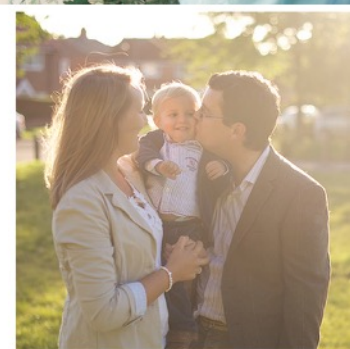


What comes after handwriting recognition?

Jon Morrey

BYU-FHTW 2020





The State of “Handwriting Recognition”

A.K.A “Automated Indexing”

80-80 Research Goal

Be able to automatically extract at least 80% of genealogical fields with 80% or greater word accuracy across our entire corpus (4B+ images)

Today



UK, City and County Directories, 1766 - 1946 for George Embley

Save ▾

DIRECTORY.]

Embley Mrs. Catherine 21 Blisworth st. Litherland
—— Geo. Huxley dispatch clerk 124 Salisbury rd 15
—— Jn. plumber 20 Briarwood rd 17
—— Wltr. Fred Courtney 231 Muirhead av 13
Embura Mark foreman 10 King's mount Oxton, B'head
Embury Mrs. 27 Mallaby st. Birkenhead
—— Rt. municipal clerk 22 Mount rd. Higher Tranmere
Emerson Ernest, J.P. 39 Ravenswood av. Rock Ferry.
T N Rock Ferry 581
—— Misses Florence & Ettie E. 30 Victoria drive West
Kirby, Wirral, Cheshire
—— Fras. Josph. cabnt. mkr. 12 Gloucester rd. Anfld.E
—— Geo. 16 Fairmead rd 11
—— Geo. 9 Smith drive 20
—— Leonard Watkins 42 Dencourt rd 11
—— Miss Margt. 5 Baytree rd. Higher Tranmere
—— Stephen F. electrl. engnr. 9 Grange rd. West
Kirby, Wirral, Cheshire. Hoylake 789
—— Wm. W., M.B., Ch.B., B.A.O. Belf. physcn. and
surgn. 83 Merton rd. Bootle. T N Bootle 969
Emery Alfd. butcher 5 Hanbury rd
—— Alfd. commrc. travllr. 47 Allangate rd 19
—— Alfd. who. warehsmn. Russell bldgs. School la 1.

ALPHABETICAL

Emmett Harold Jas. Farquhar
dard Yeast Co. Ltd.) 1
Swan 1187
—— Jas. postman 15 Merton dr
—— Mrs. Jessie 58 Webb
—— Jn. 67 Barlow la 4
—— Jn. 9 Railton rd 11
—— Jn. bacon curer 12 Cliff rd
—— Jn. boat bldr. 6 Orrell rd.
—— Jn. pork butcher 93 Wa
61 County rd 4; 98 V
Granby st 8—res. 5 Stra
—— Jn. & Sons pork butchers
Broadway, West Derby
and cooked meat purve
328); and 225 Kensington
—— Jn. W. 20 Menlove gdns. 1
—— Leonard Leslie electrl. e
Walton
—— Mrs. Louisa 42 Royden st.
—— Luther 58 Arkles la. Anfie
—— Mrs. Margt. 7 Dane st. V
—— Mrs. Norah, Westlake, 8



1257 of 2399



This image courtesy of [GenealogyBank](#)

esday at 9
Contribu
ne for
ily ap-
NDAY
4 AND

announced. Interment R
Cemetery. Friends ma
**BERRYS FUNERAL HOME, 7200
DETROIT AVE., 3-5 AND 7-9 P.M.
MONDAY.**

RAGUSA

See Monastra notice.

RHOME

James A. Rhome, age 63, of Pa-
taskala, O., beloved husband of
Evelyn (Dixon), dear father of
Peggy Corado of Seattle, Wash.,
and James Allen of Napoleon, O.,
stepbrother of Glenn Bender of
Berea, Calif., and Harold Rhome of
Dana Point, Calif. Funeral ser-
vices Monday, Nov. 12, at 1:30
p.m. Family will receive friends
at **CORRIGAN'S FUNERAL
HOME, LORAIN AVE. AT W.
208TH ST., FAIRVIEW PARK,
SUNDAY 3-5 AND 7-9 P.M.** Mem-
ber of Gimsied Falls Masonic
Lodge No. 705.

. 12, St. Peter
at 11 a.m. In-
terment Holy Cross Cemetery.
Friends may call at **CHAMBERS
FUNERAL HOME, 4420 ROCKY
RIVER DR. AT PURITAS. SUN-
DAY 3 TO 5 AND 7 TO 9 P.M.**

STEVENHAGEN

Ted C. Stevenhagen, beloved hus-
band of Dorothy (nee Cathorn),
brother of Catherine Otto of Del-
tona, Fla., two nieces and two
nephews. Nov. 7. Contributions to
the memorial fund of Lake Shore
Christian Church would be appre-
ciated. Memorial services Sunday,
Nov. 11 at The Lake Shore Chris-
tian Church (2800 Lake Shore
Blvd., Euclid, O.) at 3 p.m.

STOA

Anne M. Stoa (nee Venchiaru-
fi) age 61, wife of Vernon A.,
mother of Karen (Mrs. William)
Cordes of Elkhart, Ind., and Eric
Jon of South Russell, grandmoth-
er of Billy, Kathy and Laurel,
sister of Phyllis (Mrs Willis)
Moore of Cleveland. Dorothy

and of Barbara, father of John
Gordon, brother of Luther R. and
Mrs. Jimmie L. Smith. Family re-
ceiving friends at **STROWDER'S
FUNERAL CHAPEL, 818 EAST
105TH ST., MONDAY 8:30-9:30
P.M.** Services and burial Wednes-
day in Malvern, O. **THE MAT-
THEWS FUNERAL HOME OF
CANTON, O., IN CHARGE.**

WOLOSZ

Frank Wolosz, beloved husband of
Mary Ann (nee Sitarz), dearest
brother of Edward Wallace and the
late Charlotte Mantel, dearest un-
cle. Funeral Monday, Nov. 12, at
9:30 a.m. from **MOSINSKI FUNER-
AL HOME, 3675 E. 65TH ST., ser-
vices 10 a.m., at Our Lady of
Lourdes Church. Interment Calvary
Cemetery. VISITING HOURS SUN-
DAY 2-5 AND 7-9 P.M.**

YARMUS

Sophye R. Yarmus, dearly beloved
wife of Joseph, devoted mother of
Jane C. Dominguez, James C.,
dearest sister of Edith Weinber.

Image Index

Information

Name	Event Type	Event Date	Event Place	Gender	Age	Relationship to D...	Birth Date	Birth
Ted C Stevenhagen	Obituary	11 Nov 1973	Cleveland, Cuyahoga, Ohio, United States	Male				
Edward J Hovorka Sr	Obituary	11 Nov 1973	Cleveland, Cuyahoga, Ohio, United States	Male				

Lebenspally, als aber je
fiedel mit seiner
...

Wundergalt. *Lebenspally* ...
die Lobs ...
galtb. ...
...
...
...

Mit den ...
...
...
...
...
...
...

...
...
...

Not Yet

परदेशी
 जात गौली वासी । हींगन घाट जैन मंदि २
 देवल के पास प्राग लीर्था साठी आले . सारदा
 गौली आजा कालकाप्र साठजी . चिंपुव दुर्गा प्र
 स्त्री के सो गंगा में शंभे . और काम धामकि
 भादौ वदी १ संवत् ! २०१३ का

जात राउत वासी जोहरा तं कठ घोरा जिगवे लासपुर
 प्रा० महाजन बेटा चन्दाके फू० पिता के लीये सोम

कुरमी वासी नवागांन के नवलपुर के पास
 प्रा० रूपई बेटा दुकालर के पीता वरन के
 मुठ व बीतन व रात चन्द फूल व डीदा

जात काही वासी कटनी खासठि० कटनी जि० जबलपुर तं
 राम अधीन के अमीर रामबेलाबन सुरबनहन आये

मरव सदाशिव राव कौलीक राव पाटील मु० भालरवे डानुद पा० परचोडाज.
 जलगाव पूर्व खानदेश सा० पा चौरा कौलिकराव रामचन्द्र अस्ती
 पिता प्रकाश राव कौलिकराव पाटील याचे अस्ति विधि करायें ता० २३/६/५६

[A vertical strip of handwritten text, heavily scribbled and mostly illegible.]

[A vertical strip of handwritten text, mostly illegible.]

[A horizontal strip of handwritten text at the bottom left, mostly illegible.]

[A horizontal strip of handwritten text at the bottom right, mostly illegible.]

(Surprisingly) Not Yet

PERSON ID	LITERACY: ANSWER FOR PERSONS 11 YEARS AND OLDER		FULL TIME EDUCATION: ANSWER FOR PERSONS 3 YEARS AND OLDER		
	P11		P12a	P12b	P12c
	In what language can [NAME] read and write?		Has [NAME] ever attended school or is [NAME] attending school now?	What is the highest level of schooling [NAME] is attending now/attended in the past?	HIGHEST EDUCATIONAL GRADE COMPLETED FOR THOSE CURRENTLY ATTENDING SCHOOL OR ATTENDED IN THE PAST
	1 None (not literate) 2 English only 3 Ghanaian language only 4 English and Ghanaian language 5 English and French 6 English, French and Ghanaian language 7 Other (Specify)		IF NEVER, GO TO P13a	01 Nursery 02 Kindergarten 03 Primary 04 JSS/JHS 05 Middle 06 SSS/SHS 07 Secondary 08 Voc/technical/commercial 09 Post middle/secondary certificate (teacher training/college of education, agric, nursing, etc) 10 Post secondary Diploma (University diploma, HND, teacher training/college of education diploma, etc) 11 Bachelor degree 12 Post graduate (Cert., Diploma, Masters, PHD, etc)	What is the highest grade (form/class/level) [NAME] has completed at that level of schooling? IF 0 - 8, RECORD ACTUAL GRADE IF 9 OR MORE, CODE 9
	CODE	OTHER (SPECIFY)	N E V E R	N O W	P A S T
01	4		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
02	1		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
03	4		<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
04			<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
05			<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
06			<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
07			<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
08			<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
09			<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10			<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Infinity Project

- Volunteers are more likely to index records containing their own family
- “Just in time” quality vs. up-front perfection
- Removing other bottlenecks

[The Year 1812]

Page 253

N^o 527 } James Morrey of this Parish Cordwainer
and Mary Topham of this parish Spinster were
Married in this Church by Banns
this fourth Day of September in the Year One Thousand eight Hundred
and twelve By me John Latham Curate
This Marriage was solemnized between Us { James Morrey
Mary Topham
In the Presence of { Charlotte Robinson
Henry Morrey

Event Type: Marriage

Event Date: 4 September 1812

Event Place: St. Mary's, Nantwich, Cheshire, England, UK

Bride: Mary Topham

Groom: James Morrey

Groom's Occupation: Cordwainer

Bride's Occupation: Spinster

Marriage Officiator: John Latham

Marriage Officiator Title: Curate

Marriage Validation: Banns

Witness 1: Charlotte Robinson

Witness 2: Henry Morrey

Page Number: 253

Record Number: 527

So we're done, right???

End goals of FH records

- A family tree containing one—and only one—record of every person who has lived on the Earth
- Tell a coherent (and hopefully entertaining) story
- Render value to all people worldwide

How might we structure the data differently to achieve these goals?

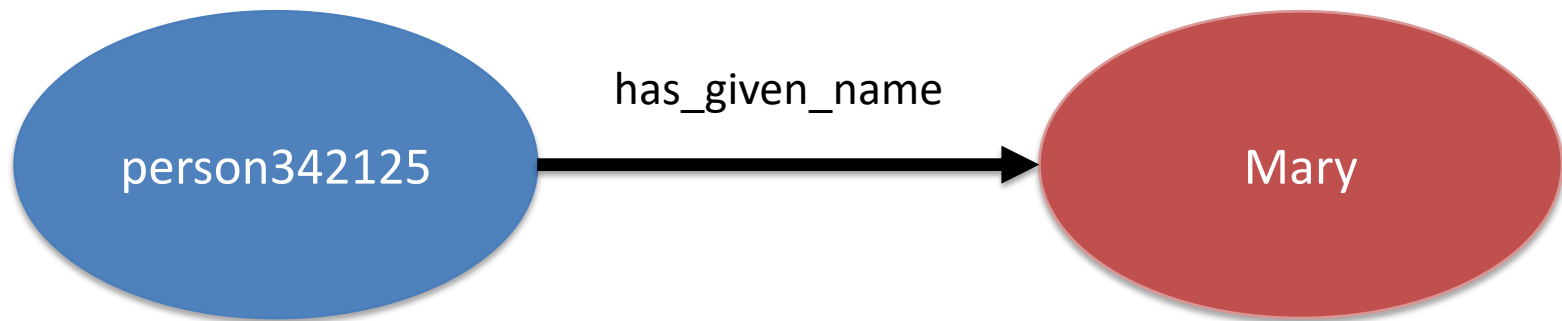
- Break open the “silos”
- Take advantage of *everything* we know
- Make *everything* queryable



Knowledge Graphs

(Rebooted)

Triple Stores



Subject

Predicate

Object

Entity1

Relationship

Entity2



- Launched 2012
- Backs Wikipedia and sister projects
- Funded by the Allen Institute for AI, Gordon & Betty Moore Foundation, and Google
- Currently contains almost 80M entities

Ancient intellectual network

This dataviz maps the relationships between master and student from Socrates to the end of the hellenistic period using data I'm currently adding to [Wikidata](#). As of today, this work is not completed (and, consequently, this graph is merely an early beta)

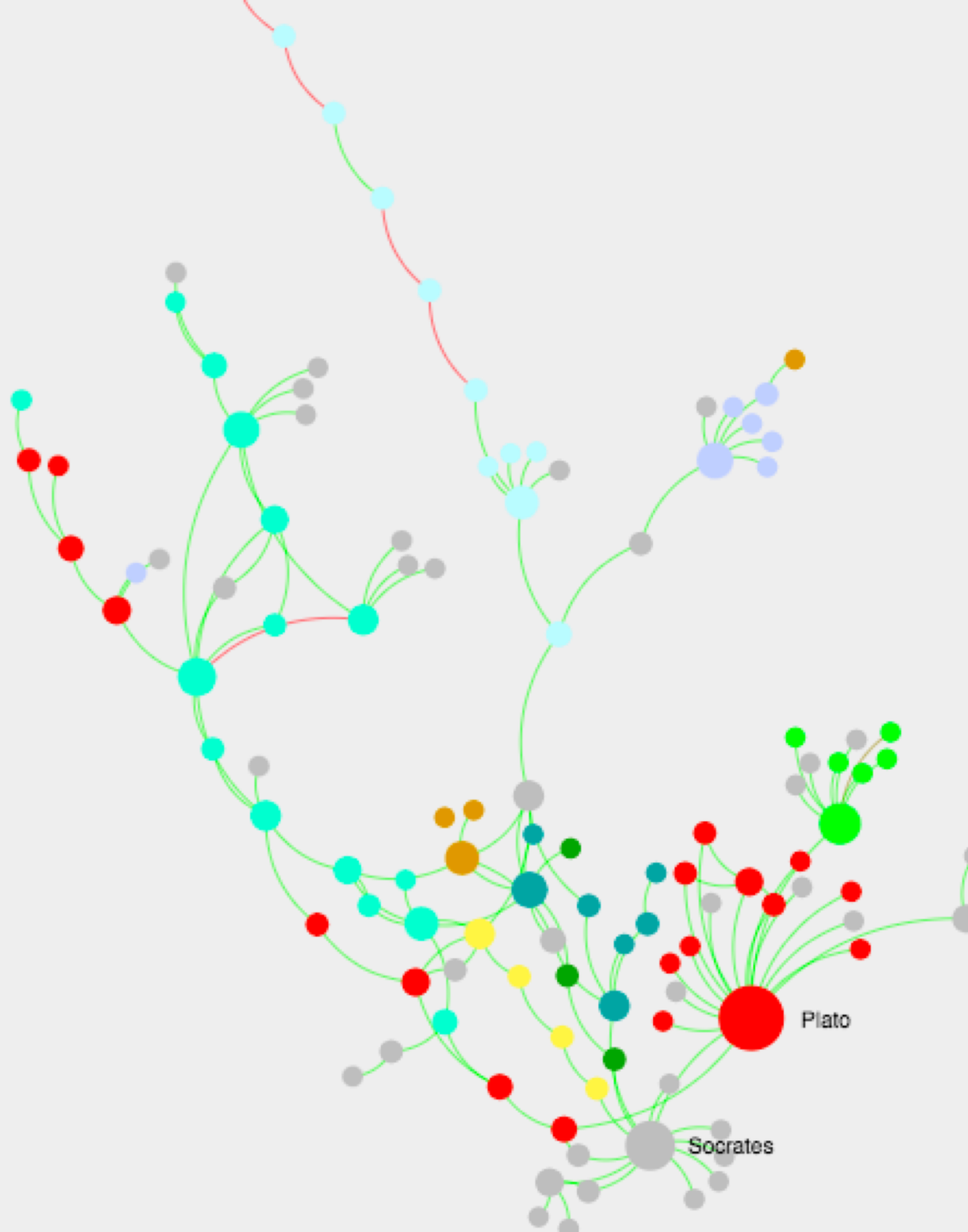
The philosophical communities are represented with the following color chart:

- Platonism
- Peripatetician (Aristotelism)
- Megarian school
- Cyrenaic school
- Epicureanism
- Stoicism
- Cynic school
- Eretrian school
- Pyrrhonism
- No known affiliation

Since the original data are quite fragmentary, the student-master links between each philosopher have been colored according to their accuracy :

- Well-established
- Likely
- Hypothetical

i [More about this visualisation](#)



Ancestors of George Washington



```
1 #George Washington and his ancestors...
2 SELECT DISTINCT ?item ?itemLabel ?dateofbirth ?image
3 WHERE
4 {
5   wd:Q23 (wdt:P22|wdt:P25)* ?item .
6   OPTIONAL { ?item wdt:P569 ?dateofbirth } .
7   OPTIONAL { ?item wdt:P18 ?image. }
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 } ORDER BY DESC(?dateofbirth)
```

4643 results in 3669 ms

</> Code

Download

Link

Search

item	itemLabel	dateofbirth	image
wd:Q23	George Washington	22 February 1732	commons:Gilbert Stuart Williamstown Portrait of George Washington.jpg
wd:Q458119	Mary Ball Washington	30 November 1708	commons:Mary Ball Washington(Pine).jpg
wd:Q768342	Augustine Washington	12 November 1694	commons:Augustine Washington.jpg
wd:Q3314035	Mildred Gale	1 January 1671	
wd:Q3219928	Lawrence Washington	1 September 1659	commons:Portrait of Lawrence Washington B&W.jpg
wd:Q16738459	Joseph Ball	1 January 1649	
wd:Q76112240	Baldwin de Redvers, 1st Earl of Devon	11 May 1645	

Ancestors of George Washington

```
1 #George Washington and his ancestors...
2 SELECT DISTINCT ?item ?gender ?itemLabel
3 WHERE
4 {
5   wd:Q23 (wdt:P22|wdt:P25)* ?item .
6   ?item wdt:P19 ?placeofbirth .
7   ?item wdt:P21 ?gender .
8   ?placeofbirth wdt:P625 ?geoloc
9   SERVICE wikibase:label { bd:servicePa
10 }
```

Submit 717 results found < ^ >

Data options

Coordinate	Color
<input type="text" value="geoloc"/>	<input type="text" value="gender"/>
Radius	Label
<input type="text" value="none"/>	<input type="text" value="none"/>
Link from coordinate	Link to coordinate
<input type="text" value="none"/>	<input type="text" value="none"/>

Display options

Canvas options

Table

Basic

Trees

Maps

More



Ancestors of George Washington



Wikidata Query Service

Examples

Help

More tools

English

```
1 #defaultView:ImageGrid
2 SELECT DISTINCT ?item ?itemLabel ?dateofbirth ?image
3 WHERE
4 {
5   wd:Q23 (wdt:P22|wdt:P25)* ?item .
6   OPTIONAL { ?item wdt:P569 ?dateofbirth } .
7   OPTIONAL { ?item wdt:P18 ?image. }
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 } ORDER BY DESC(?dateofbirth)
```



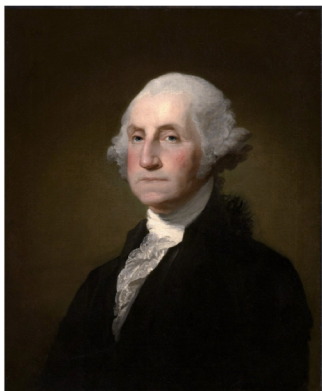
Image grid

4643 results in 4996 ms

Code

Download

Link



commons:Gilbert Stuart Williamstown ...

22 February 1732

George Washington



commons:Mary Ball Washington(Pine).jpg

30 November 1708

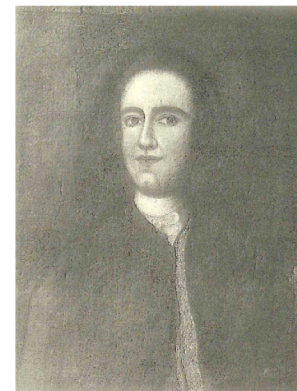
Mary Ball Washington



commons:Augustine Washington.jpg

12 November 1694

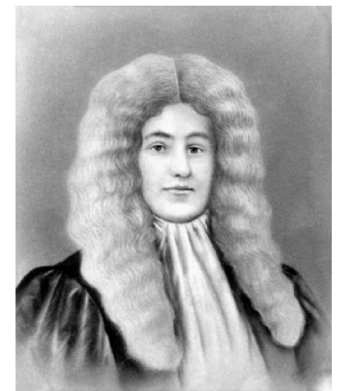
Augustine Washington



commons:Portrait of Lawrence Wa...

1 September 1659

Lawrence Washington



commons:Augustine Warner Jr.gif

1 January 1642

Augustine Warner, Jr.



George Washington

97%



Mary Ball

Descendants of Genghis Khan

```
1 # Descendants of Genghis Khan
2
3 SELECT
4 ?personItem ?englishName ?chineseName ?
5 ?childItem ?childEnglishName ?childChir
6 ?CBDB ?depth
7 WHERE
8 {
9 SERVICE gas:service {
10 gas:program gas:gasClass "com.bigda
11 gas:in wd:Q720 ;
12 gas:traversalDirection
13 gas:out ?personItem ;
14 gas:out1 ?depth ;
15 gas:maxIterations 5 ;
16 gas:linkType wdt:P40 .
17 }
```

Submit 434 results found < ^ >

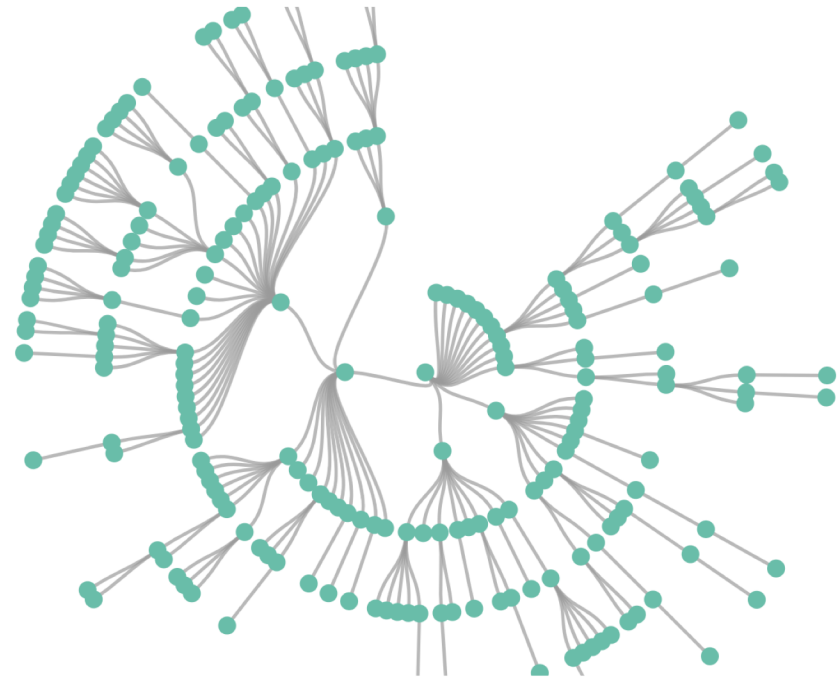
Data options

Link from	Link to
<input type="text" value="personItem"/>	<input type="text" value="childItem"/>
Label	Color
<input type="text" value="none"/>	<input type="text" value="none"/>

Display options

Canvas options

Table Basic Trees Maps More



Refresh Download Share

Wikidata Query Service

Wikidata Query Service

Examples

Help

More tools

English

```
1 SELECT ?item ?itemLabel ?fsID
2 WHERE
3 {
4   ?item wdt:P31 wd:Q5 .
5   ?item wdt:P2889 ?fsID .
6   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
7 }
```

2732 results in 1487 ms

Code

Download

Link

Search

item	itemLabel	fsID
Q26718317	Richard Alsager	LZGX-KLN
Q26928891	Feliks Szczyński	L81B-DBS
Q26972887	James Paul Boland	LT85-X5Q
Q30349633	Valentine Charles Outwater	LD2M-XN6
Q33026914	Michael Joseph Mulvihill	LTF8-NPC
Q33082977	Josephine Corliss Preston	K8F5-CMB
Q33103568	Margaret Elizabeth Tout	KWZW-KKP
Q33110665	Virgil Lee Highland	L4MJ-BQK
Q34119528	Donald George Wilhelm, Jr.	LTVG-Z4B
Q36423395	Lars Öknaberg	LDMZ-BC7

Descendants of John Lathrop who played sports

Wikidata Query Service [Examples](#) [Help](#) [More tools](#) English

```
1 #Descendants of John Lathrop who played sports
2 SELECT DISTINCT ?item ?itemLabel ?dateofbirth ?teamLabel
3 WHERE
4 {
5   wd:Q6245304 (wdt:P40)* ?item .
6   OPTIONAL { ?item wdt:P569 ?dateofbirth } .
7   ?item wdt:P54 ?team .
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
```

2 results in 482 ms [Code](#) [Download](#) [Link](#)

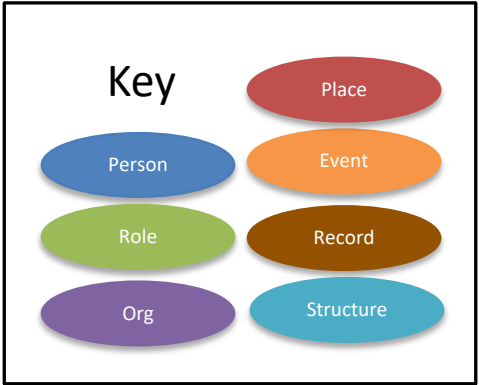
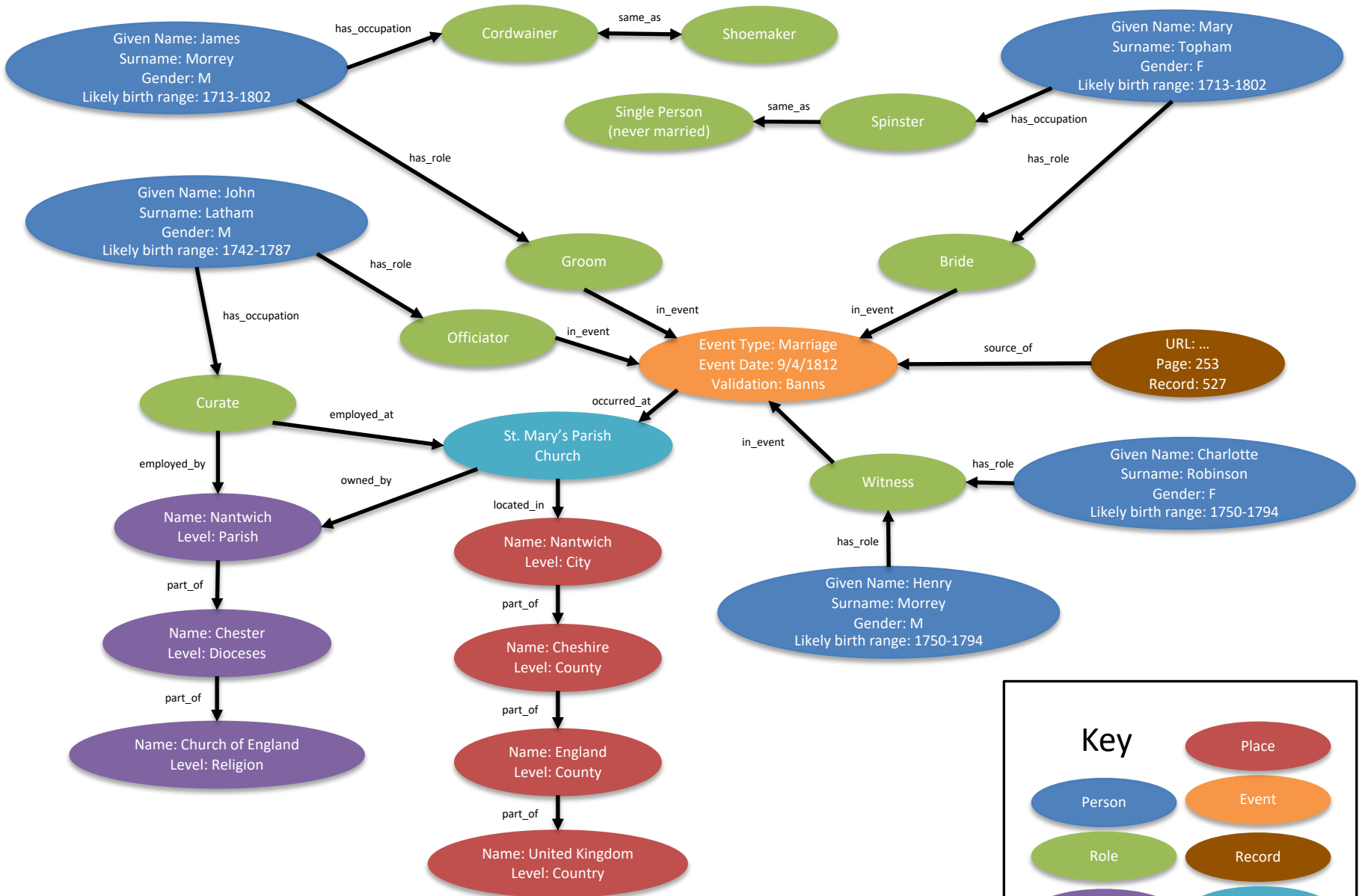
item	itemLabel	dateofbirth	teamLabel
wd:Q324742	Prescott Bush	15 May 1895	Yale Bulldogs football
wd:Q23505	George H. W. Bush	12 June 1924	Yale Bulldogs

Which facts are “genealogical”?

- Name, gender, birth date, relationships, etc.
- Occupations?
- Religion?
- Business associates?
- Whereabouts on a given day?
- Sports team?

I want to know...

- Who might have taught James Morrey the shoemaker trade?
- Were the witnesses related? How?
- Who *might* have been Henry Morrey's wife (given age, proximity, religion, social status, etc.)?
- How many *different* James Morreys are there within a 50 mile radius?
- Does this possible duplicate have an “alibi”?

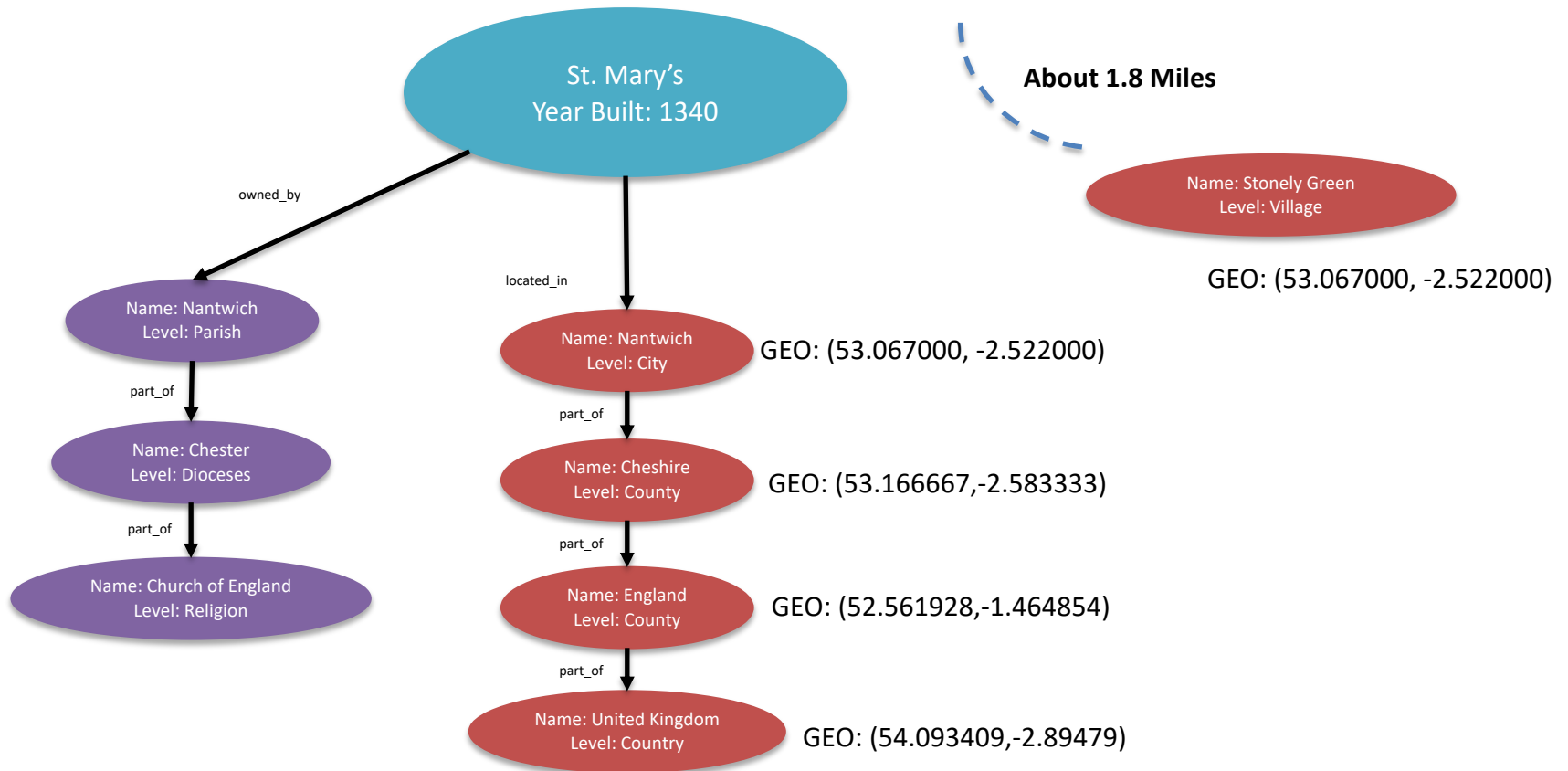


Other World Knowledge

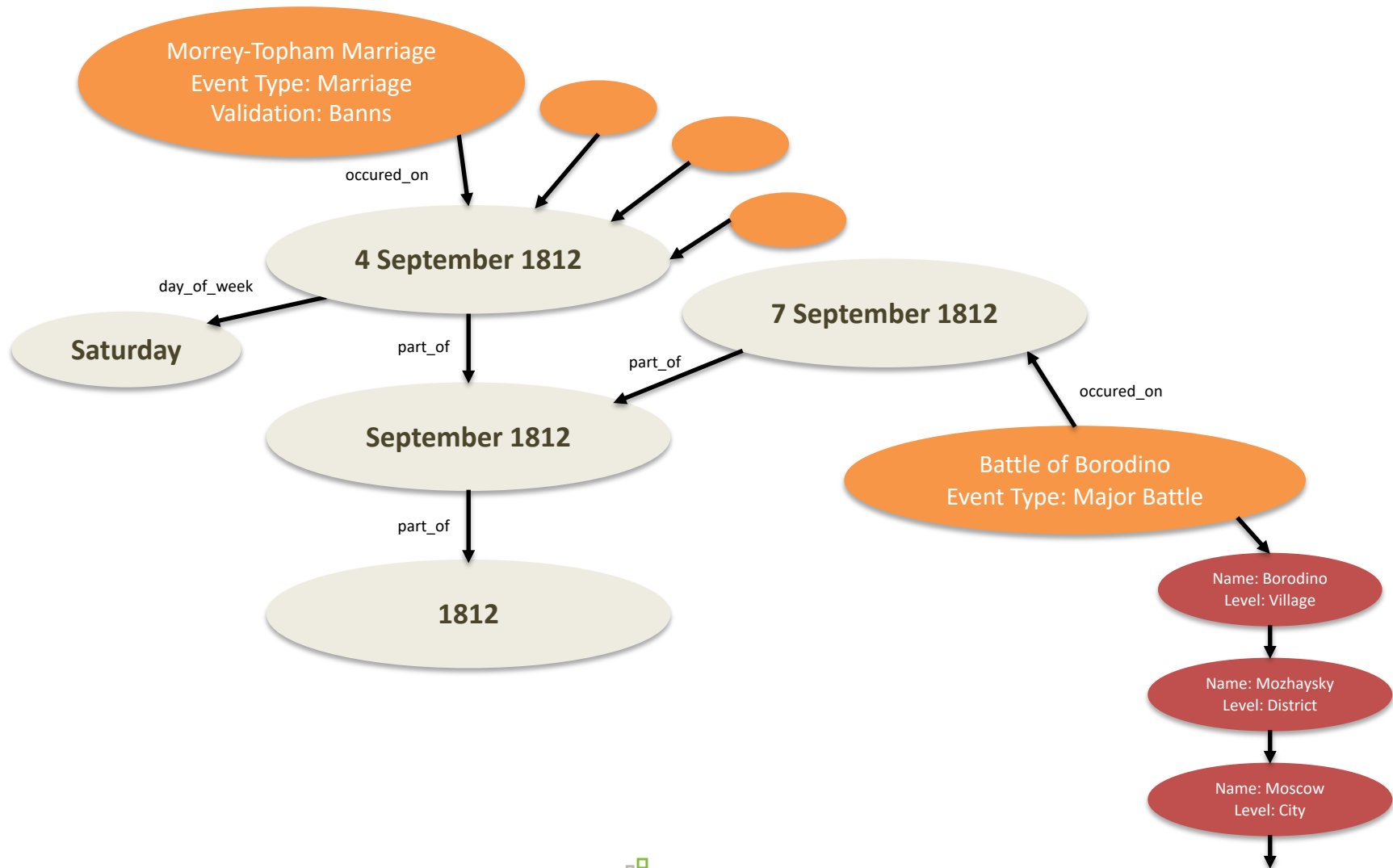
GEOPOLY: [(53.0672,-2.5210), (53.0672,-2.5202), (53.0668, -2.5202), ..]

GEO: (53.067112, -2.520631)

About 1.8 Miles



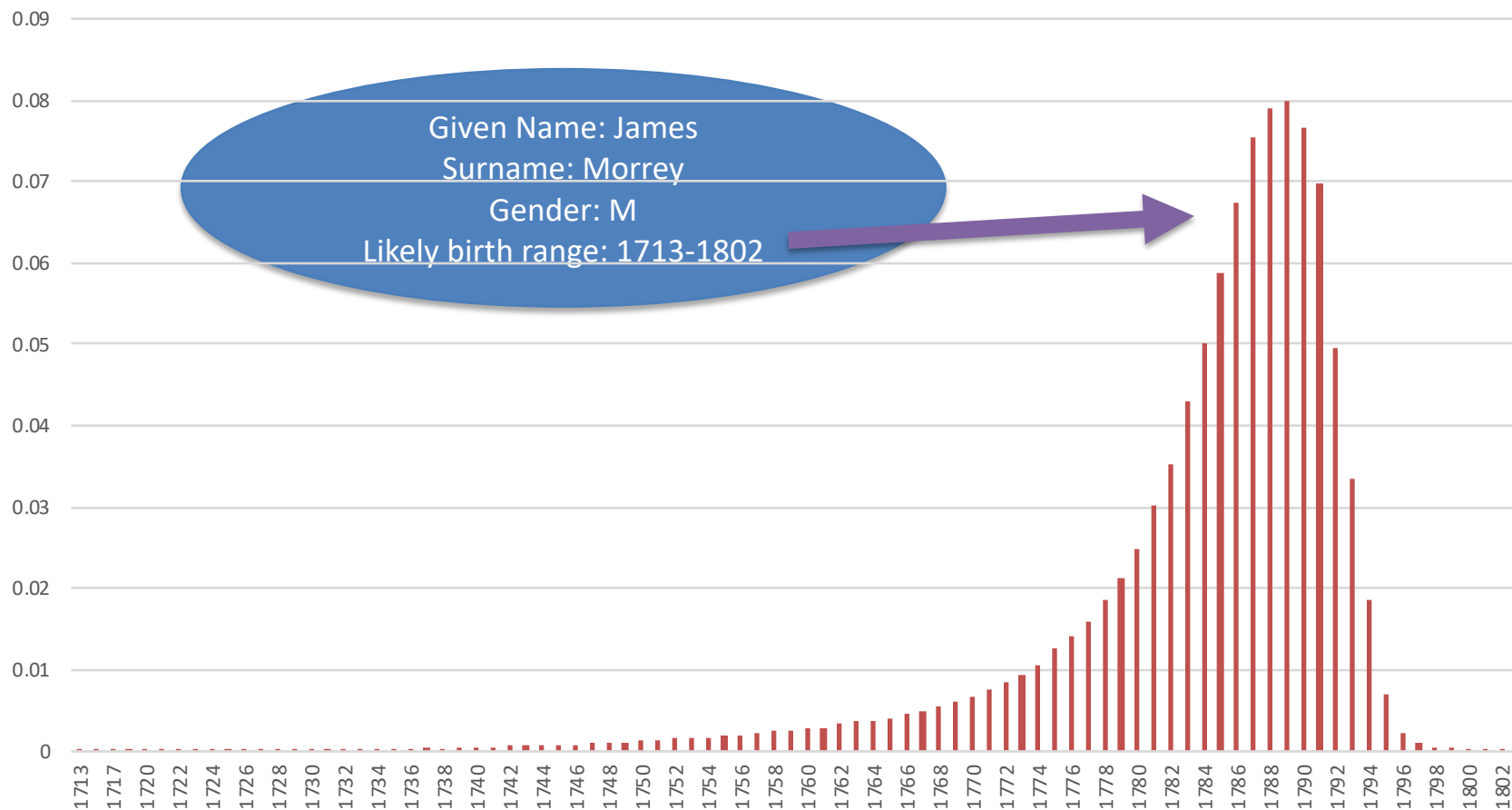
Other World Knowledge (Cont.)



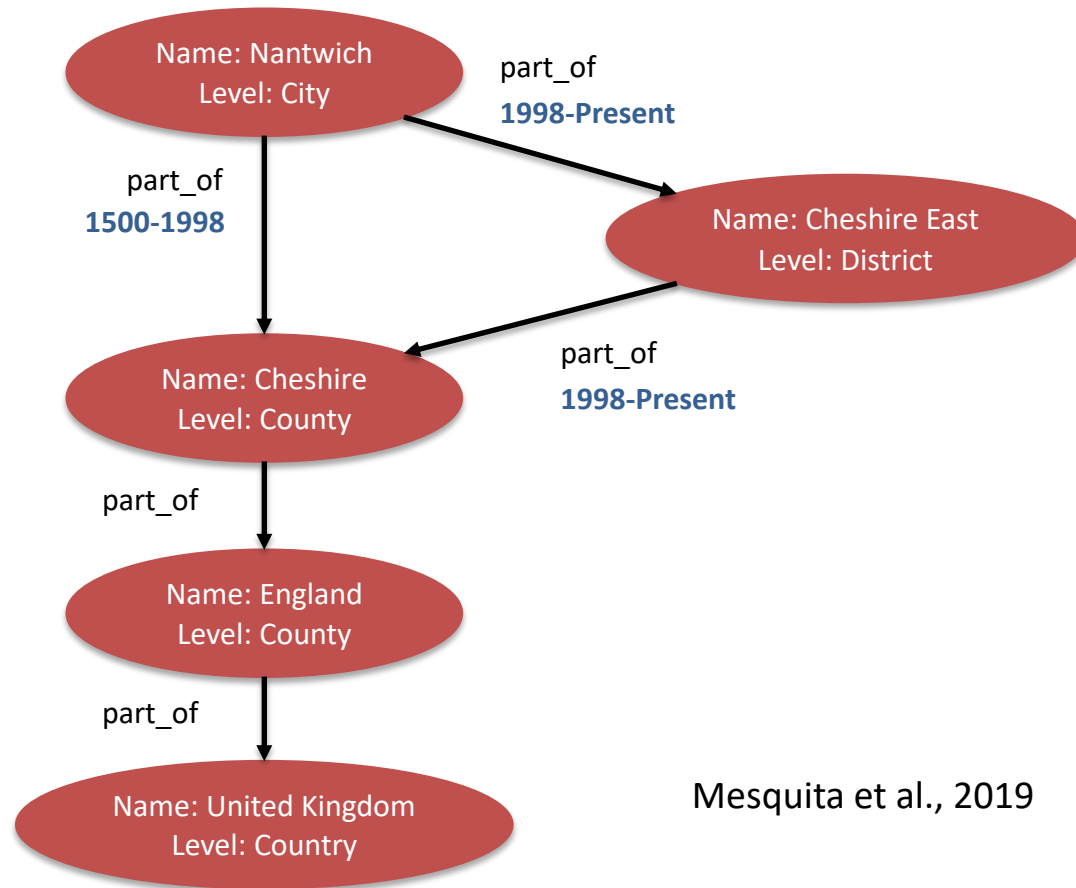
“Facts” as Probabilities

Given Name: Charlotte
Surname: Robinson
Gender: F (0.94)

Ranges as Distributions



Temporal Qualifiers



Mesquita et al., 2019

Other Digital Attributes

- Autographs
- Thumbprints
- DNA data
- Photos

Sources of knowledge

- Historical Records
- Family Tree
- Standards/Authorities
- Catalogs
- Books
- Wikipedia and Help Wikis
- Family History Guide
- Public genealogy sites
- Other knowledge graphs
- ...

How Big?

- ~20 entities per image * 4B images ~= 80B entities
- 2-3 times as many triples ~= 160B
- Not just records—Family Tree, etc... 1T triples?

Collapsing the Graph

- The knowledge graph is typically not the “system of record”
- Knowledge Fusion
- Regular rebuilds

Is this even possible?

AI

Diffbot launches AI-powered knowledge graph of 1 trillion facts about people, places, and things

KYLE WIGGERS @KYLE_L_WIGGERS AUGUST 30, 2018 8:00 AM



Image Credit: Diffbot


If you've ever performed a Google search for a celebrity, famous landmark, or

VB TRANSFORM

The AI event for
business leaders

San Francisco
July 15 - 16

[Learn More](#)

 Creative Cloud for teams

**Our experts are
on your team.**

Get business features like 1:1

Diffbot

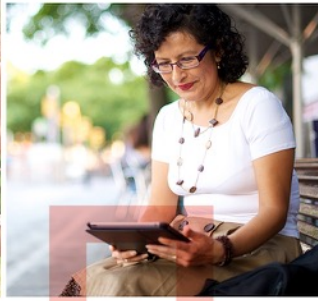
- Fully automated
- 1T triples (10B entities)
- Rebuilt from scratch every 4 days

...but good ontologies are *hard*.

- Some of the ontology already exists (Wikidata)
- Deal with it! It'll be worth it 😊

Can AI “learn on its own” from a Knowledge Graph?

- Short answer: not really (yet)
- We’ll never know if we don’t have one!



Questions/Comments?



Resources: Wikidata Query Builder

<https://query.wikidata.org/>



Resources: Diffbot TWIML&AI talk

<https://twimlai.com/twiml-talk-319>



References

- Filipe Mesquita, Matteo Cannavicchio, Jordan Schmidek, Paramita Mirza, Denilson Barbosa. 2019. KnowledgeNet: A Benchmark Dataset for Knowledge Base Population. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pages 749–758, Hong Kong, China, November 3–7, 2019*. Specifically, see “Beyond Binary Relationships” in appendix.