

# GreenFIE: A Green Form-Based Information-Extraction System for Historical Documents

(Self-improving Extraction Systems)

Tae Woo Kim  
David W. Embley  
Stephen W. Liddle



# A Green Information Extraction System

- “Green” systems improve with use
- GreenFIE
  - Green Form-based Information Extraction
  - Generates extraction rules by watching users work
- COMET

The screenshot displays the GreenFIE interface. On the left is a data table with columns: Name, GenderDesignator, BirthDate, BirthPlace, ChristeningDate, and Child. The row for 'Jean' is highlighted in red, and the 'ChristeningDate' cell '6 Mar. 1698' is highlighted in green. On the right is a document image titled 'Register of Marriages and Baptisms.' with page number 31. The document contains handwritten text with a red box highlighting 'Mar. 6 Mar. 1698' and a green box highlighting '6 Mar. 1698'. The document text includes names like Cordoner, James, and Cordoner, John, and dates like 25 Oct. 1701, 6 Oct. 1676, etc.

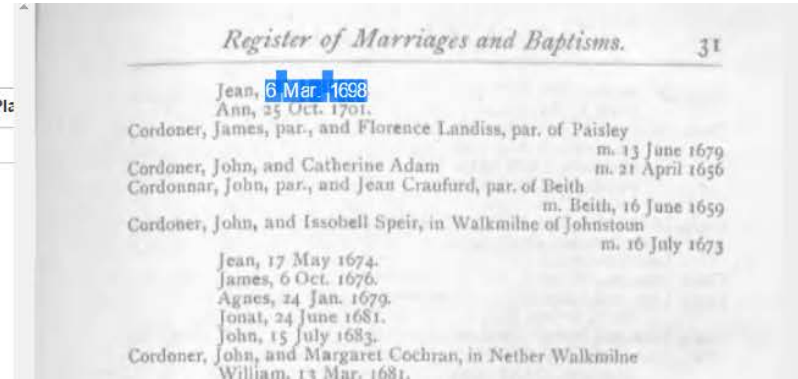
Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	Child
Jean				6 Mar. 1698	
Ann				25 Oct. 1701	
James				6 Oct. 1676	
Agnes				24 Jan. 1679	
Jonet				3 Dec. 1682	
Thomas				7 Aug. 1688	
Jean				16 Feb. 1696	
Janet				26 Feb. 1654	
John				10 Nov. 1654	
Jean				25 Sept. 1657	
Thomas				25 Sept. 1657	
Jean				2 Oct. 1674	
Agnes				15 Feb. 1691	
Jean				10 Feb. 1693	
John				6 Nov. 1769	

# DEMO

Person

Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	ChristeningPlace
Jean					

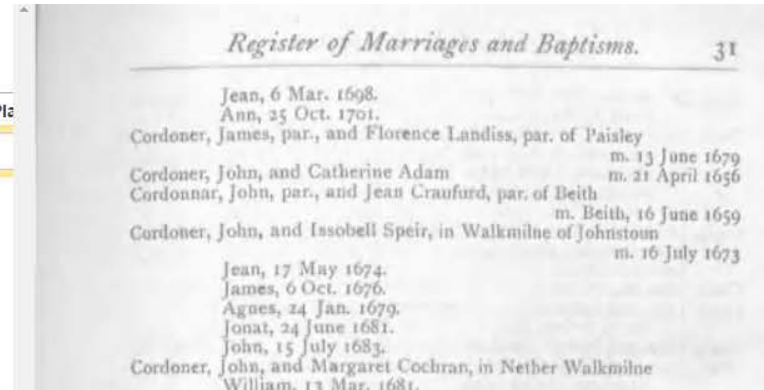
Display Mode: **Grid**



Person

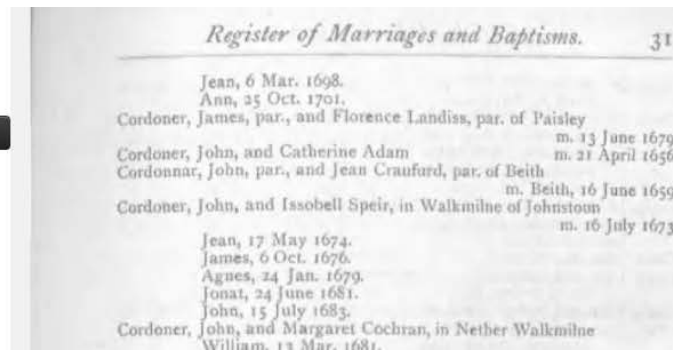
Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	ChristeningPlace
Jean				6 Mar. 1698	

Display Mode: **Grid**



Person

ChristeningDate	ChristeningPlace	DeathDate	DeathPlace	BurialDate	BurialPlace
6 Mar. 1698					



# DEMO

Person

Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	ChristeningPlace
Jean				6 Mar. 1698	
Ann				25 Oct. 1701	
Jean				17 May 1674	
James				6 Oct. 1676	
Agnes				24 Jan. 1679	
Jonat				24 June 1681	
John				15 July 1683	
William				13 Mar. 1681	
Jonet				3 Dec. 1682	
Thomas				7 Aug. 1688	
Margaret				16 Dec. 1692	
Jean				16 Feb. 1696	
William				1 Aug. 1651	
William				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	

*Register of Marriages and Baptisms.* 31

Jean, 6 Mar. 1698.  
Ann, 25 Oct. 1701.

Cordoner, James, par., and Florence Landiss, par. of Paisley  
m. 13 June 1679

Cordoner, John, and Catherine Adam  
m. 21 April 1656

Cordonnar, John, par., and Jean Craufurd, par. of Beith  
m. Beith, 16 June 1659

Cordoner, John, and Issobell Speir, in Walkmilne of Johnstoun  
m. 16 July 1673

Jean, 17 May 1674.  
James, 6 Oct. 1676.  
Agnes, 24 Jan. 1679.  
Jonat, 24 June 1681.  
John, 15 July 1683.

Cordoner, John, and Margaret Cochran, in Nether Walkmilne  
William, 13 Mar. 1681.

Cordner, John, and Jonet Cochran, in Walkinshaw, 1688 in  
Walkmiln of Johnstoun  
m. 22 April 1680

Jonet, 3 Dec. 1682.  
Thomas, 7 Aug. 1688.  
Margaret, 16 Dec. 1692.  
Jean, 16 Feb. 1696.

Cordonar, William, and Jean Cochran  
m. 7 Feb. 1651

Cordoner, William  
William, 1 Aug. 1651.  
William, 2 Jan. 1653.  
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.  
Iane, 23 April 1658.

# DEMO

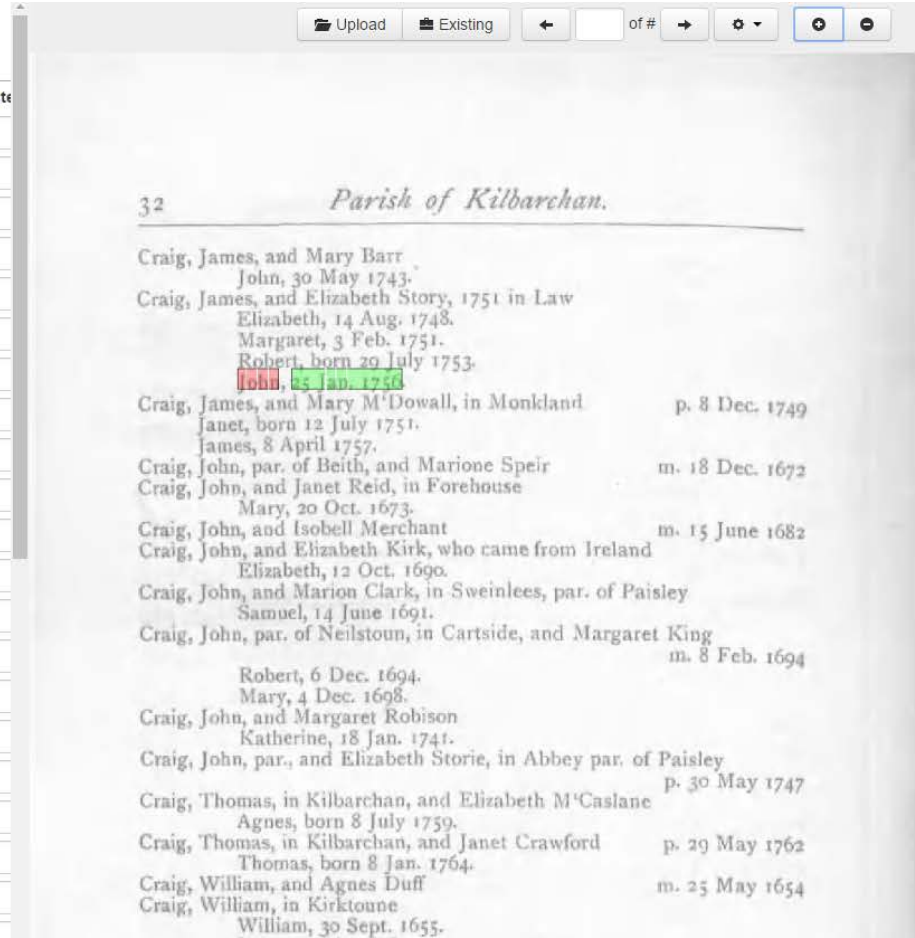
William				2 Jan. 1653	
Janet				26 Feb. 1654	
John				10 Nov. 1654	
Margaret				27 July 1655	
Jean				25 Sept. 1657	
Margaret				24 June 1660	
Jane				23 April 1658	
James				29 May 1659	
Thomas				25 Sept. 1657	
Jean				2 Oct. 1674	
Margaret				9 Dec. 1688	
Agnes				15 Feb. 1691	
Jean				10 Feb. 1693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

William, 2 Jan. 1653.  
 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.  
 Cordonnar, 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 Eliza Orr, in Netherwalkmilne of Johnstoun  
 Agnes, 15 Feb. 1691.  
 Jean, 10 Feb. 1693.  
 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, par.  
 m. 28 June 1658  
 Craig, James, Moreland in Forehouse, and Jonet Reid  
 m. Lochwinnoch, 18 Jan. 1693  
 James, 31 May 1695.  
 Margaret, 19 Sept. 1697.

Display Mode: **Grid**

# DEMO

Person					
Name	GenderDesignator	BirthDate	BirthPlace	ChristeningDate	Christe
John				30 May 1743	
Elizabeth				14 Aug. 1748	
Margaret				3 Feb. 1751	
John				25 Jan. 1756	
James				8 April 1757	
Mary				20 Oct. 1673	
Elizabeth				12 Oct. 1690	
Samuel				14 June 1691	
Robert				6 Dec. 1694	
Mary				4 Dec. 1698	
Katherine				18 Jan. 1741	
William				30 Sept. 1655	
Jean				25 July 1658	
Marion				14 May 1676	
William				5 Feb. 1682	
Elizabeth				2 Sept. 1692	
Janet				28 April 1695	
Jean				28 Dec. 1690	
William				4 Mar. 1692	
James				28 Oct. 1694	
Mary				18 April 1697	



# DEMO

Mary				4 Dec. 1698	
Katherine				18 Jan. 1741	
William				30 Sept. 1655	
Jean				25 July 1658	
Marion				14 May 1676	
William				5 Feb. 1682	
Elizabeth				2 Sept. 1692	
Janet				28 April 1695	
Jean				28 Dec. 1690	
William				4 Mar. 1692	
James				28 Oct. 1694	
Mary				18 April 1697	
William				5 Jan. 1701	
Marion				18 Feb. 1653	
Daniel				9 Feb. 1655	
Robert					

Display Mode: **Grid**



32 *Parish of Kilbarchan.*

Craig, James, and Mary Barr  
John, 30 May 1743.

Craig, James, and Elizabeth Story, 1751 in Law  
Elizabeth, 14 Aug. 1748.  
Margaret, 3 Feb. 1751.  
Robert, born 29 July 1753.  
John, 25 Jan. 1756.

Craig, James, and Mary M<sup>d</sup>Dowall, in Monkland p. 8 Dec. 1749  
Janet, born 12 July 1751.  
James, 8 April 1757.

Craig, John, par. of Beith, and Marione Speir m. 18 Dec. 1672  
Craig, John, and Janet Reid, in Forehouse  
Mary, 20 Oct. 1673.

Craig, John, and Isobell Merchant m. 15 June 1682  
Craig, John, and Elizabeth Kirk, who came from Ireland  
Elizabeth, 12 Oct. 1690.

Craig, John, and Marion Clark, in Sweinlees, par. of Paisley  
Samuel, 14 June 1691.

Craig, John, par. of Neilstoun, in Cartside, and Margaret King m. 8 Feb. 1694  
Robert, 6 Dec. 1694.  
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 1747

Craig, Thomas, in Kilbarchan, and Elizabeth M<sup>c</sup>Caslane  
Agnes, born 8 July 1759.

Craig, Thomas, in Kilbarchan, and Janet Crawford p. 29 May 1762  
Thomas, born 8 Feb. 1762.





# DEMO

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

Craig, James, and Mary Barr  
John, 30 May 1743.

Craig, James, and Elizabeth Story, 1751 in Law  
Elizabeth, 14 Aug. 1748.  
Margaret, 3 Feb. 1751.  
Robert, born 29 July 1753.  
John, 25 Jan. 1756.

Craig, James, and Mary M'Dowall, in Monkland p. 8 Dec. 1749  
Janet, born 12 July 1751.  
James, 8 April 1757.

Craig, John, par. of Beith, and Marione Speir m. 18 Dec. 1672  
Craig, John, and Janet Reid, in Forehouse  
Mary, 20 Oct. 1673.

Craig, John, and Isobell Merchant m. 15 June 1682  
Craig, John, and Elizabeth Kirk, who came from Ireland  
Elizabeth, 12 Oct. 1690.

Craig, John, and Marion Clark, in Sweinlees, par. of Paisley  
Samuel, 14 June 1691.

Craig, John, par. of Neilstoun, in Cartside, and Margaret King m. 8 Feb. 1694  
Robert, 6 Dec. 1694.  
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 1747

Craig, Thomas, in Kilbarchan, and Elizabeth M'Caslane  
Agnes, born 8 July 1759.

Craig, Thomas, in Kilbarchan, and Janet Crawford p. 29 May 1762  
Thomas, born 8 Jan. 1764.

Craig, William, and Agnes Duff m. 25 May 1654  
Craig, William, in Kirktonne  
William, 30 Sept. 1655.  
Jean, 25 July 1658.

Craig, William, and Marion Broune, in Locherside  
Marion, 14 May 1676.

# Regex Extraction Rules

Register of Marriages and Baptisms. 31

Jean, 6 Mar. 1698.

Ann, 25 Oct. 1701.

Cordoner, James, par., and Florence Landiss, par. of Paisley  
m. 13 June 1679

...

Elizabeth, 2 Sept. 1692.

- `\n([A-Z]{1}[a-z]{3}),\s(\d{1}\s[A-Z]{1}[a-z]{2}\.\s\d{4})\.`

# Regex Extraction Rules

Register of Marriages and Baptisms. 31

Jean, 6 Mar. 1698.

Ann, 25 Oct. 1701.

Cordoner, James, par., and Florence Landiss, par. of Paisley  
m. 13 June 1679

...

Elizabeth, 2 Sept. 1692.

- `\n([A-Z]{1}[a-z]{3}),\s(\d{1}\s[A-Z]{1}[a-z]{2}\.\s\d{4})\.`
- `\n([A-Z]{1}[a-z]{1,5}),\s(\d{0,2}\s[A-Z]{1}[a-z]{1,3}\.\s\d{2,6})\.`

# Regex Extraction Rules

Register of Marriages and Baptisms. 31

Name → Jean, 6 Mar. 1698. ← ChristeningDate  
Ann, 25 Oct. 1701.

Cordoner, James, par., and Florence Landiss, par. of Paisley  
m. 13 June 1679

...

Elizabeth, 2 Sept. 1692.

- `\n([A-Z]{1}[a-z]{3}),\s(\d{1}\s[A-Z]{1}[a-z]{2}\.\s\d{4})\.`
- `\n([A-Z]{1}[a-z]{1,5}),\s(\d{0,2}\s[A-Z]{1}[a-z]{1,3}\.\s\d{2,6})\.`
- `\n([A-Z]{1}[a-z]{1,10}),\s(\d{1,2}\s[JFMASOND][a-z]{1,5}\.\s\d{4})\.`

# Regex Extraction Rules

...

Margaret, 3 Feb. 1751.

Robert, born 29 July 1753.

John, 25 Jan. 1756.

Craig, James, and Mary M'Dowall, in Monkland p. 8 Dec. 1749

Janet, born 12 July 1751.

...

- `\n([A-Z]{1}[a-z]{3}),\s(\d{1}\s[A-Z]{1}[a-z]{2}\.\s\d{4})\.`
- `\n([A-Z]{1}[a-z]{1,5}),\s(\d{0,2}\s[A-Z]{1}[a-z]{1,3}\.\s\d{2,6})\.`
- `\n([A-Z]{1}[a-z]{1,10}),\s(\d{1,2}\s[JFMASOND][a-z]{1,5}\.?\s\d{4})\.`

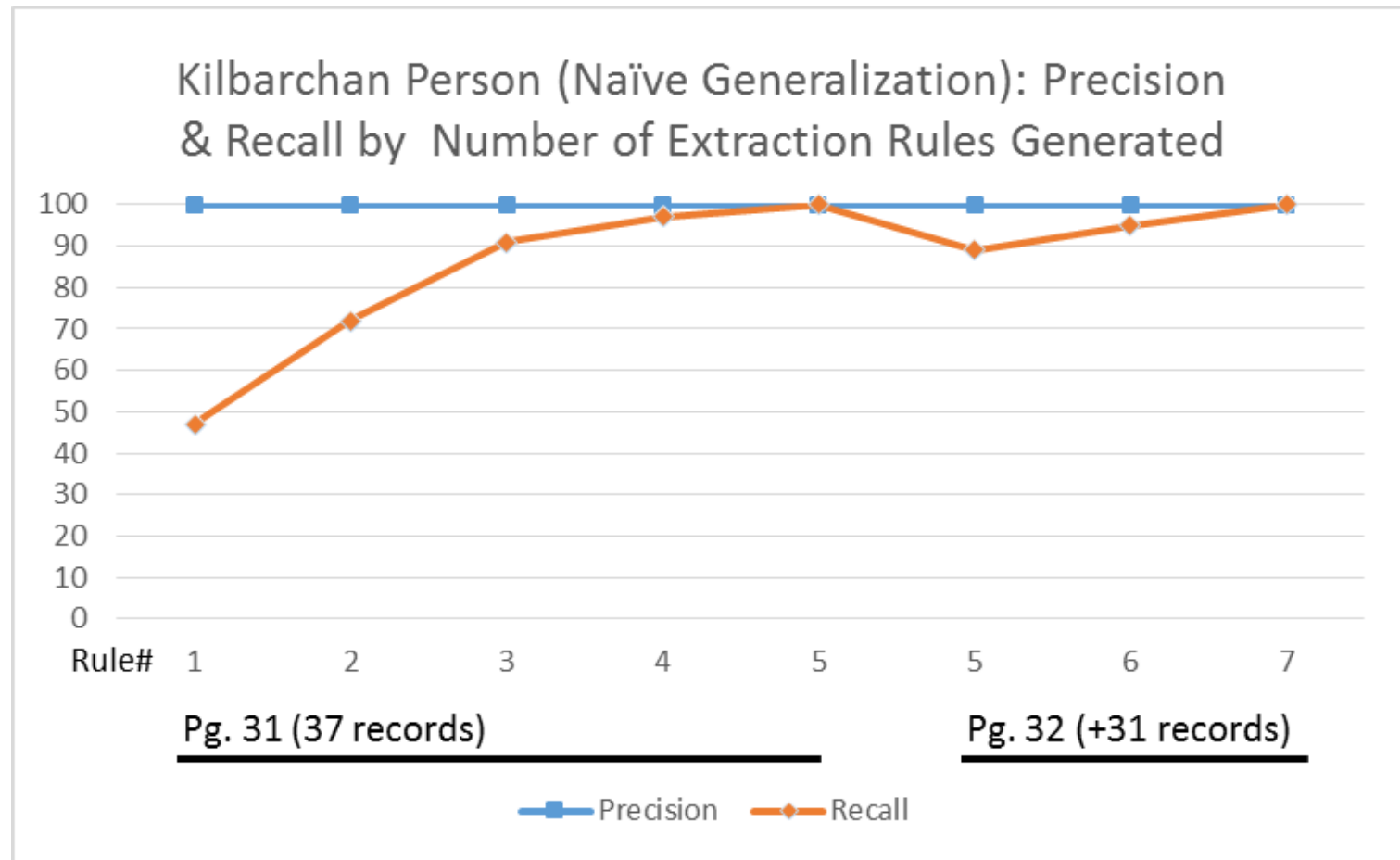
# Regex Extraction Rules

...  
Margaret, 3 Feb. 1751.  
Robert, born 29 July 1753.  
John, 25 Jan. 1756.  
Craig, James, and Mary M'Dowall, in Monkland p. 8 Dec. 1749  
Janet, born 12 July 1751.  
...

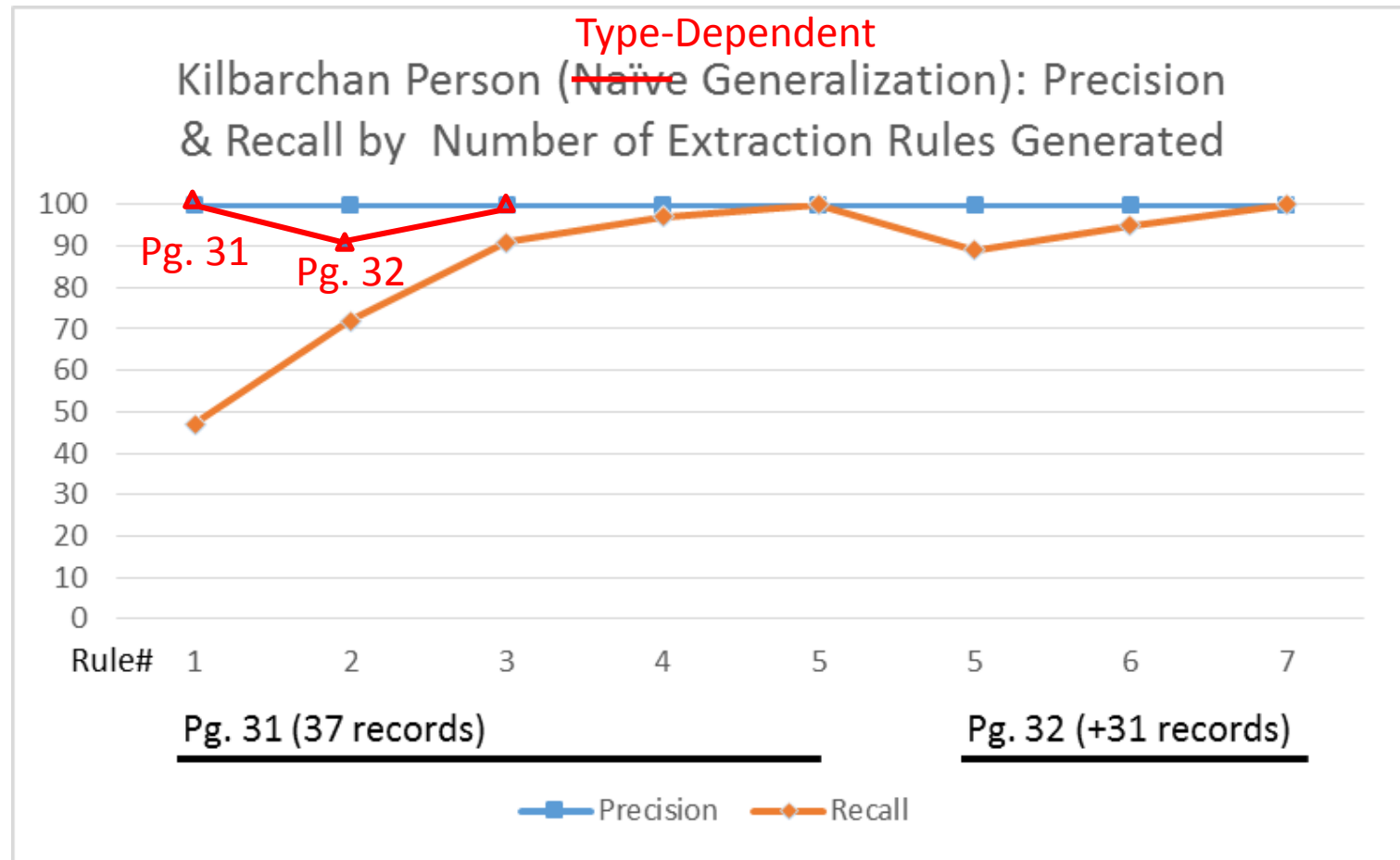
**Name** → (points to "Robert")  
**BirthDate** → (points to "29 July 1753")

- `\n([A-Z]{1}[a-z]{3}),\s(\d{1}\s[A-Z]{1}[a-z]{2}\.\s\d{4})\.`
- `\n([A-Z]{1}[a-z]{1,5}),\s(\d{0,2}\s[A-Z]{1}[a-z]{1,3}\.\s\d{2,6})\.`
- `\n([A-Z]{1}[a-z]{1,10}),\s(\d{1,2}\s[JFMASOND][a-z]{1,5}\.\s\d{4})\.`
- `\n([A-Z]{1}[a-z]{1,10}),\sborn\s(\d{1,2}\s[JFMASOND][a-z]{1,5}\.\s\d{4})\.`

# Results: Naïve Generalization



# Results: Type-Specific Generalization





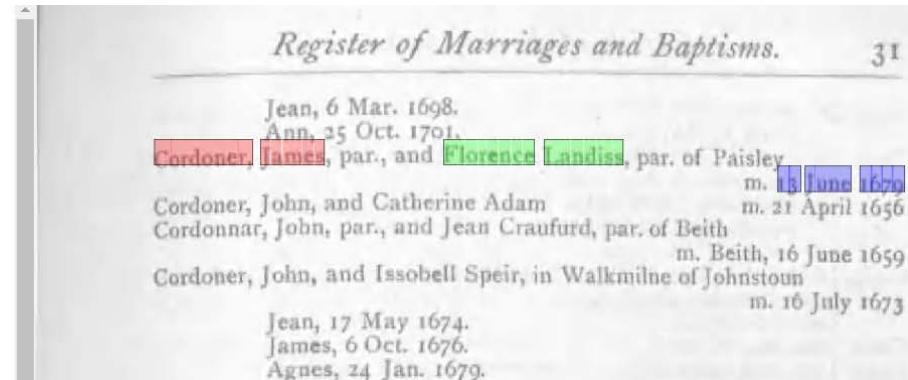
# Couple Form

Couple

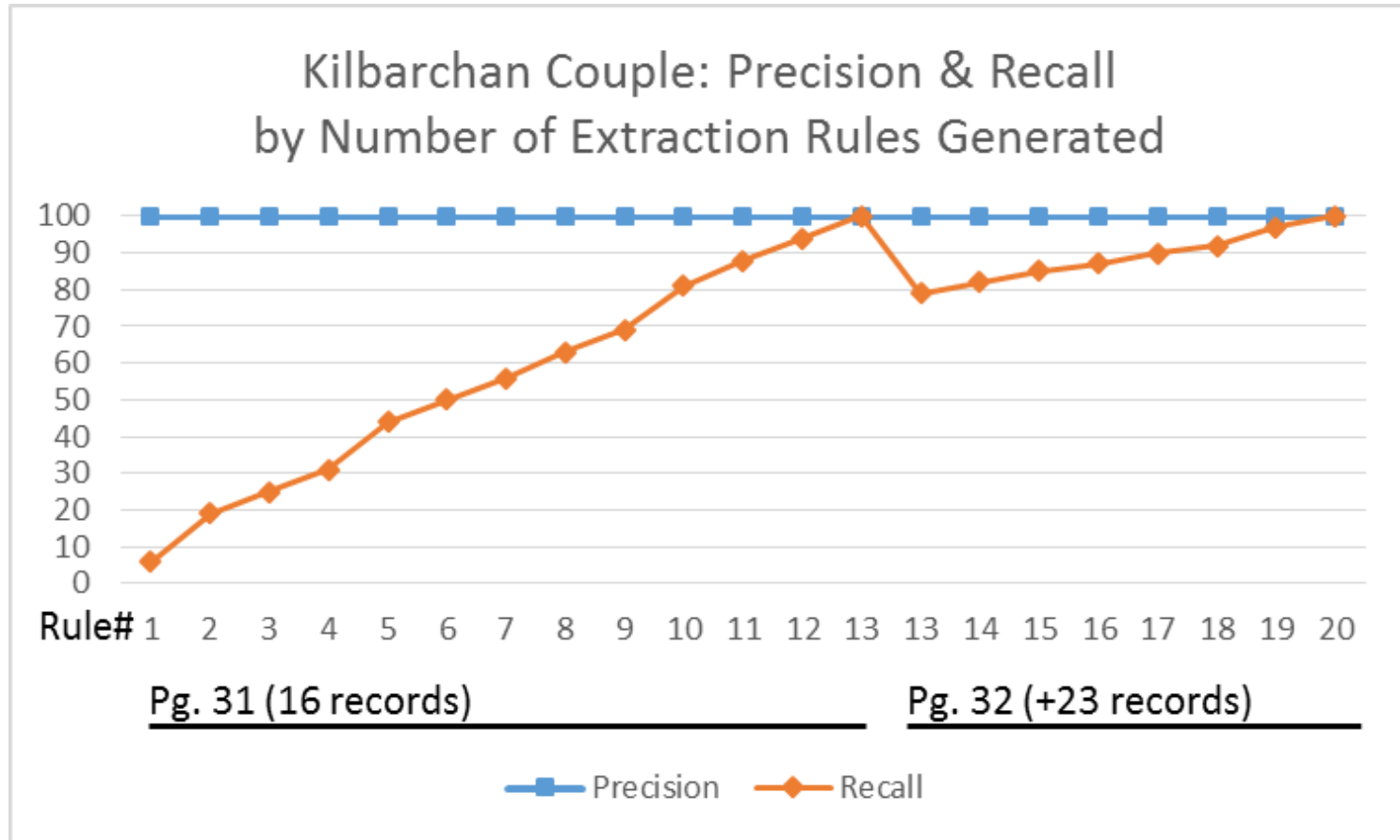
Name	SpouseName	MarriageDate	MarriagePlace
Cordoner, James	Florence Landiss	13 June 1679	

Regex

Display Mode: **Grid**



# Results: Couple Form



# Couple Patterns

Cordoner, James, par., and Florence Landiss, par. of Paisley m. 13 June 1679

Cordoner, John, and Catherine Adam m. 21 April 1656

Cordonnar, John, par., and Jean Craufurd, par. of Beith m. Beith, 16 June 1659

Cordoner, John, and Issobell Speir, in Walkmilne of Johnstoun m. 16 July 1673

Cordoner, John, and Margaret Cochran, in Nether Walkmilne

Cordner, John, and Jonet Cochran, in Walkinshaw, 1688 in Walkmiln of Johnstoun m. 22 April 1680

Cordonar, William, and Jean Cochran m. 7 Feb. 1651

Cordoner, William, and Issobell Young, in Auchnames

Cordoner, William, and Elizabeth Orr, in Netherwalkmilne of Johnstoun

Corss, John, and Jean Patison

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

Cowan, Daniel, in town par. of Paisley, and Margaret Dougal

Craig, James, par. of Kilbryde, and Jonet Cordonar, par. m. 28 June 1658

Craig, James, Moreland in Forehouse, and Jonet Reid m. Lochwinnoch, 18 Jan. 1693

# Couple Patterns

Cordoner, James, par., and Florence Landiss, par. of Paisley m. 13 June 1679

Cordoner, John, and Catherine Adam m. 21 April 1656

Cordonnar, John, par., and Jean Craufurd, par. of Beith m. Beith, 16 June 1659

Cordoner, John, and Issobell Speir, in Walkmilne of Johnstoun m. 16 July 1673

Cordoner, John, and Margaret Cochran, in Nether Walkmilne

Cordner, John, and Jonet Cochran, in Walkinshaw, 1688 in Walkmiln of Johnstoun m. 22 April 1680

Cordonar, William, and Jean Cochran m. 7 Feb. 1651

Cordoner, William, and Issobell Young, in Auchnames

Cordoner, William, and Elizabeth Orr, in Netherwalkmilne of Johnstoun

Corss, John, and Jean Patison

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

Cowan, Daniel, in town par. of Paisley, and Margaret Dougal

Craig, James, par. of Kilbryde, and Jonet Cordonar, par. m. 28 June 1658

Craig, James, Moreland in Forehouse, and Jonet Reid m. Lochwinnoch, 18 Jan. 1693

# Couple Patterns

Cordoner, James, par., and Florence Landiss, par. of Paisley m. 13 June 1679

Cordoner, John, and Catherine Adam m. 21 April 1656

Cordonnar, John, par., and Jean Craufurd, par. of Beith m. Beith, 16 June 1659

Cordoner, John, and Issobell Speir, in Walkmilne of Johnstoun m. 16 July 1673

Cordoner, John, and Margaret Cochran, in Nether Walkmilne

Cordner, John, and Jonet Cochran, in Walkinshaw, 1688 in Walkmiln of Johnstoun m. 22 April 1680

Cordonar, William, and Jean Cochran m. 7 Feb. 1651

Cordoner, William, and Issobell Young, in Auchnames

Cordoner, William, and Elizabeth Orr in Netherwalkmilne of Johnstoun

Corss, John, and Jean Patison

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

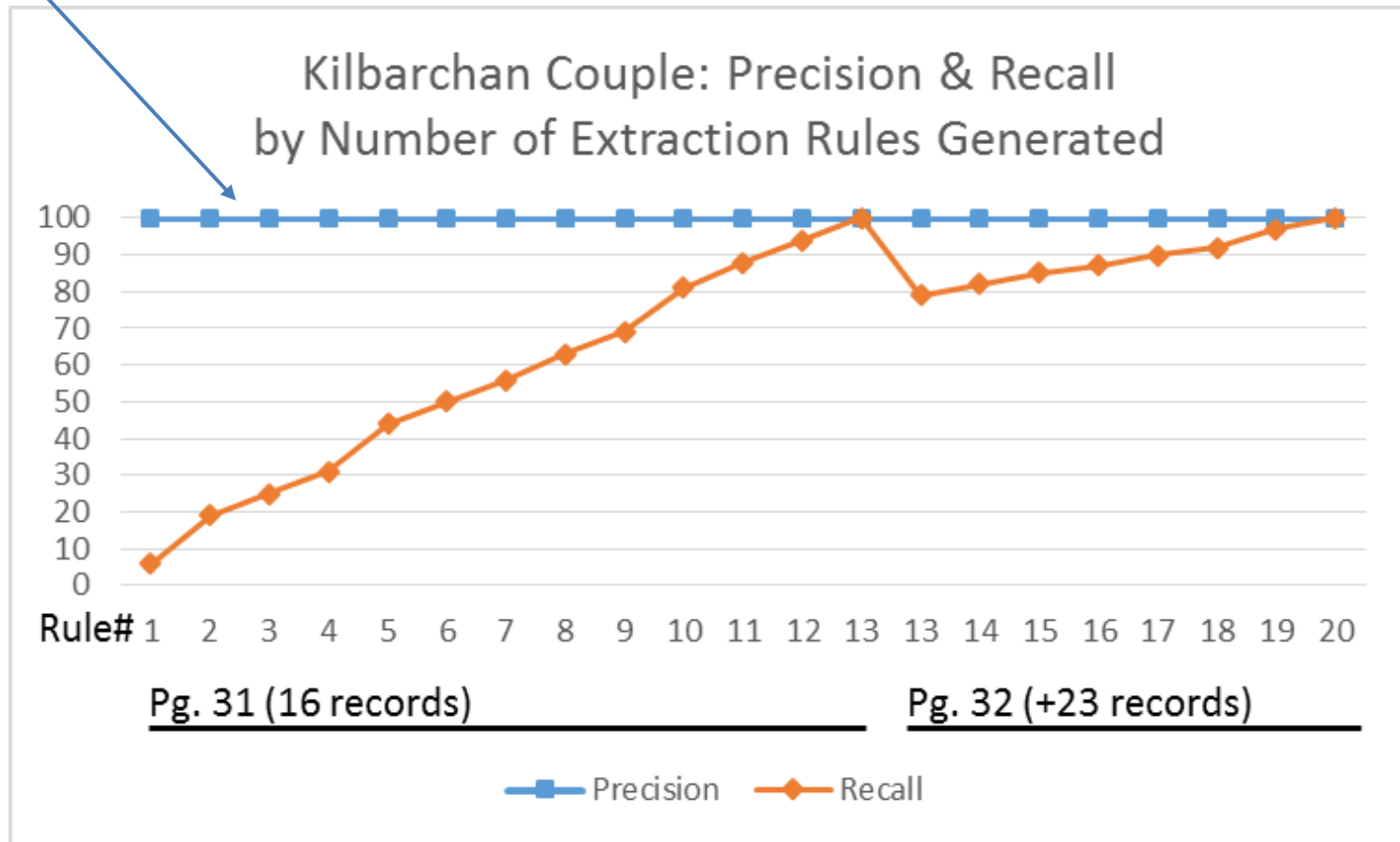
Cowan, Daniel, in town par. of Paisley, and Margaret Dougal

Craig, James, par. of Kilbryde, and Jonet Cordonar, par. m. 28 June 1658

Craig, James, Moreland in Forehouse, and Jonet Reid m. Lochwinnoch, 18 Jan. 1693

# Results: Couple Form

Precision: not necessarily 100% (over-generalization)



# Conclusions

- GreenFIE is “green”!
- Extraction-rule generation
  - Document pattern consistency
  - Number and variability of patterns
- Go green!!



[www.deg.byu.edu](http://www.deg.byu.edu)

