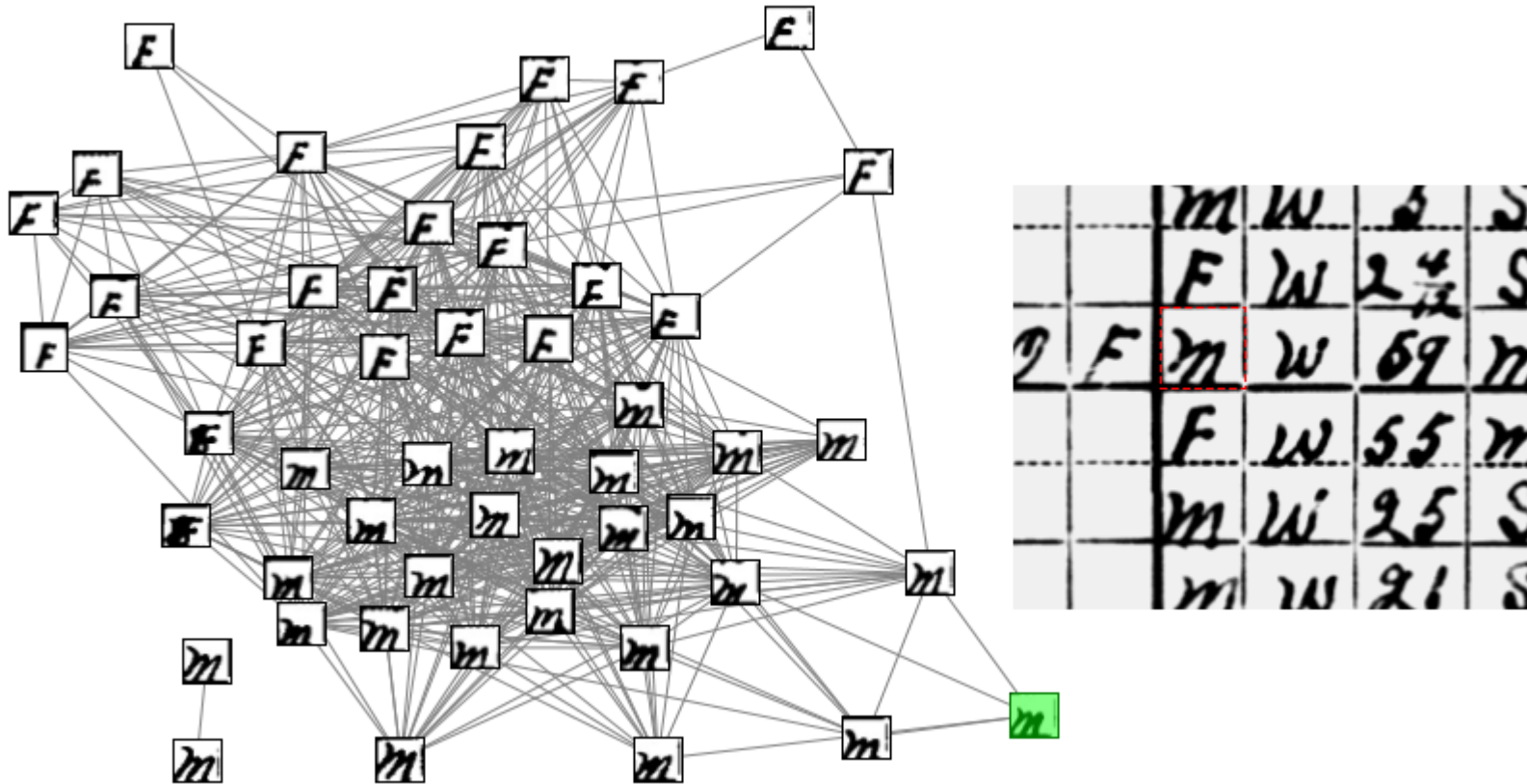
Intelligent Indexing Video

Intelligent Indexing



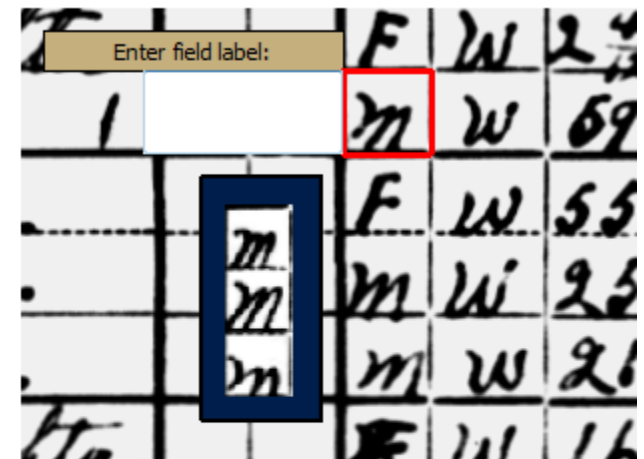
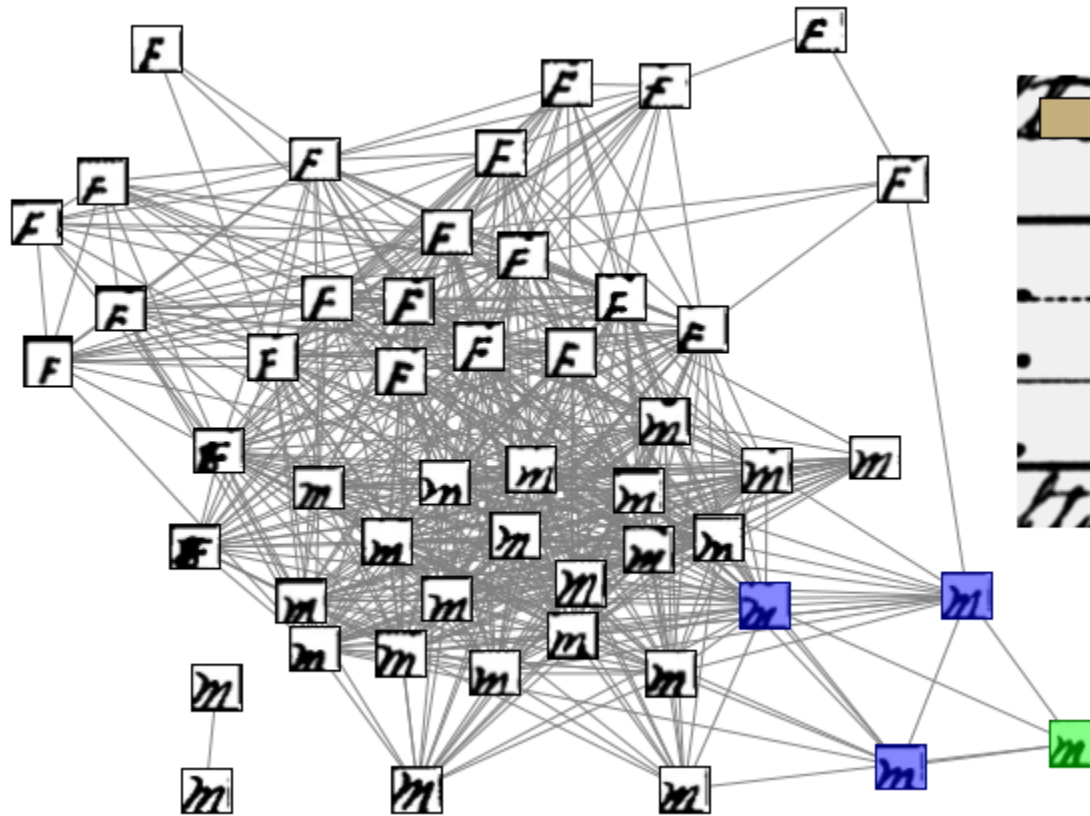
Enter field label:		Relationship of this person to the head of the family.	Sex	Color or race	Age at last in Aug.	Height	Weight	Year of issue when listed in States.	Materialism absent.	If naturalized foreigner.	Attended school from Sept. 1, 1907 to Sept. 1, 1908.	Whether ill read.	Whether ill writes.
I		Head	M	W	42	M		1891	na	1904		yes	yes
Joan M		wife	F	W	38	M						yes	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	1 1/2	S							
Noble Sarah		Head	F	W	66	Wd		1812	na	1864		yes	na
Peterson Joseph		Head	M	W	43	M						yes	yes
Blaise B		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	12	S					yes	yes	yes
Ira		daughter	F	W	9	S					yes		
			F	W	11 1/2	S							

Recommendation System



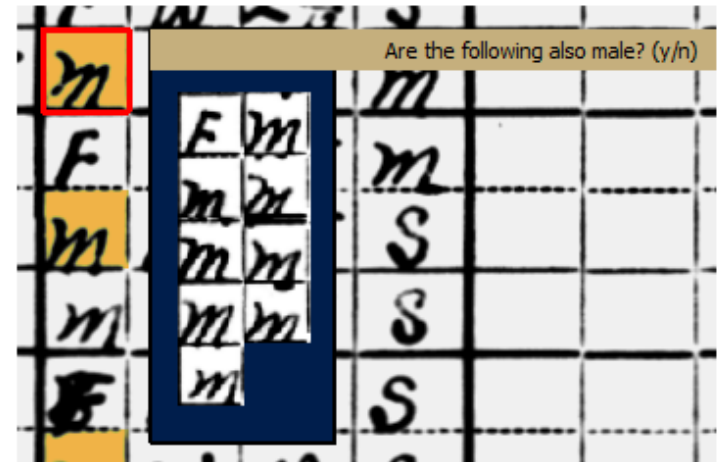
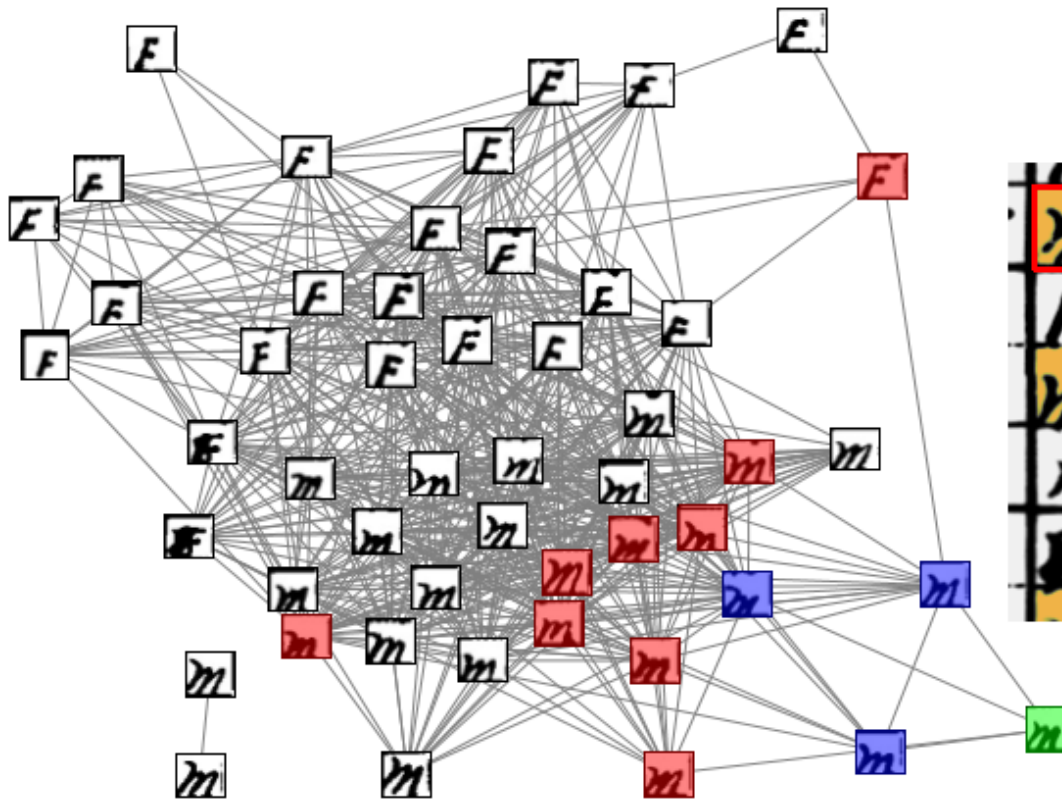
Leveraging the Label

Recommendation System



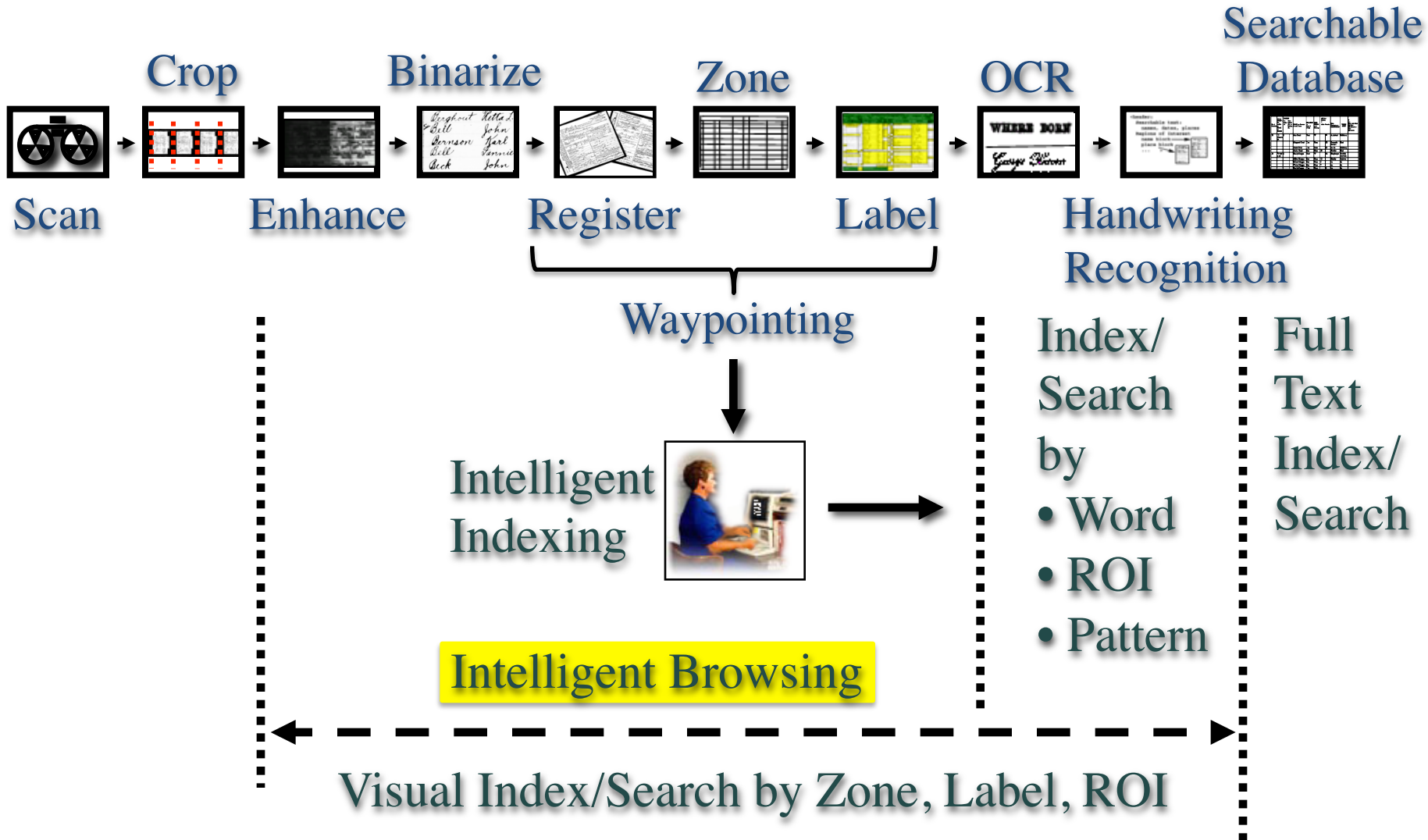
Leveraging the Label

Recommendation System



Leveraging the Label

The Digital Microfilm Pipeline



Search by Subwords/Patterns

- Davis 2018

<header :

Name fields

Place fields

Date fields

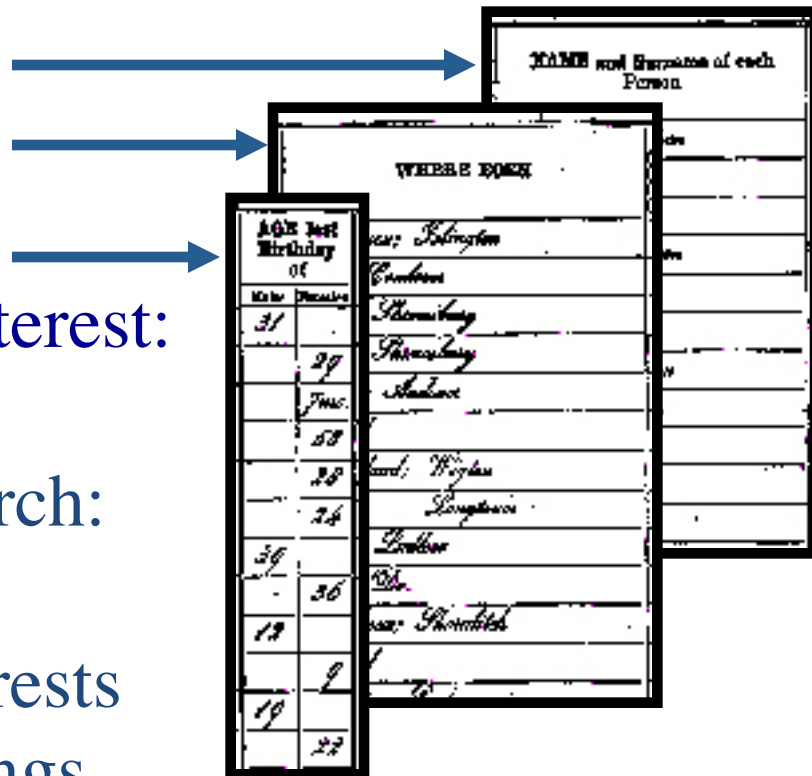
Regions of Interest:

Partial Word Search:

Subwords

Regions of Interests

Logical Groupings



N-Gram Word Spotting for

1-gram

n

Sum
new
ministrators
number
not
an
the m
been
gain
an

2-gram

pr

prices
price
to profit
proportionably
probable
prosecu-
provisions,

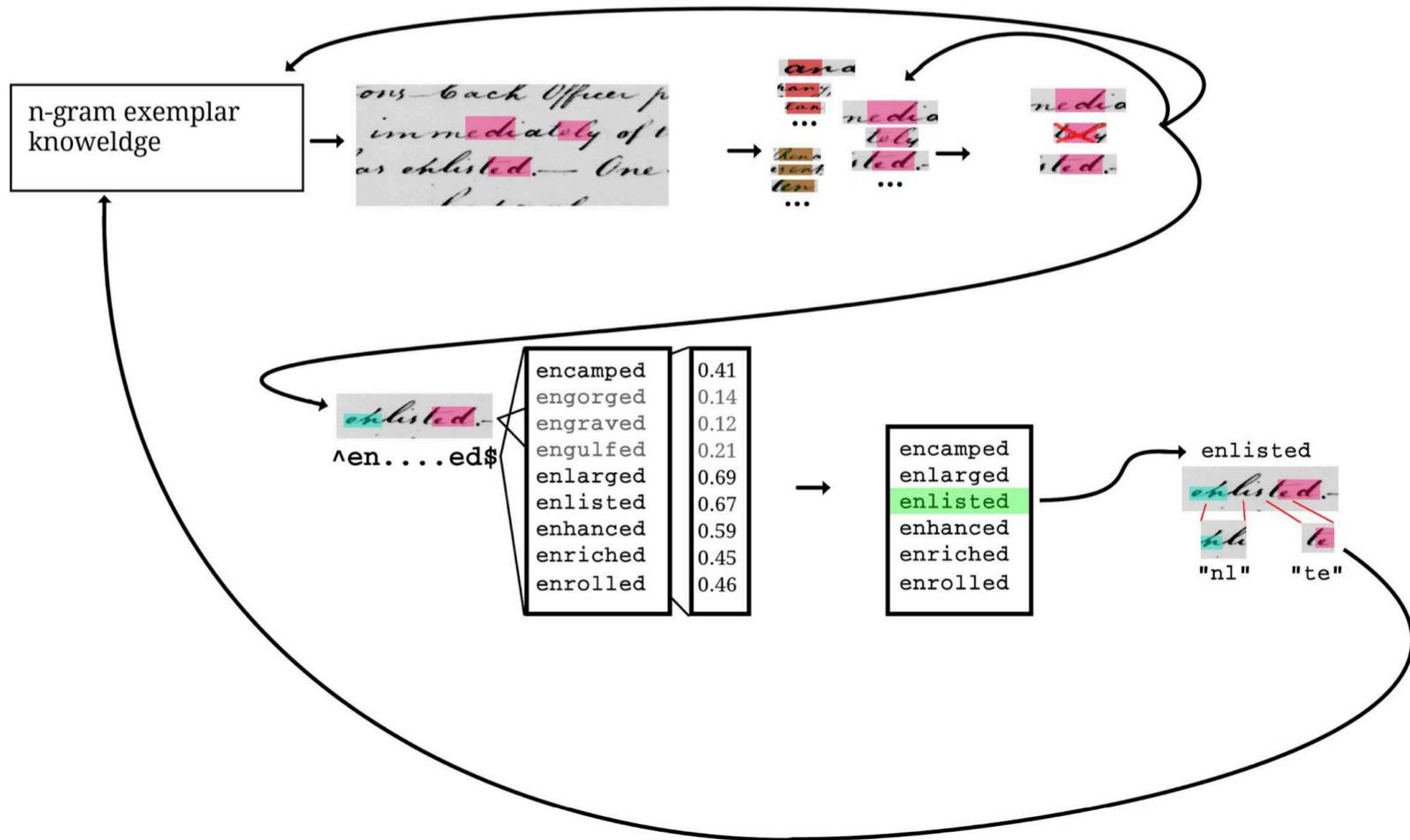
3-gram

sed

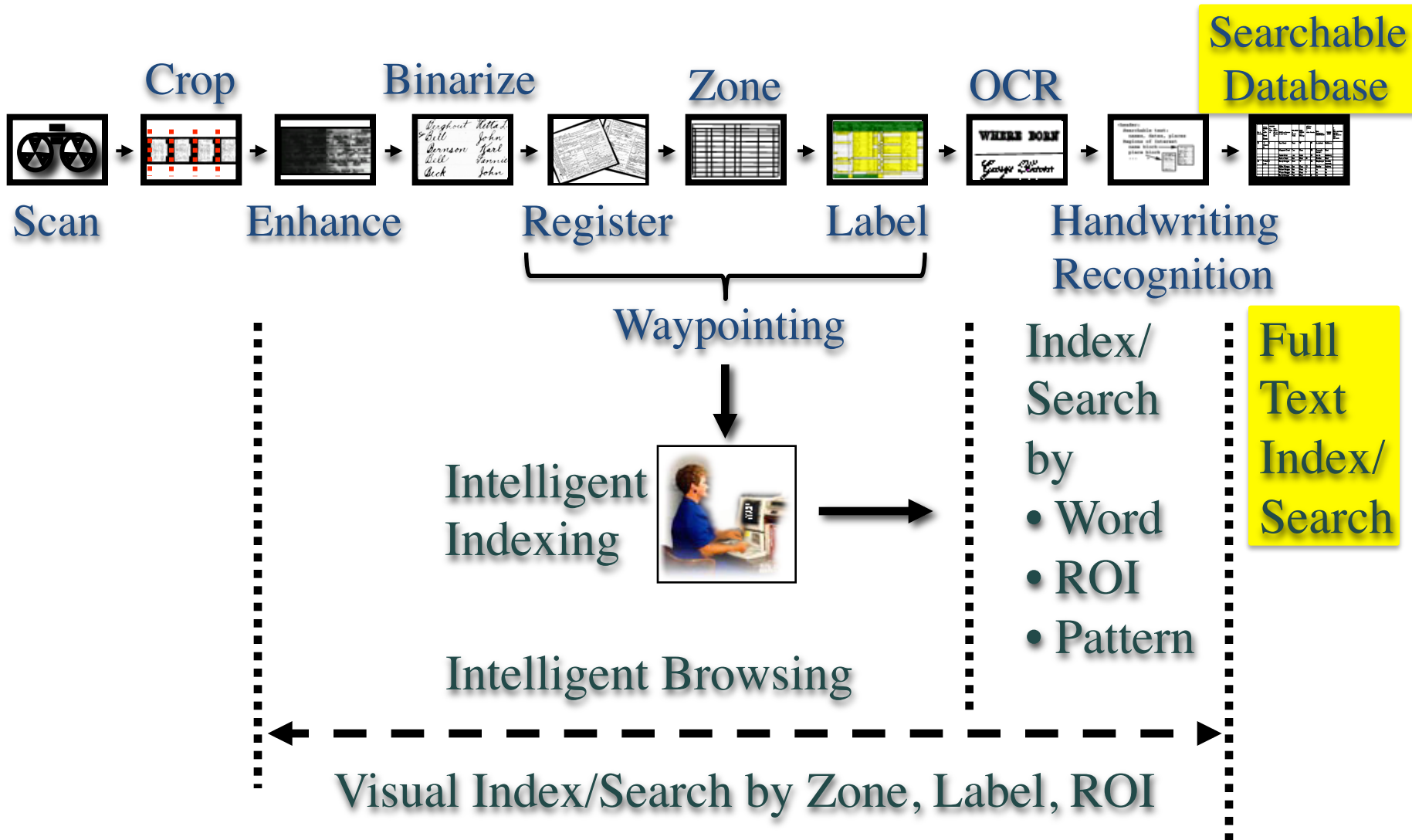
increased
unpractised
proposed
proposed
imposed
passed
disused
Ideals

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		NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of		Rank, Profession, or OCCUPATION	WHERE BORN	If (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic
		In-habit-ed	Unin-habited (U). or Building (B.)				Males	Females			
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

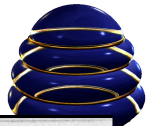
[Fan Chart](#)



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

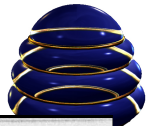
PLACE OF DEATH.
County of Huron
Township of Norwalk Registration District No. 626 File No. 11899
or
Village of Primary Registration District No. 4934 Registered No. 26
or
City of (No., St., Ward) [If death occurred in a hospital or institution, give its NAME instead of street and number.]

² FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS			MEDICAL CERTIFICATE OF DEATH	
³ SEX <u>male</u>	⁴ COLOR OR RACE <u>White</u>	⁵ SINGLE MARRIED WIDOWED OR DIVORCED <u>Married</u> <small>(Write the word)</small>	¹⁶ DATE OF DEATH <u>Feb 24</u> , 191 <u>9</u> <small>(Month) (Day) (Year)</small>	
⁶ DATE OF BIRTH <u>May 1st</u> , 18 <u>78</u> <small>(Month) (Day) (Year)</small>			¹⁷ I HEREBY CERTIFY, That I attended deceased from <u>Feb 20</u> , 191 <u>9</u> , to <u>24</u> , 191 <u>9</u> , that I last saw him alive on <u>Feb 23</u> , 191 <u>9</u> , and that death occurred, on the date stated above, at <u>1 A.M.</u>	
⁷ AGE <u>70</u> yrs. <u>9</u> mos. <u>22</u> ds. <small>If LESS than 1 day, hrs. or min.?</small>			The CAUSE OF DEATH* was as follows: <u>Right Lobar Pneumonia</u>	
⁸ OCCUPATION (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>92</u>			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	
⁹ BIRTHPLACE (State or country) <u>Ohio</u>			Contributory (SECONDARY) <u>Valvular heart lesion</u> <small>(Duration) yrs. mos. ds.</small>	
PARENTS	¹⁰ NAME OF FATHER <u>Patrick Martin</u>		(Signed) <u>J. J. Maynard</u> , M.D. <u>Feb 24</u> , 191 <u>9</u> (Address) <u>Norwalk Ohio</u>	
	¹¹ BIRTHPLACE OF FATHER (State or country) <u>Ireland</u>		*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.	
	¹² MAIDEN NAME OF MOTHER <u>Frances Nestor</u>		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	
¹³ BIRTHPLACE OF MOTHER (State or country) <u>Ireland</u>		¹⁴ THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <u>Sarah Martin</u> (Address) <u>Norwalk O</u>		
¹⁵ Filed <u>2/24</u> , 191 <u>9</u> <u>St. Mary's</u> Registrar			¹⁹ PLACE OF BURIAL OR REMOVAL <u>St. Mary's</u> ²⁰ UNDERTAKER <u>Carroll Brady</u>	
			¹⁹ DATE OF BURIAL <u>2-25</u> , 191 <u>9</u> ²⁰ ADDRESS <u>Norwalk O</u>	

11-3184

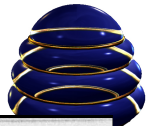
The Problem



Many documents
are heterogeneous
consisting of

Form V. S. No. 11-50M-8-20-17		STATE OF OHIO BUREAU OF VITAL STATISTICS CERTIFICATE OF DEATH	
PLACE 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>
City of _____ (No. _____, St., _____ Ward)		[If death occurred in a hospital or institution, give its NAME instead of street and number.]	
2 FULL NAME <i>Fredrick Martin</i>			
PERSONAL AND STATISTICAL PARTICULARS			MEDICAL CERTIFICATE OF DEATH
3 SEX <i>male</i>	4 COLOR OR RACE <i>White</i>	5 SINGLE MARRIED WIDOWED OR DIVORCED <i>Married</i>	16 DATE OF DEATH <i>Feb 24</i> , 191 <i>9</i>
6 DATE OF BIRTH <i>May 1st</i> , 1878			17 I HEREBY CERTIFY, That I attended deceased from <i>Feb 20</i> , 191 <i>9</i> , to <i>24</i> , 191 <i>9</i> , that I last saw him alive on <i>Feb 23</i> , 191 <i>9</i> , and that death occurred, on the date stated above, at <i>1 A.M.</i>
7 AGE <i>70</i> yrs. <i>9</i> mos. <i>22</i> ds.			The CAUSE OF DEATH* was as follows: <i>Right Lobar Pneumonia</i>
8 OCCUPATION (a) Trade, profession, or particular kind of work <i>Retired Laborer</i>			(Duration) _____ yrs. _____ mos. <i>4</i> ds.
(b) General nature of industry, business, or establishment in which employed (or employer) _____			Contributory <i>Valvular heart lesion</i>
9 BIRTHPLACE (State or country) <i>Ohio</i>			(Duration) _____ yrs. _____ mos. _____ ds.
PARENTS	10 NAME OF FATHER <i>Patrick Martin</i>	(Signed) <i>J. J. Maynard</i> , M.D.	
	11 BIRTHPLACE OF FATHER (State or country) <i>Ireland</i>	<i>Feb 24</i> , 191 <i>9</i> (Address) <i>Norwalk Ohio</i>	
	12 MAIDEN NAME OF MOTHER <i>Frances Nestor</i>	*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.	
	13 BIRTHPLACE OF MOTHER (State or country) <i>Ireland</i>	18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)	
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE			
(Informant) <i>Sarah Martin</i>		Where was disease contracted, If not at place of death? _____ Former or usual residence _____	
(Address) <i>Norwalk O</i>		19 PLACE OF BURIAL OR REMOVAL <i>St. Mary's</i> DATE OF BURIAL <i>2-25</i> , 191 <i>9</i>	
15 Filed <i>2/24</i> , 191 <i>9</i> <i>St. Martin</i> Registrar		20 UNDERTAKER <i>Carroll Brady</i> ADDRESS <i>Norwalk O</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

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.]

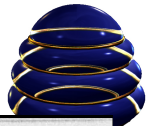
2 FULL NAME Frederick Martin

PERSONAL AND STATISTICAL PARTICULARS			MEDICAL CERTIFICATE OF DEATH	
3 SEX <u>male</u>	4 COLOR OR RACE <u>White</u>	5 SINGLE MARRIED WIDOWED OR DIVORCED <u>Married</u>	16 DATE OF DEATH <u>Feb 24</u> , 191 <u>9</u> (Month) (Day) (Year)	
6 DATE OF BIRTH <u>May 1st</u> , 18 <u>78</u> (Month) (Day) (Year)			17 I HEREBY CERTIFY, That I attended deceased from <u>Feb 20</u> , 191 <u>9</u> , to <u>24</u> , 191 <u>9</u> , that I last saw him alive on <u>Feb 23</u> , 191 <u>9</u> , and that death occurred, on the date stated above, at <u>1 A.M.</u>	
7 AGE <u>70</u> yrs. <u>9</u> mos. <u>22</u> ds.			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	
8 OCCUPATION (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>92</u>			The CAUSE OF DEATH* was as follows: <u>Right Lobar Pneumonia</u>	
9 BIRTHPLACE (State or country) <u>Ohio</u>			Contributory (SECONDARY) <u>Valvular heart lesion</u> (Duration) yrs. mos. <u>4</u> ds.	
PARENTS	10 NAME OF FATHER <u>Patrick Martin</u>	11 BIRTHPLACE OF FATHER (State or country) <u>Ireland</u>	(Signed) <u>J. J. Maynard</u> , M.D. <u>Feb 24</u> , 191 <u>9</u> (Address) <u>Norwalk Ohio</u>	
	12 MAIDEN NAME OF MOTHER <u>Frances Nestor</u>	13 BIRTHPLACE OF MOTHER (State or country) <u>Ireland</u>	*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.	
	14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <u>Sarah Martin</u> (Address) <u>Norwalk O</u>			19 PLACE OF BURIAL OR REMOVAL <u>St. Mary's</u> DATE OF BURIAL <u>2-25</u> , 191 <u>9</u>
15 Filed <u>2/24</u> , 191 <u>9</u> <u>St. Martin</u> Registrar			20 UNDERTAKER <u>Carroll Brady</u> <u>Norwalk O</u> ADDRESS	

11-3184

1919 Ohio Death Record

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

PLACE OF DEATH
County of Huron
Township of Norway
Registration District No. 626 File No. 11899
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 Friedrich Martini

PERSONAL AND STATISTICAL PARTICULARS

3 SEX male 4 COLOR OR RACE white 5 SINGLE MARRIED WIDOWED OR DIVORCED (Write the word) Married
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

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 Feb 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 (SECONDARY) Valvular heart lesion (Duration) yrs. mos. 4 ds.
(Signed) J. J. Maynard, M.D. Feb 24 1919 (Address) Norway, Ohio
*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.

PARENTS

10 NAME OF FATHER Patrick Martini
11 BIRTHPLACE OF FATHER (State or country) Ireland
12 MAIDEN NAME OF MOTHER Frances Nestor
13 BIRTHPLACE OF MOTHER (State or country) Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) Sarah Martini (Address) Norway, O.
15 Filed Feb 24 1919 G. H. Martini Registrar

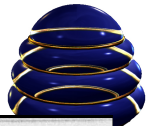
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents) At place of death yrs. mos. ds. In the State yrs. mos. ds. Where was disease contracted, If not at place of death? Former or usual residence

19 PLACE OF BURIAL OR REMOVAL St. Mary's 2-25 1919 DATE OF BURIAL
20 UNDERTAKER Charles Brady Norway, O. ADDRESS

11-3184

1919 Ohio Death Record

The Problem



Many documents are heterogeneous consisting of

- handwriting
- machine print
- solid lines

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

Village of _____ Primary Registration District No. 4934 Registered No. 26

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 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

PARENTS

10 NAME OF FATHER Patrick Martin

11 BIRTHPLACE OF FATHER (State or country) Ireland

12 MAIDEN NAME OF MOTHER Frances Nestor

13 BIRTHPLACE OF MOTHER (State or country) Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, O

15 Filed 2/24, 1919 St. Martin Registrar

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

(Duration) _____ yrs. _____ mos. 4 ds.

Contributory Valvular heart lesion
(SECONDARY)

(Duration) _____ yrs. _____ mos. _____ ds.

(Signed) J. J. Maynard, M.D.
Feb 24, 1919 (Address) Norwalk, Ohio

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

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 _____

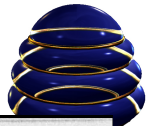
19 **PLACE OF BURIAL OR REMOVAL** St. Mary's DATE OF BURIAL 2-25, 1919

20 UNDERTAKER Carroll Brady ADDRESS Norwalk, O

11-3184

1919 Ohio Death Record

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

PLACE OF DEATH
County of Huron

Township of Norwalk Registration District No. 626 File No. 11899

Village of _____ Primary Registration District No. 4934 Registered No. 26

City of _____ (No. _____) St., _____ Ward _____

FULL NAME Fredrick Martin

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

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

PARENTS

10 NAME OF FATHER Patrick Martin

11 BIRTHPLACE OF FATHER (State or country) Ireland

12 MAIDEN NAME OF MOTHER Frances Nestor

13 BIRTHPLACE OF MOTHER (State or country) Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, O

15 Filed 2/24, 1919 Stok Martin Registrar

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

(Duration) _____ yrs. _____ mos. 4 ds.

Contributory (SECONDARY) Valvular heart disease
(Duration) _____ yrs. _____ mos. _____ ds.

(Signed) J. J. Maynard, M.D.
Feb 24, 1919 (Address) Norwalk, Ohio

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

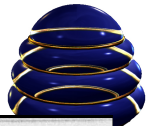
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
At place of death _____ yrs. _____ mos. _____ ds. In the State _____ yrs. _____ mos. _____ ds.
Where was disease contracted, If not at place of death? _____
Former or usual residence _____

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

20 UNDERTAKER Carroll Brady ADDRESS Norwalk, O

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: Huron

County of Huron Registration District No. 626 File No. 11899

Township of Norwalk or Village of _____ Primary Registration District No. 4934 Registered No. 26

City of _____ (No. _____) St., _____ Ward _____

FULL NAME: Fredrick Martin

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

PERSONAL AND STATISTICAL PARTICULARS

3 SEX: male 4 COLOR OR RACE: white 5 SINGLE MARRIED WIDOWED OR DIVORCED: married

6 DATE OF BIRTH: May 1st 1878

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

8 OCCUPATION: Retired laborer

9 BIRTHPLACE: Ohio

PARENTS

10 NAME OF FATHER: Patrick Martin

11 BIRTHPLACE OF FATHER: Ireland

12 MAIDEN NAME OF MOTHER: Frances Nestor

13 BIRTHPLACE OF MOTHER: Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE

(Informant) Sarah Martin (Address) Norwalk, O

15 Filed 2/24, 1919 Stok Martin Registrar

MEDICAL CERTIFICATE OF DEATH

16 DATE OF DEATH: Feb 24, 1919

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

Contributory Valvular heart disease

(Signed) J. Maynard, M.D. Feb 24, 1919 (Address) Norwalk, Ohio

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

18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)

At place of death _____ yrs. _____ mos. _____ ds. In the State _____ yrs. _____ mos. _____ ds.

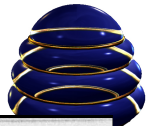
Where was disease contracted, If not at place of death? _____ Former or usual residence _____

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

20 UNDERTAKER: Carroll Brady ADDRESS: Norwalk, O

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

STATE OF OHIO
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH

PLACE OF DEATH: Huron

County of Huron Registration District No. 626 File No. 11899

Township of Norwalk or Village of _____ Primary Registration District No. 4934 Registered No. 26

City of _____ (No. _____) St., _____ Ward _____

FULL NAME: Fredrick Martin

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

PERSONAL AND STATISTICAL PARTICULARS

3 SEX: male 4 COLOR OR RACE: white 5 SINGLE MARRIED WIDOWED OR DIVORCED: married

6 DATE OF BIRTH: May 1st 1878

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

8 OCCUPATION: Retired laborer

9 BIRTHPLACE: Ohio

PARENTS

10 NAME OF FATHER: Patrick Martin

11 BIRTHPLACE OF FATHER: Ireland

12 MAIDEN NAME OF MOTHER: Frances Nestor

13 BIRTHPLACE OF MOTHER: Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE

(Informant) Sarah Martin (Address) Norwalk, O

15 Filed 2/24 1919 Stok Martin Registrar

MEDICAL CERTIFICATE OF DEATH

16 DATE OF DEATH: Feb 24, 1919

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

(Duration) _____ yrs. _____ mos. 4 ds.

Contributory (SECONDARY) Valvular heart disease

(Duration) _____ yrs. _____ mos. _____ ds.

(Signed) J. J. Maynard, M.D. Feb 24, 1919 (Address) Norwalk, Ohio

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

18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)

At place of death _____ yrs. _____ mos. _____ ds. In the State _____ yrs. _____ mos. _____ ds.

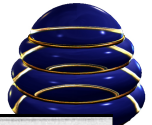
Where was disease contracted, If not at place of death? _____ Former or usual residence _____

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

20 UNDERTAKER: Carroll Brady ADDRESS: Norwalk, O

11-3184

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: Huron

County of Huron Registration District No. 626 File No. 11899

Township of Norwalk or Village of _____ Primary Registration District No. 4934 Registered No. 26

City of _____ (No. _____) St., _____ Ward _____

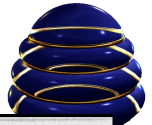
FULL NAME: Fredrick Martin

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

PERSONAL AND STATISTICAL PARTICULARS			MEDICAL CERTIFICATE OF DEATH	
3 SEX <u>male</u>	4 COLOR OR RACE <u>White</u>	5 SINGLE MARRIED WIDOWED OR DIVORCED <u>Married</u>	16 DATE OF DEATH <u>Feb 24</u> , 191 <u>9</u>	
6 DATE OF BIRTH <u>May 1st 1848</u>			17 I HEREBY CERTIFY, That I attended deceased from <u>Feb 20</u> , 191 <u>9</u> , to <u>24</u> , 191 <u>9</u> , that I last saw him alive on <u>Feb 23</u> , 191 <u>9</u> , and that death occurred, on the date stated above, at <u>1 A.M.</u>	
7 AGE <u>70</u> yrs. <u>9</u> mos. <u>22</u> ds.			The CAUSE OF DEATH* was as follows: <u>Right Lobar Pneumonia</u>	
8 OCCUPATION (a) Trade, profession, or particular kind of work <u>Retired Laborer</u>			Contributory (SECONDARY) <u>Valvular heart lesion</u>	
9 BIRTHPLACE (State or country) <u>Ohio</u>			(Duration) _____ yrs. _____ mos. <u>4</u> ds.	
PARENTS	10 NAME OF FATHER <u>Patrick Martin</u>	11 BIRTHPLACE OF FATHER (State or country) <u>Ireland</u>	(Signed) <u>9 J. Maynard</u> , M.D. <u>Feb 24 1919</u> (Address) <u>Norwalk Ohio</u>	
	12 MAIDEN NAME OF MOTHER <u>Frances Weston</u>	13 BIRTHPLACE OF MOTHER (State or country) <u>Ireland</u>	*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.	
	14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <u>Sarah Martin</u> (Address) <u>Norwalk O</u>			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 _____
15 Filed <u>2/24</u> , 191 <u>9</u> Registrar <u>Stok Martin</u>			19 PLACE OF BURIAL OR REMOVAL <u>St. Mary's</u>	
			20 UNDERTAKER <u>Carroll Brady</u>	
			DATE OF BURIAL <u>2-25</u> , 191 <u>9</u> ADDRESS <u>Norwalk O</u>	

11-3184

The Problem



The interleave and overlap, greatly compounds the difficulty of OCR and handwriting 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
Village of _____ Primary Registration District No. 4934 Registered No. 26
City of _____ (No. _____) St., _____ Ward _____
FULL NAME Fredrick 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 Nestor

13 BIRTHPLACE OF MOTHER (State or country) Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, O

15 Filed 2/24, 1919 Stok Martin Registrar

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

Contributory (SECONDARY) Valvular heart disease
(Duration) _____ yrs. _____ mos. 4 ds.

(Signed) J. J. Maynard, M.D. Feb 24, 1919 (Address) Norwalk Ohio

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

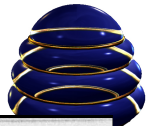
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
At place of death _____ yrs. _____ mos. _____ ds. In the State _____ yrs. _____ mos. _____ ds.
Where was disease contracted, If not at place of death? _____
Former or usual residence _____

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

20 UNDERTAKER Charles Brady ADDRESS Norwalk, O

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

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 Fredrick 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

PARENTS

10 NAME OF FATHER Patrick Martin
11 BIRTHPLACE OF FATHER (State or country) Ireland
12 MAIDEN NAME OF MOTHER Frances Nestor
13 BIRTHPLACE OF MOTHER (State or country) Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, O

15 Filed 2/24 1919 Stok Martin Registrar

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

Contributory Valvular heart disease
(Duration) _____ yrs. _____ mos. 4 ds.
(Signed) J. J. Maynard, M.D. (Duration) _____ yrs. _____ mos. _____ ds.
Feb 24 1919 (Address) Norwalk Ohio

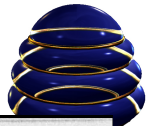
*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.

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 DATE OF BURIAL 2-25 1919
20 UNDERTAKER Charles Brady ADDRESS Norwalk, O

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

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: Fredrick Martin

PERSONAL AND STATISTICAL PARTICULARS			MEDICAL CERTIFICATE OF DEATH	
3 SEX <u>male</u>	4 COLOR OR RACE <u>White</u>	5 SINGLE MARRIED WIDOWED OR DIVORCED <u>Married</u> (Write the word)	16 DATE OF DEATH <u>Feb 24</u> , 191 <u>9</u> (Month) (Day) (Year)	
6 DATE OF BIRTH <u>May 1st 1878</u> (Month) (Day) (Year)			17 I HEREBY CERTIFY, That I attended deceased from <u>Feb 20</u> , 191 <u>9</u> , to <u>24</u> , 191 <u>9</u> , that I last saw him alive on <u>Feb 23</u> , 191 <u>9</u> , and that death occurred, on the date stated above, at <u>1 A.M.</u>	
7 AGE <u>70</u> yrs. <u>9</u> mos. <u>22</u> ds. If LESS than 1 day, _____ hrs. or _____ min.?			The CAUSE OF DEATH* was as follows: <u>Right Lobar Pneumonia</u>	
8 OCCUPATION (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>92</u>			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 _____	
9 BIRTHPLACE (State or country) <u>Ohio</u>			Contributory (SECONDARY) <u>Valmian heart disease</u> (Duration) _____ yrs. _____ mos. _____ ds.	
PARENTS	10 NAME OF FATHER <u>Patrick Martin</u>	19 (Signed) <u>97 Mayord</u> , M.D. <u>Feb 24</u> , 191 <u>9</u> (Address) <u>Norwalk Ohio</u>		
	11 BIRTHPLACE OF FATHER (State or country) <u>Ireland</u>	*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.		
	12 MAIDEN NAME OF MOTHER <u>Frances Nestor</u>	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 _____		
13 BIRTHPLACE OF MOTHER (State or country) <u>Ireland</u>			10 PLACE OF BURIAL OR REMOVAL <u>St Mary's</u> DATE OF BURIAL <u>2-25</u> , 191 <u>9</u>	
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <u>Sarah Martin</u> (Address) <u>Norwalk O</u>			20 UNDERTAKER <u>Carroll Brady</u> ADDRESS <u>Norwalk O</u>	
15 Filed <u>2/24</u> , 191 <u>9</u> <u>Flores Martin</u> Registrar				

11-3184

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

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

PARENTS

11 BIRTHPLACE OF FATHER (State or country) Ireland

12 MAIDEN NAME OF MOTHER Frances Nestor

13 BIRTHPLACE OF MOTHER (State or country) Ireland

14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Sarah Martin
(Address) Norwalk, O

15 Filed 2/24, 1919 Stok Martin Registrar

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

(Duration) _____ yrs. _____ mos. 4 ds.

Contributory Valvular heart lesion
(SECONDARY)

(Duration) _____ yrs. _____ mos. _____ ds.

(Signed) J. J. Maynard, M.D.
Feb 24, 1919 (Address) Norwalk, Ohio

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

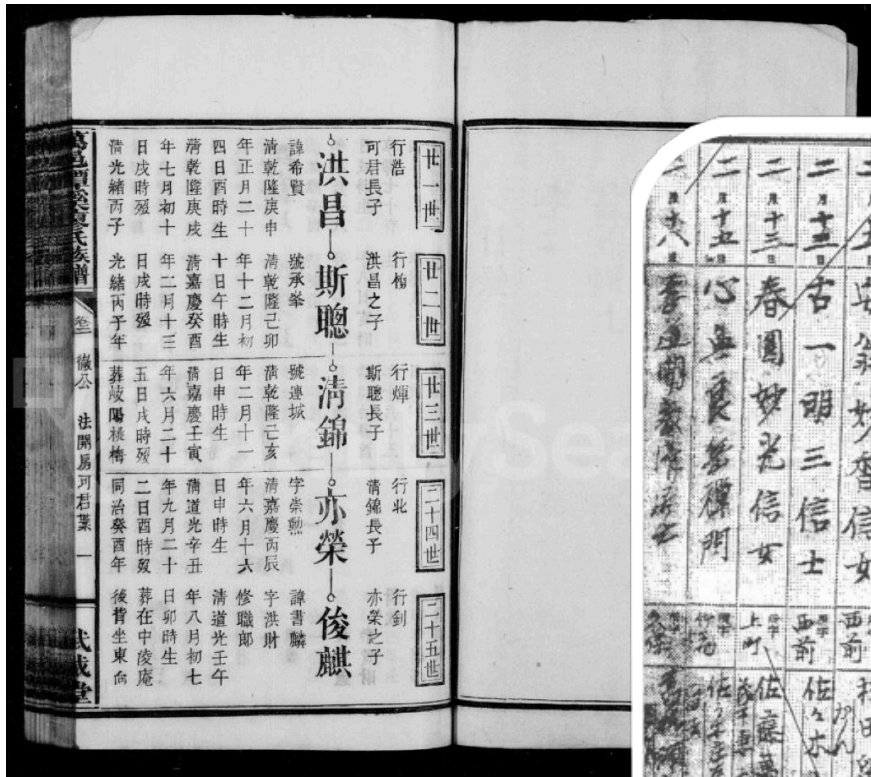
18 LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients, or Recent Residents)
At place of death _____ yrs. _____ mos. _____ ds. In the State _____ yrs. _____ mos. _____ ds.
Where was disease contracted, If not at place of death? _____
Former or usual residence _____

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

20 UNDERTAKER Carroll Brady ADDRESS Norwalk, O

11-3184

What about different alphabets?



Jia Pu



Koseki



ARTS

- Asian Records Transcription System

Applications Places

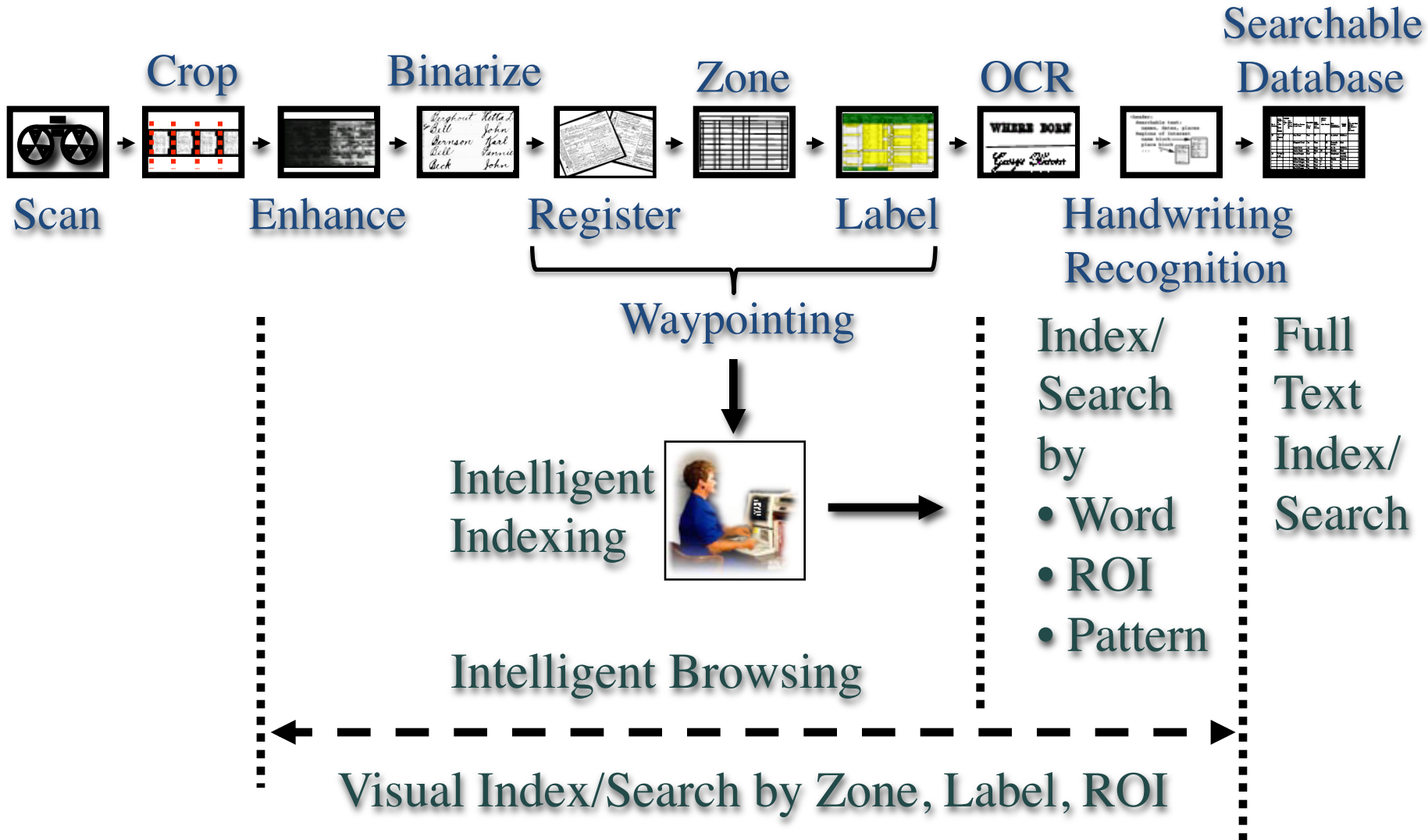
file:///home/ubuntu/Desktop/jiapu-recordbuilder/static/index.html?https://familysearch.org/ark:/61903/3:1:3Q57-L9SJ-V9FG-S.jpg

FamilySearch
Prev Next

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

[bin] rb-prototype - Goog... [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

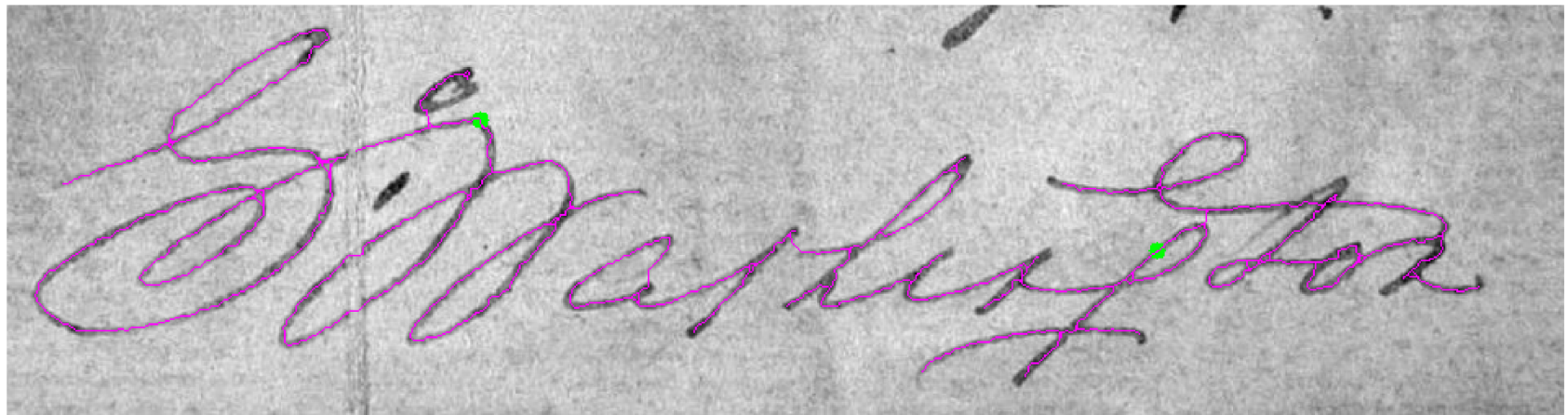
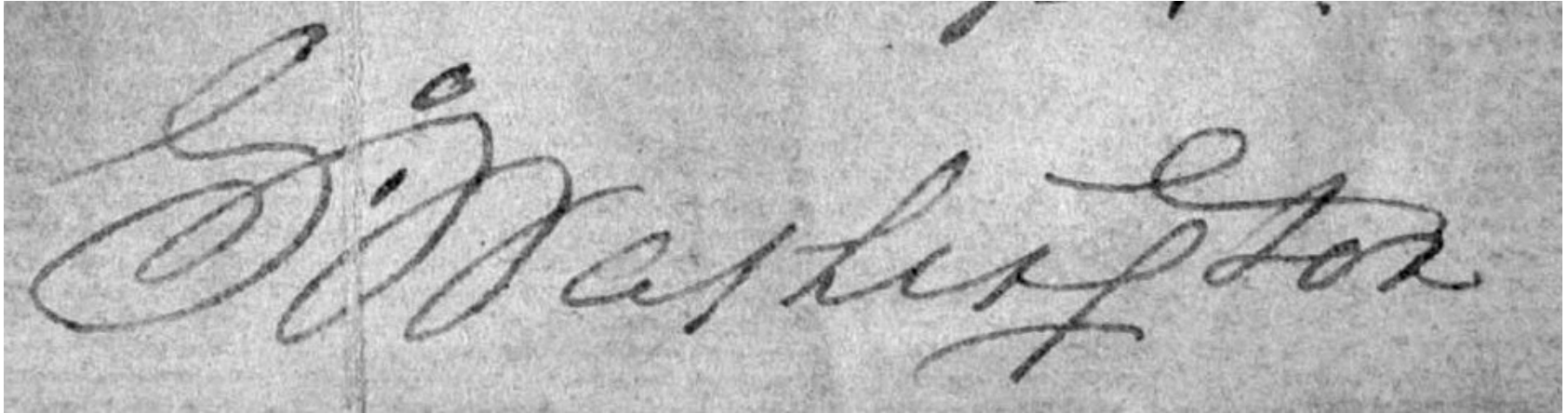
Zhi-han Tsai

Intelligent Pen



Kevin's Slides

Intelligent Pen



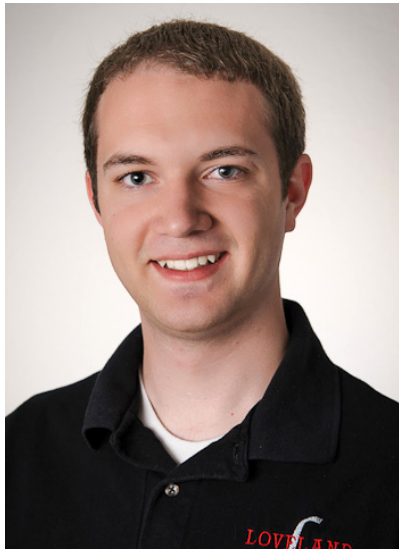
Intelligent Pen



Brooks, Martin L.

Brooks, Martin L.

The Students



Ryan Cheatham
Sophomore
Loveland, Colorado

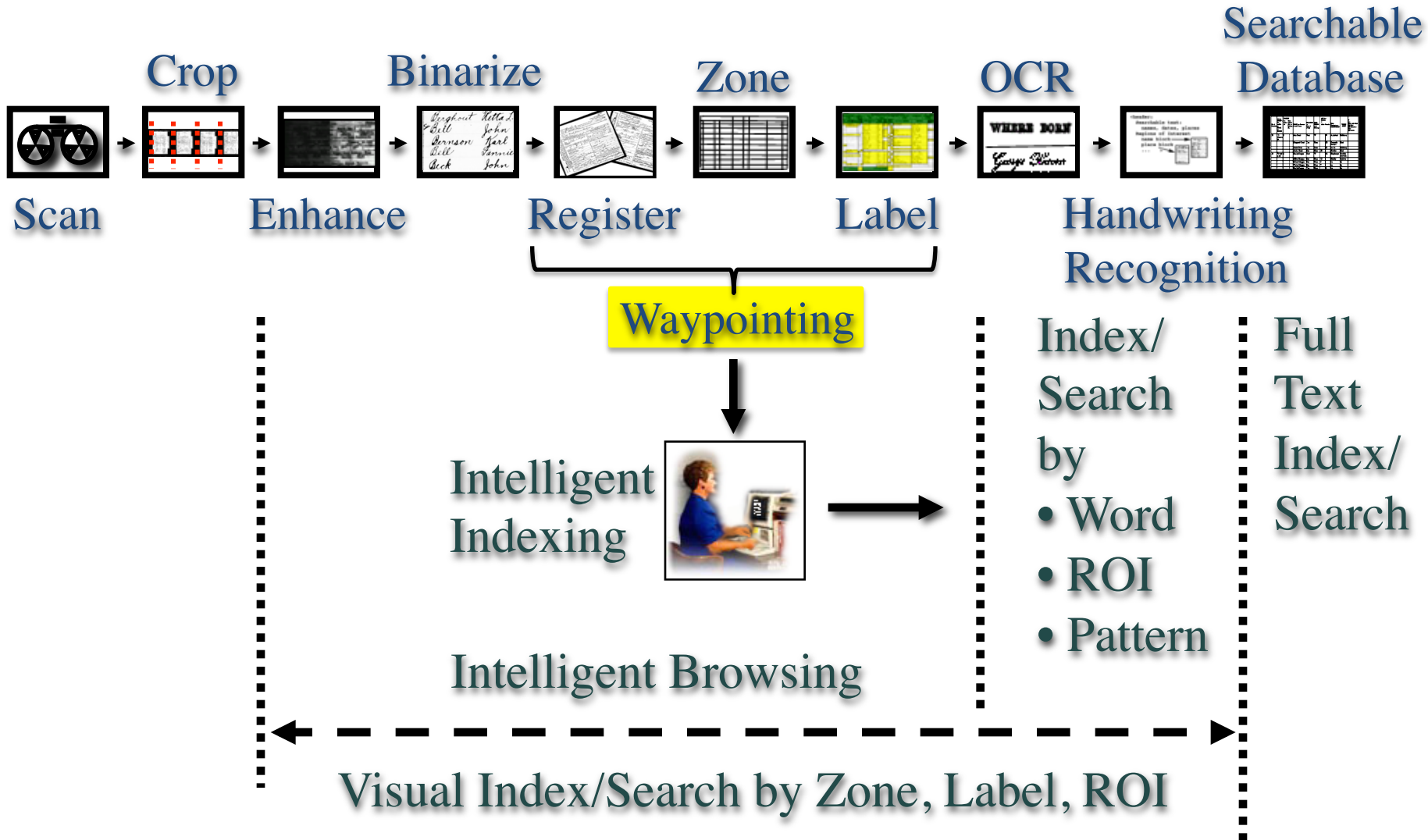


Robert Clawson
MS Student
San Diego, California



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

FILMADO POR LA SOCIEDAD GENEALOGICA		LOCALIDAD QUE CUBRE EL REGISTRO	
EN MEXICO		PARROQUIA DE SANTIAGO TEMAPACHE VER. DIOCESIS DE TUXPAM VER. MEX.	
FOTOGRAFO	REDUCCION (X)	TITULO DEL REGISTRO	
SANTIAGO MEJIA M	40	BAUTISMOS	
FECHA DE FILMACION	EXPOSICION	ITEM 1	
JUN 15 1984	20		
NUMERO DE EMULSION	NUMERO DE CAMARA		
7456-452-1	51896		
NUMERO DEL PROYECTO	NUMERO DEL ROLLO		
MEXC-2-059	1		

OCR/Index the text

* David Ouimette, Jake

Gehring


Automated Waypointing: Group titleboards uniformly

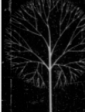


ENGLAND	BISHOP'S TRANSCRIPTS
COUNTY DERBYSHIRE	Baptisms
PARISH Hayfield	Marriages
	Burials
	YEARS 1662-1810
<small>FILMED AT LICHFIELD, STAFFORDSHIRE, ENGLAND, BY THE GENEALOGICAL SOCIETY, SALT LAKE CITY, UTAH, U.S.A.</small> <small>JANUARY 1966</small>	

Archivo de la Parroquia
- de -
SANTA MARIA de los RAYOS
PARISH
BAUTISMO DE HIJOS LEGITIMOS
VOLUMEN NUM. 104 1897 1899
<small>Academia Mexicana de Genealogía y Heráldica</small>
<small>Reductor</small> 111 <small>Fecha</small> SEP - 23 DE 1957

...

 <small>MICROFILMED BY THE GENEALOGICAL SOCIETY SALT LAKE CITY, UTAH AT</small> STATE DEPT. HEALTH HONOLULU HI 96801	<small>LOCALITY OF RECORD</small> HAWAII
<small>OPERATOR</small> STEVE BALDRIDGE	<small>REDUCCION X</small> 16 X
<small>DATE FILMED</small> 17 AUG 1978	<small>LIGHT METER SETTING</small> 64-2
<small>FILM EMULSION NUMBER</small> 7252 11	<small>FILM UNIT SER NO</small> GSI 5481
<small>PROJECT NUMBER</small> HA00009	<small>ROLL NUMBER</small> 17
<small>TITLE OF RECORD</small> HAWAII, MAUI, KAUAI, NIIHAU BIRTHS, DEATHS, MARRIAGE	<small>VOLUME/S</small> Item 3
<small>YEAR S INCLUDED</small> 1862 - 1896	

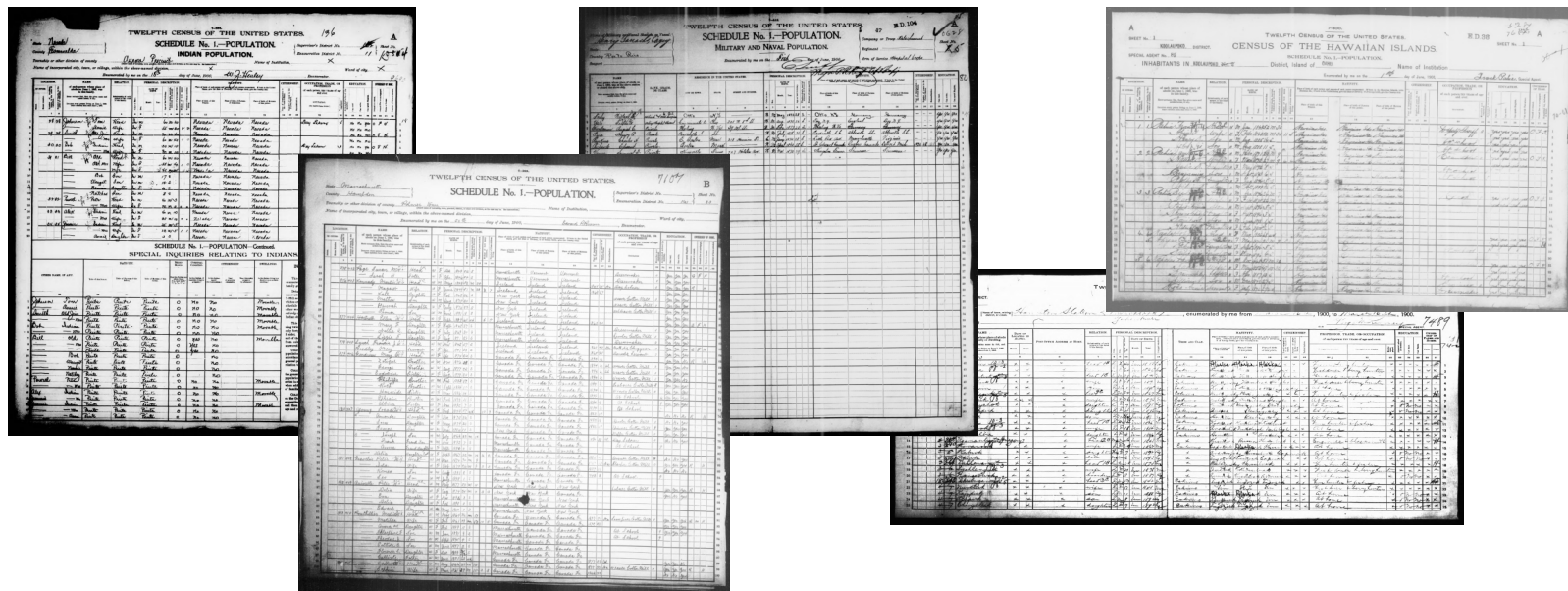
 <small>FILMADO POR LA SOCIEDAD GENEALOGICA</small> EN MEXICO	<small>LOCALIDAD QUE COBRE EL REGISTRO</small> PARROQUIA DE SANTIAGO TEMAPACHE VER. DIOCESIS DE TUXPAM VER. MEX.
<small>FOTOGRAFO</small> SANTIAGO MEJIA M	<small>REDUCCION (X)</small> 40
<small>FECHA DE FILMACION</small> JUN 15 1984	<small>EXPOSICION</small> 20
<small>NUMERO DE EMULSION</small> 7456-452-1	<small>NUMERO DE CAMARA</small> 51896
<small>NUMERO DEL PROYECTO</small> MEXC-2-059	<small>NUMERO DEL ROLLO</small> 1
<small>TITULO DEL REGISTRO</small> BAUTISMOS	ITEM 1

...

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

