

# End-to-End Robokeying of “Born Paper” Obituaries

Patrick Schone, Heath Nielson

(FamilySearch, [patrickjohn.schone@ldschurch.org](mailto:patrickjohn.schone@ldschurch.org),  
[nielsonhe@familysearch.org](mailto:nielsonhe@familysearch.org))

# Image to Index

THE CLEVELAND PLAIN DEALER, FRIDAY, SEPTEMBER 26, 1936

**Only Three Remain**  
**AGREE ON TRAFFIC PLANS FOR SERIES**  
 City and Suburban Clubs to Ramp Idea

**Only Three Remain**  
 The city and suburban clubs have agreed on a traffic plan for the series of races to be held at the new track. The plan provides for a series of races to be held at the new track. The plan provides for a series of races to be held at the new track.

**Archbishop Holds Third Term Wrono**  
 Archbishop Healy held his third term wrono on Friday. The wrono was held at the cathedral. The archbishop presided over the wrono. The wrono was held at the cathedral.

**AS STARTLING AS THE FIRST STREAM-LINERS!**  
 NEWER HAS BUILT A NEW KIND OF CAR FOR THE LOWEST-PRICE FIELD  
**GO NASH**  
 WE SAVE MONEY EVERY WHERE!

**Compare! See These Spectacular Bargains Specially Planned for Friday and Saturday!**  
 200 Beautifully Furined CLOTH COATS \$28  
 Regularly Would Be \$37 to \$55

**104 Hollywood-Inspired FUR COATS \$88**  
 Regularly Would Be \$125 to \$225

**PERSIAN LAMBS \$165**  
**BEAVER COATS \$195**  
**SILVER FOX \$44**  
**PERSIAN FUR COATS \$18**  
**FUR JACKETS \$37**

**I. J. FOX**  
 1000 Broadway, New York City

**4 WAYS TO BUY! Cash - Charge - Budget - Layaway ... A Small Deposit Reserves Selection**



**Only Three Remain**  
 The city and suburban clubs have agreed on a traffic plan for the series of races to be held at the new track. The plan provides for a series of races to be held at the new track.

**Compare!**  
 Because of FUR SPECIALIZATION... because of being BUYING facilities... because of HOLLYWOOD STYLE sources... today, tomorrow and every day

**YOU GET MORE FOR YOUR MONEY AT I. J. FOX!**

**Compare! See These Spectacular Bargains Specially Planned for Friday and Saturday!**

**200 Beautifully Furined CLOTH COATS \$28**  
 Regularly Would Be \$37 to \$55

**104 Hollywood-Inspired FUR COATS \$88**  
 Regularly Would Be \$125 to \$225

**I. J. FOX**  
 1000 Broadway, New York City



**GERMAN**  
 Berlin

# Image to Index

THE CLEVELAND PLAIN DEALER, FRIDAY, SEPTEMBER 19, 1940

Only Three Months  
Left in 'Blitz' Dies

AGREE ON TRAFFIC  
PLANS FOR SERIES

WILL BE OPEN FRIDAY AND SATURDAY UNTIL NINE IN THE EVENING

**Compare!**  
Because of FUR SPECIALIZATION... because of  
being BUYING facilities... because of HOLLYWOOD  
STYLE sources... today, tomorrow and every day  
**YOU GET MORE FOR YOUR  
MONEY AT I. J. FOX!**

Compare! See These Spectacular Bargains  
Specially Planned for Friday and Saturday!

200 Beautifully Fur-lined  
**CLOTH COATS**  
\$28  
Regularly Would Be \$37 to \$55

104 Hollywood-Inspired  
**FUR COATS**  
\$88  
Regularly Would Be \$125 to \$225

1 of Murder, Inc.  
Are Found Guilty

AS STARTLING  
AS THE FIRST  
STREAM-  
LINERS!  
POWER HAS BUILT  
A NEW KIND  
OF CAR  
FOR THE LOWEST-  
MILE FIELD  
GO NASH  
AND SAVE MONEY  
EVERY MILE!

4 WAYS TO BUY! Cash—Charge—Budget—Layaway... A Small Deposit Reserves Selection

| Deceased              |                            |
|-----------------------|----------------------------|
| Name                  | Edwin A. Johnson           |
| Event Type            | Obituary                   |
| Event Date            | 19 Sep 1940                |
| Event Place           | Ohio, United States        |
| Gender                | Male                       |
| Age                   | 99                         |
| Birth Year            | 1841                       |
| Birthplace            | Montville, Geauga, Ohio    |
| Death Year            | 1940                       |
| Newspaper             | The Cleveland Plain Dealer |
| Spouses and Children  |                            |
| Jennett               | Wife                       |
| Mrs. Millie Leggett   | Sister                     |
| Mrs. James A. Jones   | Daughter                   |
| Mrs. H. R. Lynn       | Daughter                   |
| Mrs. Chester H. Jones | Daughter                   |
| Stuart E.             | Son                        |

# Image to Index



## Robokeyer

- Entity Tag
- Name chunk
- Relation Tag

| Deceased              |                            |
|-----------------------|----------------------------|
| Name                  | Edwin A. Johnson           |
| Event Type            | Obituary                   |
| Event Date            | 19 Sep 1940                |
| Event Place           | Ohio, United States        |
| Gender                | Male                       |
| Age                   | 99                         |
| Birth Year            | 1841                       |
| Birthplace            | Montville, Geauga, Ohio    |
| Death Year            | 1940                       |
| Newspaper             | The Cleveland Plain Dealer |
| Spouses and Children  |                            |
| Jennett               | Wife                       |
| Mrs. Millie Leggett   | Sister                     |
| Mrs. James A. Jones   | Daughter                   |
| Mrs. H. R. Lynn       | Daughter                   |
| Mrs. Chester H. Jones | Daughter                   |
| Stuart E.             | Son                        |

# Image to Index



OCR

Robokeyer

- Entity Tag
- Name chunk
- Relation Tag

| Deceased              |                            |
|-----------------------|----------------------------|
| Name                  | Edwin A. Johnson           |
| Event Type            | Obituary                   |
| Event Date            | 19 Sep 1940                |
| Event Place           | Ohio, United States        |
| Gender                | Male                       |
| Age                   | 99                         |
| Birth Year            | 1841                       |
| Birthplace            | Montville, Geauga, Ohio    |
| Death Year            | 1940                       |
| Newspaper             | The Cleveland Plain Dealer |
| Spouses and Children  |                            |
| Jennett               | Wife                       |
| Mrs. Millie Leggett   | Sister                     |
| Mrs. James A. Jones   | Daughter                   |
| Mrs. H. R. Lynn       | Daughter                   |
| Mrs. Chester H. Jones | Daughter                   |
| Stuart E.             | Son                        |

# Image to Index



Zone

OCR

Robokeyer

- Entity Tag
- Name chunk
- Relation Tag

| Deceased              |                            |
|-----------------------|----------------------------|
| Name                  | Edwin A. Johnson           |
| Event Type            | Obituary                   |
| Event Date            | 19 Sep 1940                |
| Event Place           | Ohio, United States        |
| Gender                | Male                       |
| Age                   | 99                         |
| Birth Year            | 1841                       |
| Birthplace            | Montville, Geauga, Ohio    |
| Death Year            | 1940                       |
| Newspaper             | The Cleveland Plain Dealer |
| Spouses and Children  |                            |
| Jennett               | Wife                       |
| Mrs. Millie Leggett   | Sister                     |
| Mrs. James A. Jones   | Daughter                   |
| Mrs. H. R. Lynn       | Daughter                   |
| Mrs. Chester H. Jones | Daughter                   |
| Stuart E.             | Son                        |



# Zoning

THE CLEVELAND PLAIN DEALER, FRIDAY, SEPTEMBER 20, 1936

**Only Three Known 'BUY IN BLUE' Dies**  
 (Ohio State Grotesque)  
 (Ohio State Grotesque)  
 (Ohio State Grotesque)

**AGREE ON TRAFFIC PLANS FOR SERIES**  
 City and Ball Officials Give 10 Ramp Idea

**Ohio State Grotesque**  
 Opens Akron Rally

**GERMAN**  
*Gardin*

**WE'LL BE OPEN FRIDAY AND SATURDAY UNTIL NINE IN THE EVENING**

**Compare!**  
 Because of FUR SPECIALIZATION... because of huge BUYING facilities... because of HOLLYWOOD STYLE sources... today, tomorrow and every day

**YOU GET MORE FOR YOUR MONEY AT I. J. FOX!**

**Compare! See These Spectacular Bargains Specially Planned for Friday and Saturday!**

**240 Beautifully Furrowed CLOTH COATS \$28**  
 Regularly Would Be \$37 to \$55

**184 Hollywood-Inspired FUR COATS \$88**  
 Regularly Would Be \$125 to \$225

**PERSON LAINES \$165**  
**HEAVY COATS \$195**  
**FUR JACKETS \$37**

**I. J. FOX**  
 114 HOLLYWOOD BLVD. CLEVELAND, OHIO

**4 WAYS TO BUY! Cash—Charge—Budget—Layaway... A Small Deposit Reserves Selection**

THE CLEVELAND PLAIN DEALER, FRIDAY, SEPTEMBER 20

**STREETWAY IN BLUE**  
 (Ohio State Grotesque)  
 (Ohio State Grotesque)  
 (Ohio State Grotesque)

**DOUBLE FEATURES**  
 (Ohio State Grotesque)

**MILLS RESTAURANT**  
 (Ohio State Grotesque)

**STARTLING**  
 (Ohio State Grotesque)

**THE FIRST**  
 (Ohio State Grotesque)

**STREAMLINERS**  
 (Ohio State Grotesque)

**NEW WIN**  
 (Ohio State Grotesque)

**OF CAR**  
 (Ohio State Grotesque)

**IN THE LOWEST**  
 (Ohio State Grotesque)

**OF CASH**  
 (Ohio State Grotesque)

**IN ALL MONEY**  
 (Ohio State Grotesque)

**OF MONEY**  
 (Ohio State Grotesque)



**AGREE ON TRAFFIC PLANS FOR SERIES**  
 City and Ball Officials Give 10 Ramp Idea

**Ohio State Grotesque**  
 Opens Akron Rally

**GERMAN**  
*Gardin*

**GERMAN**  
*Gardin*

**WE'LL BE OPEN FRIDAY AND SATURDAY UNTIL NINE IN THE EVENING**

**Compare!**  
 Because of FUR SPECIALIZATION... because of huge BUYING facilities... because of HOLLYWOOD STYLE sources... today, tomorrow and every day

**YOU GET MORE FOR YOUR MONEY AT I. J. FOX!**

**Compare! See These Spectacular Bargains Specially Planned for Friday and Saturday!**

**240 Beautifully Furrowed CLOTH COATS \$28**  
 Regularly Would Be \$37 to \$55

**184 Hollywood-Inspired FUR COATS \$88**  
 Regularly Would Be \$125 to \$225

**PERSON LAINES \$165**  
**HEAVY COATS \$195**  
**FUR JACKETS \$37**

**I. J. FOX**  
 114 HOLLYWOOD BLVD. CLEVELAND, OHIO

**4 WAYS TO BUY! Cash—Charge—Budget—Layaway... A Small Deposit Reserves Selection**

# Zoning Challenges

n Capitol  
it's five-  
sought to  
of shoring  
nt across  
haps to-  
President  
today in  
i destiny  
ie giant"  
e nations  
an plunge  
ited  
sly, White  
adminis-  
gness to  
bolster-  
nst Com-

tion ba  
President  
tion toda  
after he  
gree from  
The pr  
Washingt  
three and  
turn from  
Mexico C  
Senator  
him back  
Cold  
Cold, w  
anticipate  
to pack h  
his line o  
cold, dov

- Newspaper content is very dense
  - Distance between words within a column can be greater than distance between columns.



# Content Filtering

- QuickOCR
  - Less accurate
  - More performant
- Text tiling
  - Group adjacent zones together
  - Uses a cosine similarity metric to predict when blocks from different zones should merge
- BMD detector
  - Identify any content containing Birth/Marriage/Death information
  - Uses support vector machines on ngrams of characters to predict which blocks of data appear to be BMDs and which appear to NOT be.

# Image to Index



Zone

Quick OCR

Text tile

BMD  
Detect

OCR

Robokeyer

- Entity Tag
- Name chunk
- Relation Tag

| Deceased              |                            |
|-----------------------|----------------------------|
| Name                  | Edwin A. Johnson           |
| Event Type            | Obituary                   |
| Event Date            | 19 Sep 1940                |
| Event Place           | Ohio, United States        |
| Gender                | Male                       |
| Age                   | 99                         |
| Birth Year            | 1841                       |
| Birthplace            | Montville, Geauga, Ohio    |
| Death Year            | 1940                       |
| Death Place           | The Cleveland Plain Dealer |
| Spouses and Children  |                            |
| Jennett               | Wife                       |
| Mrs. Millie Leggett   | Sister                     |
| Mrs. James A. Jones   | Daughter                   |
| Mrs. H. R. Lynn       | Daughter                   |
| Mrs. Chester H. Jones | Daughter                   |
| Stuart E.             | Son                        |

# Results

- Proof of concept which met our expectations
- Would require more work to improve accuracy
- Production-based system would require 90-95% F-score
- We believe target is attainable