City Directories Automated Indexing

Collin Zoeller | czoeller@byu.edu

Brigham Young University

Jonathan Palmer | jpalmer@hbs.edu

Harvard Business School



Data

- R.L. Polk city directories, business and alphabetical
- Demographic survey for automotive and insurance companies
- Rochester, NY (1879-2000)
- Available in most major cities

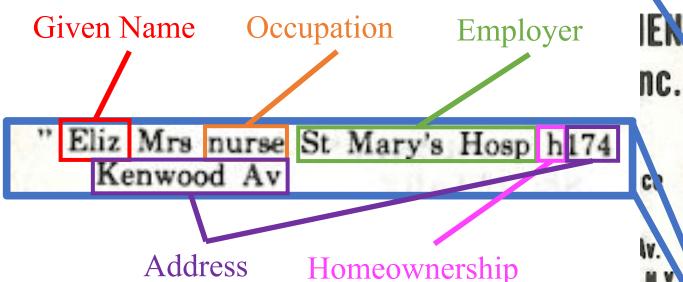
Alphabetical

1979 Rochester, NY

Family Name

176

Spouse



Also included in dataset but not shown:

- Race (select locations and early time periods only)
- Military Branch of involvement
- Resident location if not in Rochester

LAZERSON

- " Emanuel H & Rita; sec-treas Monroe Surgical Sup rBrighton NY
- " Judy Lynn studt r691 Parsells Av
- " Lee studt r691 Parsells Av
- " Nathan M retd r53 Tyler St
- " Patti B tchr Boces h17 Portsmouth Ter Apt 2

Lazier E retd h241 Alexander St Apt 213 Lazinski P h214 Berkeley St Apt 4b

Lazlo Pal h144 Weld St Lazore Dorothy h167 Spur Harbor Ct

- Eliz Mrs nurse St Mary's Hosp h174 Kenwood Av
- Joseph retd h240 Hazelwood Ter
- Patricia (Blue & White Grill) h44 Lyell
- " Robt M & Elizabeth; iron wkr h174 Kenwood Av

Lazur Jake & Ethel h425 Lake Av Lazus Laura rSt Johns Hm For Agng LEAHY

III YOUI POCKEL!

400 LAKE AVENUE - BETWE

Sharon Mrs Leak David r2

- " Delbert & V
- Wks h237
- " Delbert Jr r

" Wallace & A Honing h3

Leake Herbert

Whiteford " Margt L clk

Leakey Debora

Dorothy S re

Gordon Rev

Ch r648 C " Paul A & K

Prod h290 Leaper D A r7

Debbie mgr

" Patricia med

Business 1979 Rochester, NY

Business Category

Business Name

Municipality

Address

Also included in dataset but not shown:

- Professionals
- executive leadership
- Number of employees

ICE CREAM AND FROZEN DESSERTS— RETAIL

Abbott's Frozen Custard 2195 Empire Blvd	
(W)	
Abbott's Frozen Custard 3071 W Henrietta	0.51
Rd (H)	
Abbott's Frozen Custard 1170 Brooks Av	
(G)	
Abbotts Custard 1220 Fairport Rd (F)	
Banana Cabana 1475 E Ridge Rd (I)	22
Baskin-Robbins Greece Towne Mall	
Baskin & Robbins 2044 Chili Av (G)	
Baskin-Robbins Ice Cream 525 Titus Av	
(I)	33
Buckman's Ice Cream Village 2576 W	
Ridge Rd (GR)	
Carvel 689 E Ridge Rd (I)	
Carvel 2513 E Henrietta Rd (H)	

- Text segmented and read via OCR
- Text easily parsed through regular expressions
- Current process is ~13 seconds/page (compared to some 10-15 minutes/page to index manually)



	year	image_id	block_id	is_ad	raw_string	last name	person1	suffix	person2	military	occupation	employer	type	address	city_state
6	1979	6	7	TRUE	Insurance William Combs General Manager Edward Tackaberry Supervisor Suite 1100 Alliance Bldg 183 E Main St (14604) Tel 546-6100	Insurance	William					14604			
7	1979	6	8	FALSE	Aevarato Enrico studt r309 Murray St	Aevarato	Enrico				studt		r	309 Murray St	
8	1979	6	9	FALSE	Afen Anthony G & Marie N retd h129 Westview Ter	Afen	Anthony G		Marie N		retd h		h	129 Westview Ter	
9	1979	6	10	FALSE	Afenko Peter & Afansia retd h85 Branch St	Afenko	Peter		Afansia		retd		h	85 Branch St	
10	1979	6	11	FALSE	Afentakis Panayiotis & Sofia studt h302 University Pk	Afentakis	Panayiotis		Sofia		studt		h	302 University Pk	
12	1979	6	13	FALSE	Affel Herman A pres Computer Consoles rPhiladelphia PA	Affel	Herman A				pres	Computer Consoles	r		Philadelphia PA
14	1979	6	15	TRUE	Affinity Woodworks Inc Newell White Pres 15 Henrietta St	Affinity	Woodworks								
16	1979	6	17	FALSE	Affonso Arthur & Hazel h345 Alexander St Apt 10	Affonso	Arthur		Hazel				h	345 Alexander St Apt 10	

Advantages and Impact

- Fast unsupervised indexing
- A treasure trove of data not included in census records
 - Race (some locations)
 - Employer
 - Searchable address information
 - Tracks individuals and businesses across time

