# A Green Form-Based Information Extraction System for Historical Documents

Tae Woo Kim

#### Motivation

Large Volume of Historical Documents



Automated solutions: our only hope for success

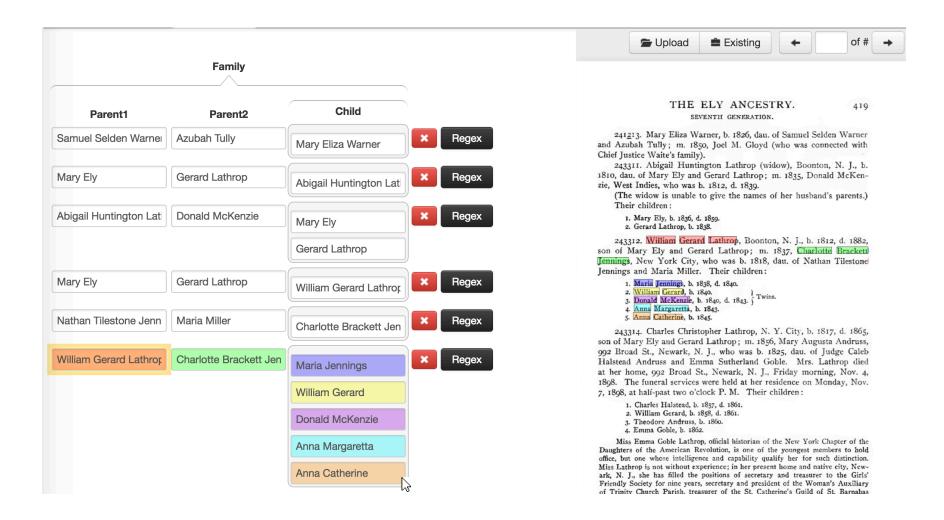
#### GreenFIE

"Green": systems that improve with use

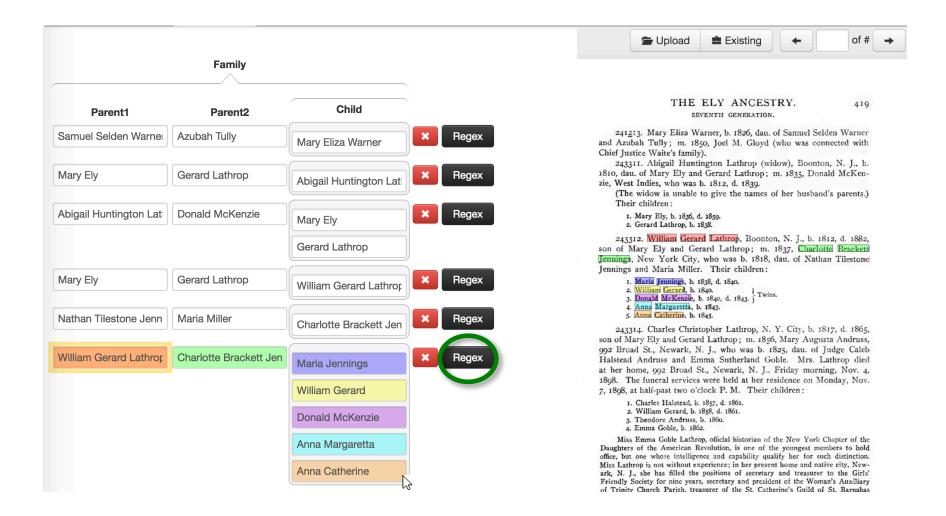
"FIE": Form-based Information Extraction

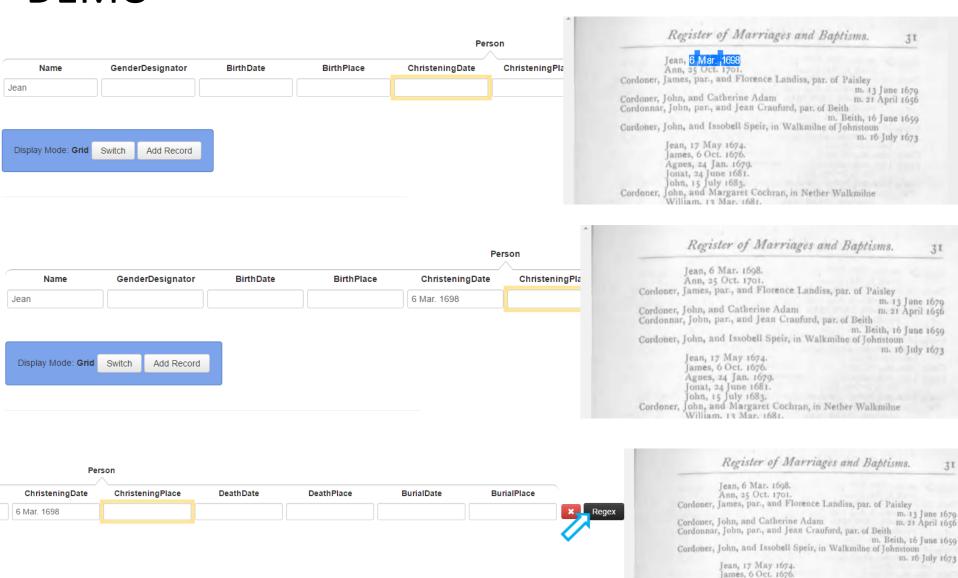
- GreenFIE
  - "watches" users fill in forms
  - learns & executes rules for form filling

# Extraction by Form Filling



# "Green" Extraction by Form Filling





Agnes, 24 Jan. 1679. Jonat, 24 June 1681. John, 15 July 1683.

William, 13 Mar, 1681.

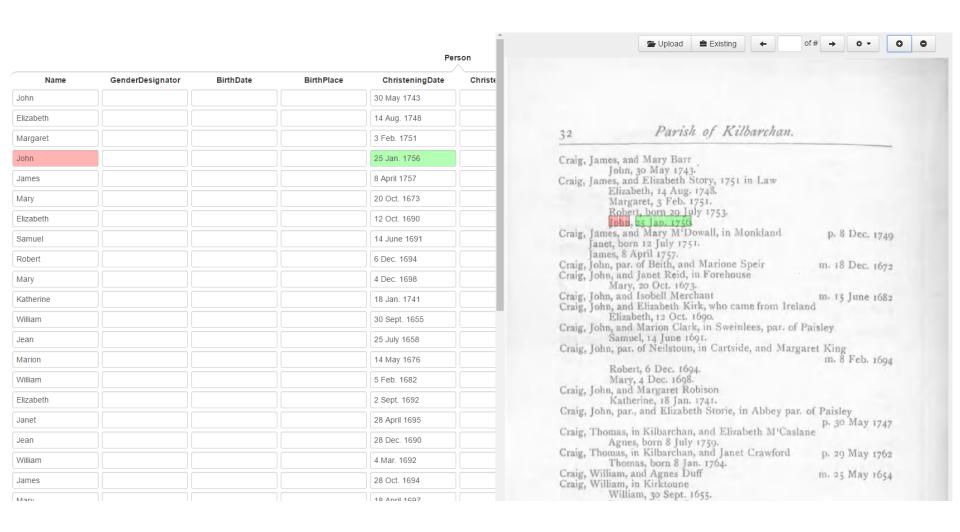
Cordoner, John, and Margaret Cochran, in Nether Walkmilne

				Per	rson	Register of Marriages and Baptisms. 31
Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	ChristeningPlace	Jean, 6 Mar. 1698.
Jean				6 Mar. 1698		Ann. 25 Oct. 1701.
Ann				25 Oct. 1701		Cordoner, James, par., and Florence Landiss, par. of Paisley m. 13 June 1679
Jean				17 May 1674		Cordoner, John, and Catherine Adam m. 21 April 1656 Cordonnar, John, par., and Jean Craufurd, par. of Beith
James				6 Oct. 1676		m. Beith, 16 June 1650
Agnes				24 Jan. 1679		Cordoner, John, and Issobell Speir, in Walkmilne of Johnston m. 16 July 1673
Jonat				24 June 1681		Jean, 17 May 1674.
John				15 July 1683		James, 6 Oct. 1676. Agnes, 24 Jan. 1679.
William				13 Mar. 1681		Jonat, 24 June 1681. John, 15 July 1683.
Jonet				3 Dec. 1682		Cordoner, John, and Margaret Cochran, in Nether Walkmilne
Thomas				7 Aug. 1688		William, 13 Mar. 1681. Cordner, John, and Jonet Cochran, in Walkinshaw, 1688 in
Margaret				16 Dec. 1692		Walkmiln of Johnstoun m. 22 April 1680 Jonet, 3 Dec. 1682.
Jean				16 Feb. 1696		Thomas, 7 Aug. 1688.
William				1 Aug. 1651		Margaret, 16 Dec. 1692. Jean, 16 Feb. 1696.
William				2 Jan. 1653		Cordonar, William, and Jean Cochran m. 7 Feb. 1651
Janet				26 Feb. 1654		Cordoner, William William, 1 Aug. 1651.
John				10 Nov. 1654		William, 2 Jan. 1653. Janet, 26 Feb. 1654.
Margaret				27 July 1655		Cordoner, William, in Achindinane
Jean				25 Sept. 1657		John, 10 Nov. 1654. Cordonar, William, in Over Wakmilne of Johnstoun
Margaret				24 June 1660		Margaret, 27 July 1655.
Jane				23 April 1658		Margaret, 24 June 1660
James				29 May 1659		Cordonner, William, in Achinames. Iane. 23 April 1658.

vviiiam	2 Jan. 1653
Janet	26 Feb. 1654
John	10 Nov. 1654
Margaret	27 July 1655
Jean	25 Sept. 1657
Margaret	24 June 1660
Jane	23 April 1658
James	29 May 1659
Thomas	25 Sept. 1657
Jean	2 Oct. 1674
Margaret	9 Dec. 1688
Agnes	15 Feb. 1691
Jean	10 Feb. 1693
Eliza	28 July 1695
Jean	31 July 1698
Jonet	28 July 1682
John	6 Nov. 1769
James	31 May 1695
Margaret	19 Sept. 1697

Display Mode: **Grid** Switch Add Record

William, 2 Jan. 1053. Janet, 26 Feb. 1654. Cordoner, William, in Achindinane John, 10 Nov. 1654. Cordonar, William, in Over Wakmilne of Johnstoun Margaret, 27 July 1655. Jean, 25 Sept. 1657. Margaret, 24 June 1660. Cordonner, William, in Achinames. Jane, 23 April 1658. James, 29 May 1659. Cordownar, William, in Nether Walkmylne Thomas, 25 Sept. 1657. Cordoner, William, and Issobell Young, in Auchnames Jean, 2 Oct. 1674. Cordoner, William, at the Wakmilns of Johnstoun Margaret, 9 Dec. 1688. Cordoner, William, and I liza Orr, in Netherwalkmilne of Johnstonn Agnes, 15 Feb. 1691. Jean, 10 Feb. 1603. Eliza, 28 July 1695. Jean, 31 July 1698. Corss, John, and Jean Patison Jonet, 28 July 1682. Couper, James, and Issobell Load m. 30 Nov. 1682 Couper, James, par. of Erskine, and Mary Black, par. 30 Mar. 1744 Coupar, William, in Kilbarchan, and Janet Caldwell p. 29 Dec. 1768 John, 6 Nov. 1769. Cowan, Daniel, in town par, of Paisley, and Margaret Dougal p. 15 Jan. 1763 Craig, James, par. of Kilbryde, and Jonet Cordonar, parm. 28 June 1658 Craig, James, Moreland in Forehouse, and Jonet Reid m. Lochwinnoch, 18 Jan. 1693 James, 31 May 1695. Margaret, 19 Sept. 1697.



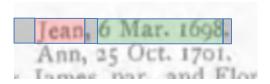
Mary	4 Dec. 1698	Parish of Kilbarchan.
Catherine	18 Jan. 1741	
Villiam	30 Sept. 1655	Craig, James, and Mary Barr John, 30 May 1743.
ean	25 July 1658	Craig, James, and Elizabeth Story, 1751 in Law
Marion	14 May 1676	Elizabeth, 14 Aug. 1748. Margaret, 3 Feb. 1751.
Villiam	5 Feb. 1682	Robert, born 29 July 1753 John, 25 Jan. 1756.
Elizabeth	2 Sept. 1692	Craig, James, and Mary M'Dowall, in Monkland p. 8 Dec. 174
anet	28 April 1695	Janet, born 12 July 1751. James, 8 April 1757.
ean	28 Dec. 1690	Craig, John, par. of Beith, and Marione Speir m. 18 Dec. 167 Craig, John, and Janet Reid, in Forehouse
Villiam	4 Mar. 1692	Mary, 20 Oct. 1673.
ames	28 Oct. 1694	Craig, John, and Isobell Merchant m. 15 June 168 Craig, John, and Elizabeth Kirk, who came from Ireland
ntary ( )	18 April 1697	Elizabeth, 12 Oct. 1690.
Villiam	5 Jan. 1701	Craig, John, and Marion Clark, in Sweinlees, par. of Paisley Samuel, 14 June 1691.
Marion	18 Feb. 1653	Craig, John, par. of Neilstoun, in Cartside, and Margaret King m. 8 Feb. 169
Daniel	9 Feb. 1655	Robert, 6 Dec. 1694.
Robert		Mary, 4 Dec. 1698. Craig, John, and Margaret Robison
		Katherine, 18 Jan. 1741. Craig, John, par., and Elizabeth Storie, in Abbey par. of Paisley p. 30 May 174
Display Mode: Grid Switch Add Record		Craig, Thomas, in Kilbarchan, and Elizabeth M'Caslane Agnes, born 8 July 1759.
		Craig, Thomas, in Kilbarchan, and Janet Crawford p. 29 May 176

				Person		Craig
Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	Christe	Craig
John				30 May 1743		0.180
Elizabeth				14 Aug. 1748		
Margaret				3 Feb. 1751		Crai
Robert		29 July 1753				
John				25 Jan. 1756		Crai Crai
Janet		12 July 1751				Craig
James				8 April 1757		Crai
Mary				20 Oct. 1673		Crai
Elizabeth				12 Oct. 1690		Crai
Samuel				14 June 1691		
Robert				6 Dec. 1694		Crai
Mary				4 Dec. 1698		Crai
Katherine				18 Jan. 1741		Crai
Agnes		8 July 1759				Crai
Thomas		8 Jan. 1764				Crai Crai
William				30 Sept. 1655		Crai
Jean				25 July 1658		Crai

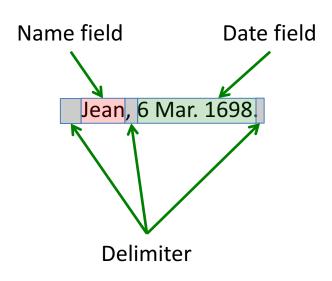
ig, James, and Mary Barr John, 30 May 1743. g, James, and Elizabeth Story, 1751 in Law Elizabeth, 14 Aug. 1748. Margaret, 3 Feb. 1751. Robert, born 29 July 1753 John, 25 Jan, 1756. ig, James, and Mary M'Dowall, in Monkland p. 8 Dec. 1749 Janet, born 12 July 1751. James, 8 April 1757. g, John, par. of Beith, and Marione Speir m. 18 Dec. 1672 g, John, and Janet Reid, in Forehouse Mary, 20 Oct. 1673. g, John, and Isobell Merchant m. 15 June 1682 g, John, and Elizabeth Kirk, who came from Ireland Elizabeth, 12 Oct. 1690. g, John, and Marion Clark, in Sweinlees, par. of Paisley Samuel, 14 June 1691. g, John, par. of Neilstoun, in Cartside, and Margaret King m. 8 Feb. 1694 Robert, 6 Dec. 1694. Mary, 4 Dec. 1698. g, John, and Margaret Robison Katherine, 18 Jan. 1741. ig, John, par., and Elizabeth Storie, in Abbey par. of Paisley p. 30 May 1747 g, Thomas, in Kilbarchan, and Elizabeth M'Caslane Agnes, born 8 July 1759. g, Thomas, in Kilbarchan, and Janet Crawford p. 29 May 1762 Thomas, born 8 Jan. 1764. g, William, and Agnes Duff m. 25 May 1654 g, William, in Kirktonne William, 30 Sept. 1655. Jean, 25 July 1658. g, William, and Marion Broune, in Locherside Marion, 14 May 1676.

#### **GreenFIE Rule Creation**

#### image text



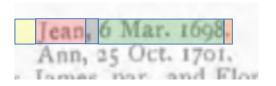
#### underlying OCR text



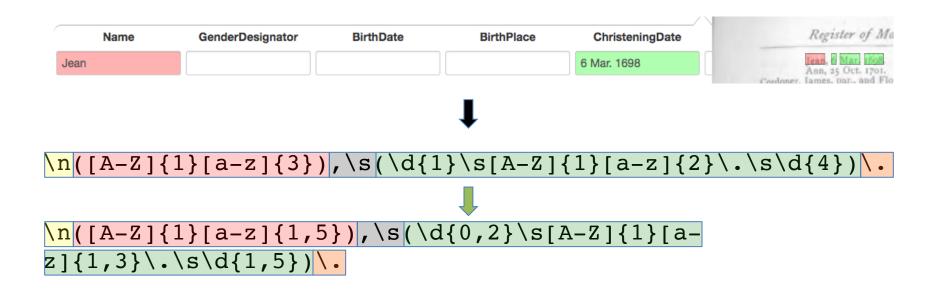
#### Rule Creation Basics

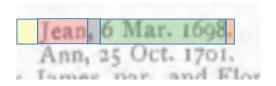




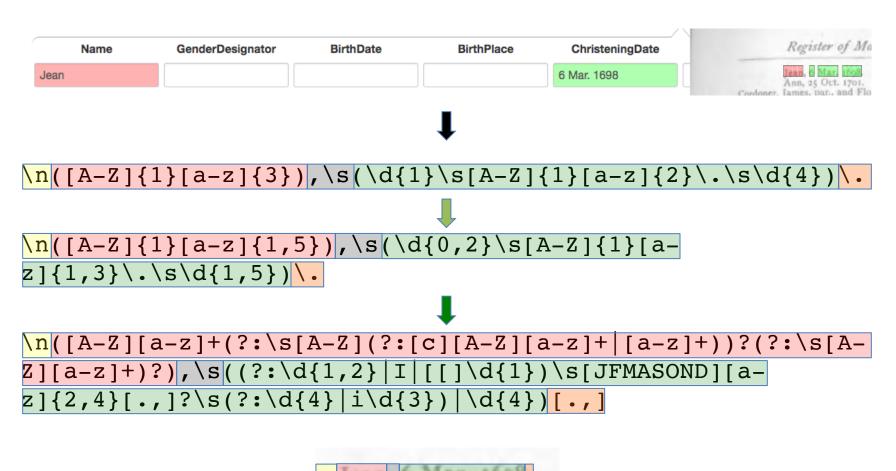


#### Rule Creation Basics





#### Rule Creation Basics



# Rule Creation for Simple Records

```
Their children:

1. Mary Ely, b. 1836, d. 1859.
2. Gerard Lathrop, b. 1838.

243312. William Gerard Lathrop, 1
```

# Rule Creation for Complex Records

```
\n\d{5,7}[.,]\s\([A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]\s[\s\S]{45,99}?\d{4}[.,]\s\([A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]\s[\s\S]{62,136}?\n1[.,]\s([A-Z][a-z]+)?(?:\s[A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]
```

```
2. Gerard Lathrop, b. 1838.

243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882, son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone Jennings and Maria Miller. Their children:

1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840.
3. Donald McKenzie, b. 1840, d. 1843.
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
```

# Rule Creation for Complex Records

```
\n\d{5,7}[.,]\s\([A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]\s[\s\S]{45,99(?)}d{4}[.,]\s\([A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]\s[\s\S]{62,136\}?)n1[.,]\s\([A-Z][a-z]+(?:\s[A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]
```

```
2. Gerard Lathrop, b. 1838.

243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882, son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone Jennings and Maria Miller. Their children:

1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840.
3. Donald McKenzie, b. 1840, d. 1843.
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
```

# Rule Creation for Complex Records

```
\n\d{5,7}[.,]\s([A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-
z_{1+})?(?:\s[A-Z][a-z]+)?)[.,]\s[\s\s]{45,99}?\d{4}[.,]\s
([A-Z][a-z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z]
Z][a-z]+)?)[.,]\s[\s\S]{62,136}?\n1[.,]\s[?:[A-Z][a-x]]
z]+(?:\s[A-Z](?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]
z]+)?)[.,]\s[\s\S]{0,43}?\n2[.,]\s(?:[A-Z][a-z]+(?:\s[A-
Z | (?:[c][A-Z][a-z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,] 
[\s\s] \{0,43\}? \n3[.,]\s([A-Z][a-Z]+(?:\s[A-Z](?:[c][A-Z][a-Z])
z]+|[a-z]+))?(?:\s[A-Z][a-z]+)?)[.,]
               2. Gerard Lathrop, b. 1838.
             243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
         son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
         Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
         Jennings and Maria Miller. Their children:
               1. Maria Jennings, b. 1838, d. 1840.
               2. William Gerard, b. 1840.
               3. Donald McKenzie, b. 1840, d. 1843.
               4. Anna Margaretta, b. 1843.
               5. Anna Catherine, b. 1845.
```

# Field Experiments

Experimental setup

Development test set

Blind test set

THE ELY ANCESTRY. 41

241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with Chief Justice Waite's family).
243311. Abjezil Huntington Lathron (widow). Roonton, N. L. b.

243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b. 1810, dau. of Mary Ely and Gerard Lathrop; m. 1835, Donald McKenzie, West Indies, who was b. 1812, d. 1839.

(The widow is unable to give the names of her husband's parents.)
Their children:

Mary Ely, b. 1836, d. 1859.
 Gerard Lathrop, b. 1838.

243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882, son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone Jennings and Maria Miller. Their children:

- 1. Maria Jennings, b. 1838, d. 1840.
- 2. William Gerard, b. 1840. 3. Donald McKenzie, b. 1840, d. 1843.
- 4. Anna Margaretta, b. 1843. 5. Anna Catherine, b. 1845.

243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop; m. 1856, Mary Augusta Andruss, 992 Broad St., Newark, N. J., who was b. 1825, du. of Judge Caleb Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died

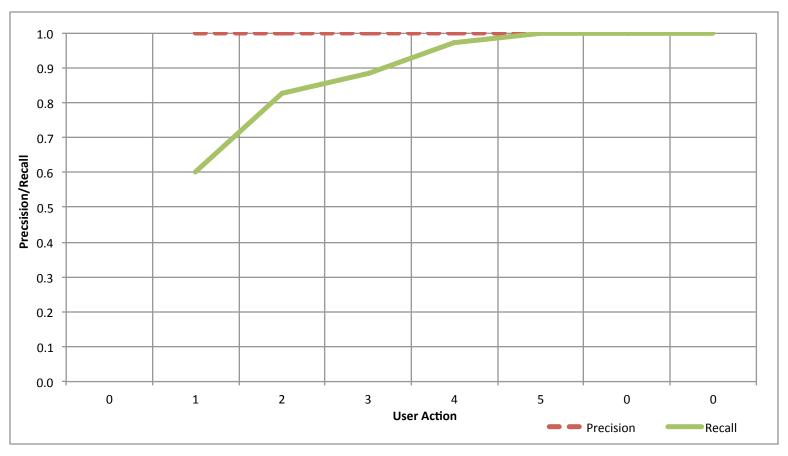
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4, 1898. The funeral services were held at her residence on Monday, Nov. 7, 1898, at hali-past two o'clock P. M. Their children:

- Charles Halstead, b. 1857, d. 1861.
   William Gerard, b. 1858, d. 1861.
- William Gerard, D. 1858, d.
   Theodore Andruss, b. 1860.
- Theodore Andruss, b. 18
   Emma Goble, b. 1862.

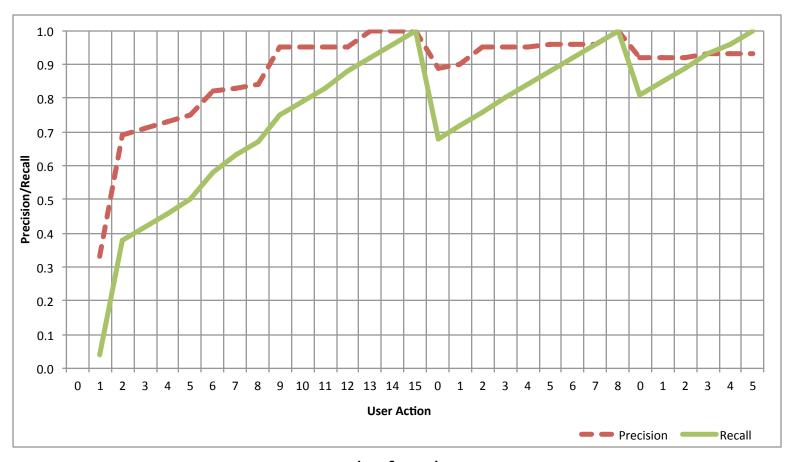
Miss Remsa Gohle Lathrop, edicial historian of the New York Chapter of the Daughters of the American Revolution, is one of the youngest sensibers to hold office, but one whose intelligence and capability qualify her for such distinction. Miss Lathrop is not visitoust experience; in her present home and native city, Newark, N. J., abe has filled the positions of secretary and resaurer to the Girls of Trimity Camera Parish, treasurer of the S. Cathroin's Golf of S. Barushas Houghtal, and manager of several of Newark's charitable institutions which here are adjusted to the S. Cathroin's Golf of S. Barushas Houghtal, and manager of several of Newark's charitable institutions which here are adjusted to the S. Cathroin's Golf of S. Barushas Houghtal, and manager of several of Newark's charitable institutions which here are adjusted to the several products and the several products are several to the several products and the several products are several to the several products and the several products are several to the several products and the several products are several products are several products and the severa

Register of Marringes and Baptims. 29

Gehran, Nimin, in Palatey
Agen, 32 Dec. 1956.
Gehran, Sager, 32 Dec. 1956.
Gehran, Bohert, 1954 in Bayeryas
Bayer, 32 Dec. 1958.
Gehran, Bohert, 1954 in Bayeryas
John Sager, 1956.
Gehran, Bohert, 1954 in Bayeryas
John Sager, 1956.
Gehran, Bohert, 1957 and John Sager, 1957.
Helen, 22 Aug. 1952.
Gehran, Bohert, 1957.
Helen, 22 Aug. 1958.
Gehran, Bohert, 1957.
Gehran, Robert, 1957.
Gehran, Bohert, 1957.
Gehran, 1957.
Gehran,



Person Form Results for Kilbarchan Pages 33 – 35

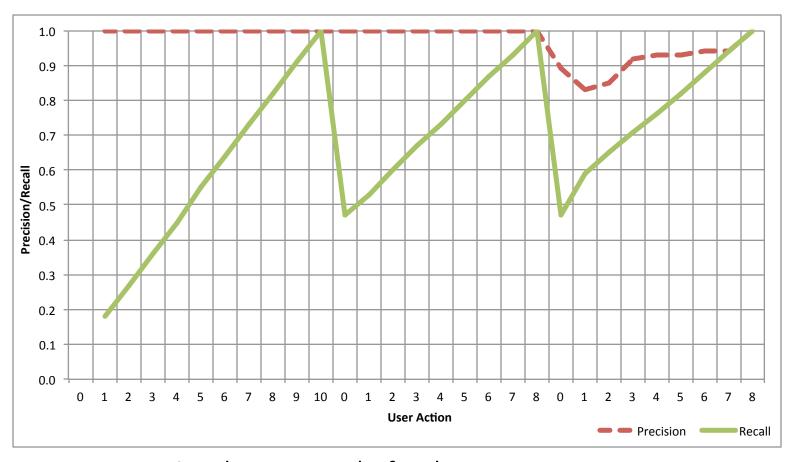


Person Form Results for Ely Pages 575 – 577

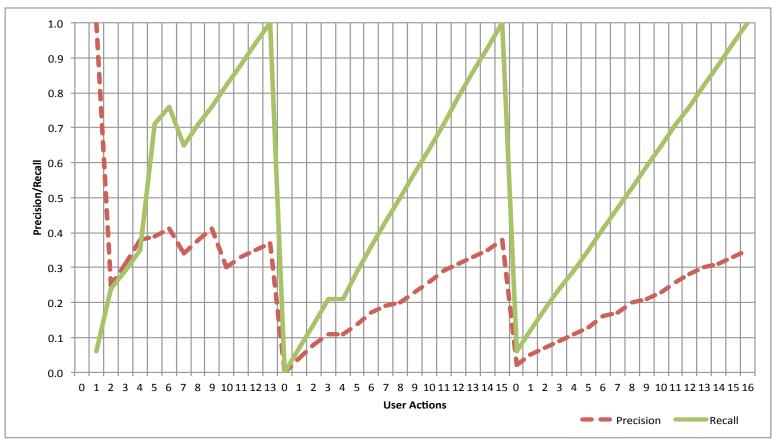
30 Parish of Kilbarchan.	THE ELY ANCESTRY. 421
	SEVENTH GENERATION.
Cochran, William, and decd. Janet Barbor, in Sandyflatt, par. of Paisley Alexander, 2 Mar. 1675.	
Cochrane, William, in Sandlmure, and Jonet King daughter, 4 Feb. 1677.	James Harvey Patterson, who was b. 1817, son of Thomas Patterson and Jane Clark. Their children:
Cochran, William, par. of Paisley, and Annas Cuming, in	and Jane Clark. Their children.
Sandflat, par. of Paisley m. 5 Oct. 1676 Elizabeth, 6 Dec. 1678.	1. Thomas Ely, b. 1843.
, son, 24 April 1681.	2. Mary, b. 1855.
James, 18 Mar. 1683. Elizabeth, 6 Sept. 1689.	3. Jeannie St. Clair, b. 1861.
Cochran, William, and Janet Cochran m. 8 July 1682	4. Harry, b. 1863.
Cochran, William, and Jonet Baxter, in Forty Shillingworth Jonet, 16 July 1682.	243327. Rev. Ben Ezra Stiles Ely, Ottumwa, Ia., b. 1828, son of
Cochran, William, and Janet Akin, in Barr Janet, 20 Feb. 1698.	Rev. Ezra Stiles Ely and Mary Ann Carswell; m. 1848, Elizabeth Eu-
William, 26 June 1702.	dora McElroy, West Ely, Mo., who was b. 1829, d. 1871, dau. of Abra-
William, 2 July 1704. Cochran, William, in Sandyflat, and Anne Hall	ham McElroy and Mary Ford Radford; m. 2nd, 1873, Abbie Amelia
William, 1 Nov. 1700.	Moore, Harrison, Ill., who was b. 1852, dau. of Porter Moore and Har-
Cochran, William, in Sandflatt, and Jane Henderson Anne, 20 Mar. 1710.	riet Leonard. Their children:
Cochran, William, and Margaret Paterson	
Alexander, 7 May 1710. Cochran, William, and Margaret Gibson, 1745 in Barr	I. Elizabeth B., b. 1849.
Jane, 1 Mar. 1741.  Mary, born 15-15 Sept. 1745.	2. Ben-Ezra Stiles, b. 1856. 3. George Everly Montgomery, b. 1858, d. 1877.
Elizabeth, born 3 Oct. 1747	4. Laura Elizabeth, b. 1859.
Cochran, William, in Shillingworth, and Janet Paterson	5. LaRose DeForest, b. 1861.
Hugh, 30 Oct. 1743.	6. Charles Wadsworth, b. 1863.
Martha, born 4 July 1749.  Cochran, William, par., and Mary Swan, par. of Beith p. 28 Mar. 1767	<ul><li>7. Mary Anita, b. 1865.</li><li>8. Francis Argyle, b. 1876.</li></ul>
Mary, 21 July 1769 (father dead). Colquhoun, Alexander, par. of Wester Kilpatrick, and	
Sussanna Brisban m. I Iune 1670	243320. Harriet Clarissima Ely, b. 1848, dau. of Ezra Stiles Ely
Colquboun, John, in Blackstoun, and Euphemia Paterson Anne, 12 Feb. 1702.	and Caroline Thompson Holmes; m. Beale Steenberger Blackford,
Catherine, 24 Jan. 1704.	Parkersburg, W. Va. Their children:
Alexander, 31 Mar. 1706. Margaret, 11 Jan. 1708.	1. Caroline Holmes Ely.
William, 25 June 1710. Colquhoun, Robert, and Frances Honyman p. 27 April 1771	(Mrs. Caroline Blackford, 1014 Broad St., Richmond, Va., or Rev. Robert
Colvin, John, in Kilbarchan, and Margaret Dickson	Dixon, Rector Trinity Church, Parkersburg, Va., may be able to give the date of
James, born 3 May 1769. Cordoner, Andrew, 1655 in Johnstoun	marriage).
Jeane, 29 May 1653.	7.1.1. D. Francis Ed. (See March Ca. Dhile deletie De.)
Andrø, 22 April 1655. Cordoner, Andro, and Agnes Barbor	243322. Zebulon DeForest Ely (831 No. 7th St., Philadelphia, Pa.),
Jean, 19 Jan 1673. Cordoner, Andro, and Margaret Arthur	b. 1852, son of Ezra Stiles Ely and Caroline Thompson Holmes; m. 1874, Clara Vanola Major, who was b. 1856, d. 1876, dau. of William Major
Jonet, 13 Feb. 1674.	and Deborah A. Beatty; m. 2nd, 1878, Mamie Anna Souder, who was b.
Cordoner, Andro, and Agnes Baird, at Walkmiln of Johnston.  John, 2 June 1678.	1855, dau. of Joseph William Souder and Anna Maria Walker. Their
Thomas, 23 Jan. 1681. Cordoner, Andrew, and Elizabeth King, in Netherwalkmilne of Johnston,	children:
1696 in Upper Walkmilne	
Margaret, 24 Mar. 1689. Andrew, 4 May 1690.	1. William Major, b. 1875, d. 1876.
John, 24 Jan. 1692.	<ol> <li>Zebulon DeForest, b. 1879.</li> <li>Edmund Holmes, b. 1880.</li> </ol>
Elizabeth, 8 July 1694. Andrew, 3 May 1696.	4. Edna Holmes, b. 1881



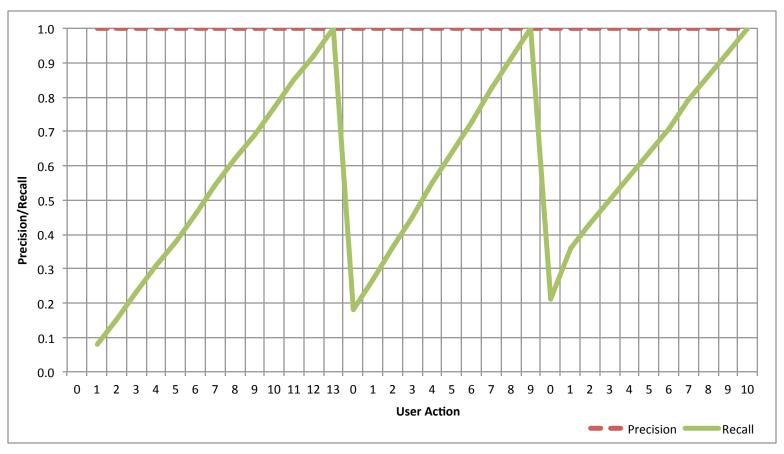
Couple Form Results for Kilbarchan Pages 33 – 35



Couple Form Results for Ely Pages 575 – 577



Family Form Results for Kilbarchan Pages 33 – 35

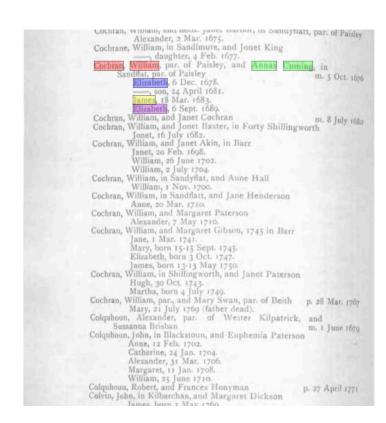


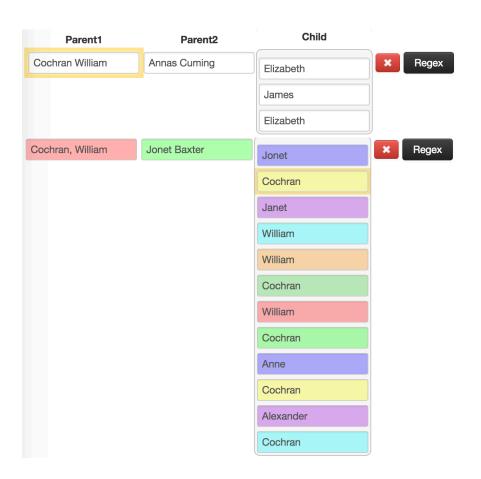
Family Form Results for Ely Pages 575 – 577

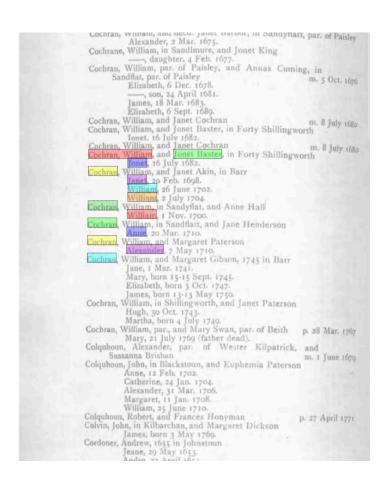
# **Results Summary**

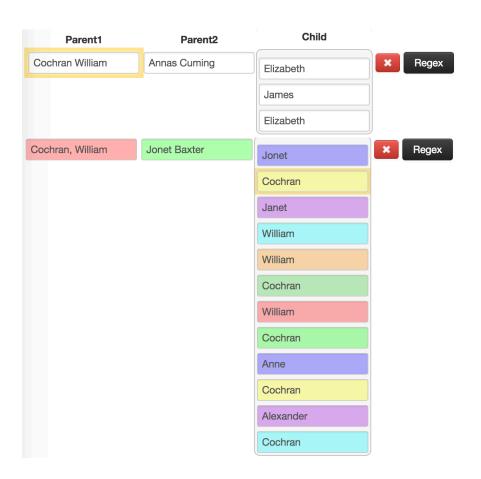
	User Action	Correct	Correct / User Action
Ely	86	157	1.83
Person	28	76	2.71
Couple	26	43	1.65
Family	32	38	1.19
Kilbarchan	64	176	2.75
Person	5	98	19.60
Couple	15	30	2.00
Family	44	48	1.09
Overall	150	333	2.22

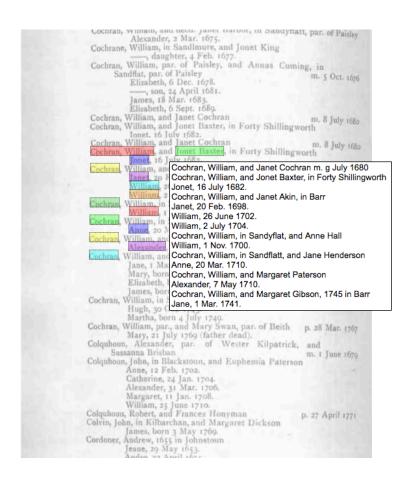


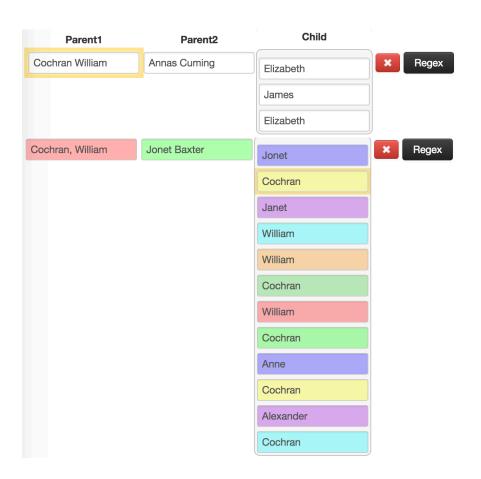


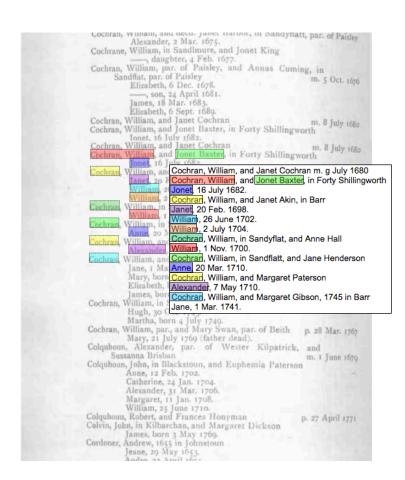






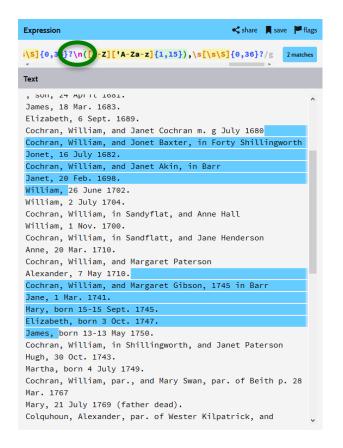


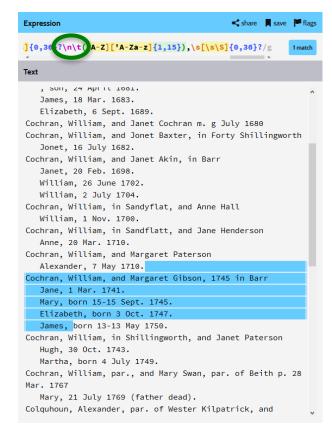




# Family Form Resolutions

- Enumerators (e.g. 1. 2., ..., n in Ely)
- Visual indentation





- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
3. Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
3. Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
  Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abiqail Huntinqton Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
3. Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
2. William Gerard, b. 18<u>5</u>8, d. 1861.
3. Theodore Andruss, b. 1860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
3. Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
3. Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY, 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family)
243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings. New York Citu. who was b. 1818. dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
3. Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss,
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7. 1898. at half-past two o'clock P. M. Their children:
1. Charles Halstead, b. 1857, d. 1861.
William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

- Value-type generalizations
- Historical documents
  - OCR errors
  - White-space layout
  - End-of-line hyphens
- Commentary
  - Pattern exceptions
  - Skip lengths

```
THE ELY ANCESTRY. 419
SEVENTH GENERATION.
241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner
and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with
Chief Justice Waite's family),
243311. Abiqail Huntington Lathrop (widow), Boonton, N. J., b.
1810, dau. of Mary Ely and Gerard Lathrop ; m. 1835, Donald McKen-
zie. West Indies, who was b. 1812, d. 1839.
(The widow is unable to give the names of her husband's parents.)
Their children :
1. Mary Ely, b, 1836, d. 1859.
2. Gerard Lathrop, b. 1838.
243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882,
son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett
Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone
Jennings and Maria Miller. Their children:
1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840. ) .
  Donald McKenzie, b. 1840, d. 1843. ]
4. Anna Margaretta, b. 1843.
5. Anna Catherine, b. 1845.
243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865,
son of Mary Ely and Gerard Lathrop ; m. 1856, Mary Augusta Andruss.
992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb
Halstead Andruss and Emma Sutherland Goble. Mrs. Lathrop died
at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4,
1898. The funeral services were held at her residence on Monday, Nov.
7, 1898, at half-past two o'clock P. M. Their children:

    Charles Halstead, b. 1857, d. 1861.

2. William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. i860.
4. Emma Goble, b. 1862.
Miss Emma Goble Lathrop, official historian of the New York Chapter of the
Daughters of the American Revolution, is one of the youngest members to hold
office, but one whose intelligence and capability qualify her for such distinction.
Miss Lathrop is not without experience; in her present home and native city, New-
ark, N. J., she has filled the positions of secretary and treasurer to the Girls'
Friendly Society for nine years, secretary and president of the Woman's Auxiliary
of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas
Hospital, and manager of several of Newark's charitable institutions which her
grandparents were instrumental in founding. Miss Lathrop traces her lineage
back through many generations of famous progenitors on both sides. Her maternal
ancestors were among the early settlers of New Jersey, among them John Ogden,
who received patent in 1664 for the purchase of Elizabethtown, and who in 1673 was
```

#### Conclusion

- GreenFIE is "green"
  - Learns from examples
  - Improves itself with use
- GreenFIE diminishes labor
  - 333 records extracted with 150 user actions
  - 19.6 records per user action (Kilbarchan, Person)
- Future work: tech-transfer to FamilySearch