



Автоматическая идентификация репрессированных семей

Илья Воронцов, Глеб Суворов, Москва, 18.06.2017

Задача

хан, 1873 г. р. Житель а. Аблай Черлакского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Райса, 45 л., дети: Байзиит, 9 л., Канын, 7 л., Кулбану, 5 л., Лукман, 2 г. Выслан вместе с семьей в 1931 г. Реабилитирован УВД по Омской обл. (ГАОО. Ф. 1007. Оп. 2. Д. 71 ИЦ УВД по Омской обл. Ф. 8. Оп. 19. Д. 455)

1878 г. р. Житель с. Николаевка Николаевского с/с Исилькульского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Евгения, 31 л., дочь Таисия, 19 л., сын Леонид, 19 л., дочь Евгения, 1919 г. р., сын Юрий, 1929 г. р., сын Александр, 3 г. Расстрелян в 1931 г. Реабилитирована УВД по Омской обл. АВДЕЕВА Пелагея Николаевна, 19. 03. 2012 г. (ГАОО. Ф. 1011. Оп. 2. Д. 1, 4, 54, ИЦ УВД по Омской обл.)

онон Иванович, 1877 г. р. Житель Ивановского с/с Черлакского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Тасия, 1912 г. р., сын Василий, 1920 г. р., сын Михаил, 1928 г. р. Выслан вместе с семьей в 1930 г. на спецпоселение в Восточный Западно-Сибирского края. Реабилитированы УВД по Омской обл. АВДИЕНКО (ОВДИЕНКО): Конон Иванович, Анна Ефимовна 03. 12. 1888 г. (ГАОО. Ф. 1007. Оп. 2. Д. 3, 9, 291 Ф. 1120. Оп. 1. Д. 80. Т. 11 ИЦ УВД по Омской обл. Ф. 8. Оп. 50. Д. 720)

ч. Житель с. Нагибино Нагибинского с/с Тюкалинского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена, 31 г. вместе с семьей. Реабилитирован УВД по Омской обл. АВЕРКОВ Орест Павлович 13. 03. 2012 г. (ГАОО. Ф. 1539. Оп. 2. Д. 41)

ович, 1896 г. р. Житель с. Воскресенское Воскресенского с/с Иконниковского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Тасия, 65 л., сестра Вера, 16 л., сын Павел, 11 л., дочь Анна, 8 л., сын Федор, 2 г. Выслан в 1930 г., возвращен. Раскулачен в Омской обл. АВИЛКИН Сергей Михайлович 28. 02. 2012 г. (ГАОО. Ф. 1409. Оп. 2. Д. 71 ИЦ УВД по Омской обл.)

олаевич. Житель с. Еремеевка Еремеевского с/с Полтавского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Елена, 1917 г. р. Выслан вместе с семьей в 1931 г. на спецпоселение в Нарымский край, позднее переселены в Кузбасс. Реабилитирована (ГАОО. Ф. 998. Оп. 2. Д. 9, 671 ИЦ УВД по Омской обл. Ф. 8. Оп. 5. Д. 103)

ич (Минович), 1862 г. р. Житель пос. Новодонский Новодонского с/с Исилькульского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Елизавета, 45 л., сын Иван, 19 л., сын Дмитрий, 1913 г. р., сын Василий, 14 л., сын Семен, 1892 г. р., сноха Татьяна Абрамовна, 1892 г. р., внучка Мария, 1915 г. р., внук Николай, 1918 г. р., внук Михаил, 1920 г. р. Выслан вместе с семьей в 1931 г. на спецпоселение в УВД по Омской обл. АВРАМЕНКО: Семен Иванович, Татьяна Абрамовна, Анастасия Семеновна, Мария Семеновна, Николай Семенович 1918 г. (ГАОО. Ф. 1011. Оп. 2. Д. 2, 18, 37, 56, 73, 74, 76, 781 ИЦ УВД по Омской обл. Ф. 8. Оп. 1. Д. 40)

нович, 1877 г. р. Житель с. Павловка Павловского с/с Калачинского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Елизавета, 45 л., дочь Прасковья, 1911 г. р. Выслан вместе с семьей в 1930 г. Реабилитированы УВД по Омской обл. 07. 06. 1999 г. (ИЦ УВД по Омской обл.)

1884 г. р. Житель с. Колесовка Колесовского с/с Колесовского р-на Омского окр. Западно-Сибирского края. Состав семьи: жена Елизавета, 45 л., дочь Анна, 1918 г. р., сын Александр, 1921 г. р., дочь Аграфена, 3 г., сын Иван, 1928 г. р., брат Николай, 1899 г. р., сестра Елизавета, 1928 г. р. Выслан вместе с семьей в 1931 г. на спецпоселение в Средне-Расибонскую комендатуру Нарымского края.

Почему это важно?

из **2 700 000** записей в базе открытого листа
отмечены только **3600** родственников!
притом, что мы знаем, что в реальности их было
НАМНОГО больше



Какие базы использовались?



Какие подходы?



фото предоставлено организаторами

Инструменты

ruby + levenshtein.gem

Python + pyMorphy2

R+httr,XML,data.table



Что мы успели сделать за хакатон?

VorontsovIE committed an hour ago

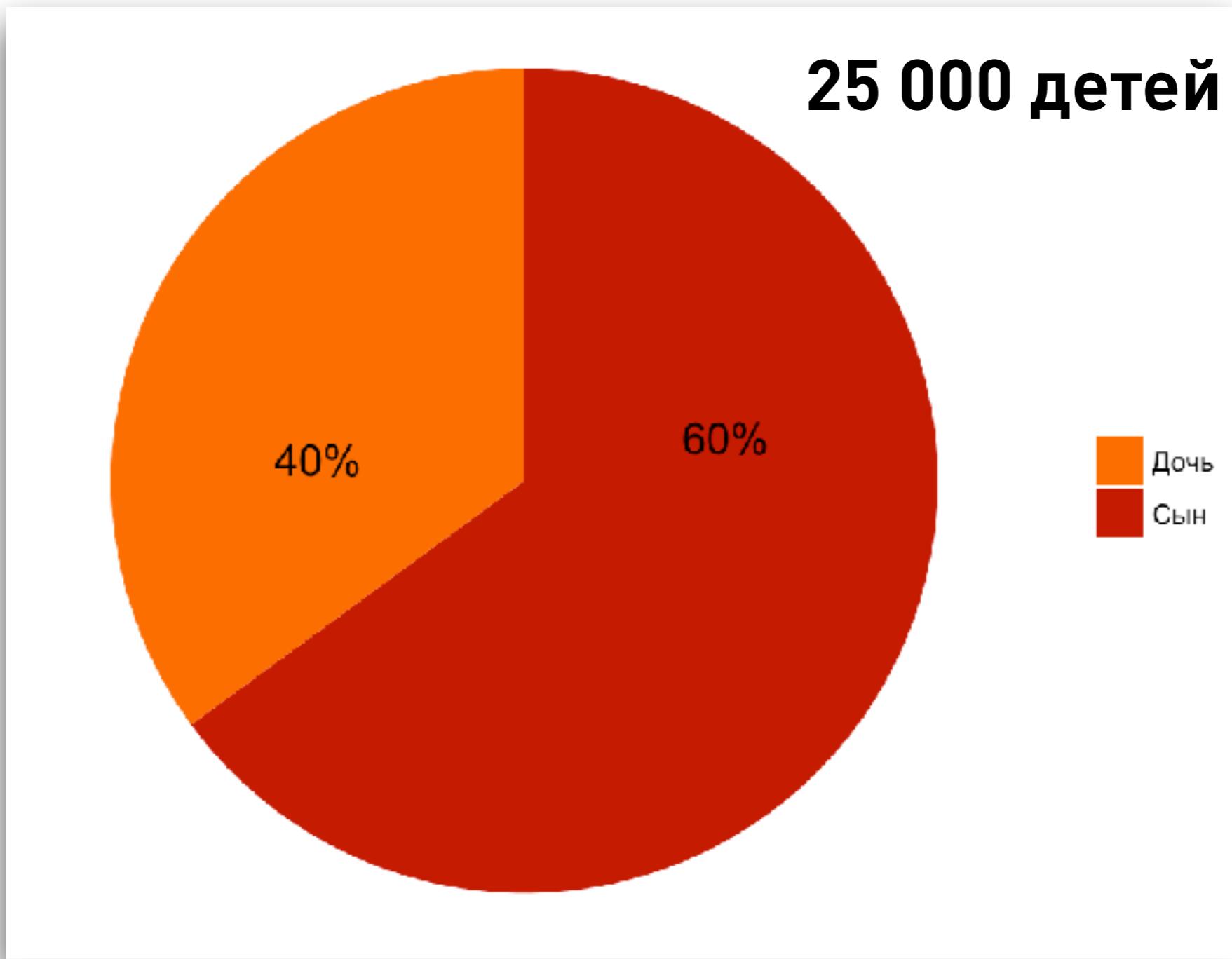
-  **normalize surnames in 'algir' list** 7727c73
-  **readme upd** 56a443d
-  **add additional information; fix bugs in search; separate lib from script** 4e643d5
-  **surnames normalized** 5ef58be
-  **requirements** 66335c1

Commits on Jun 17, 2017

-  **parse memolist** 64cc20a
-  **added families** 2d734b5
-  **added femails from algir** 1085b99



https://github.com/VorontsovIE/memo_family/commits/master



Подозрительные результаты из первой версии

Михайлов Михаил Карпович (1856)	85
Васильев Василий Яковлевич (1853)	81
Захаров Захар (1847)	79
Тимофеев Тимофей Васильевич (1870)	65
Александров Александр Александрович (1869)	61
Павлов Павел Лукьянович (1865)	56
Попов Иван Александрович (1847)	42
Ильин Илья Петрович (1879)	42
Степанов Пётр Фёдорович (1894)	40
Иванов Иван Егорович (1836)	39
Егоров Егор Филиппович (1852)	38
Матвеев Матвей Антонович (1863)	36
Никитин Никита Иванович (1850)	36

Хорошие результаты из 4 версии

FirstName N	SecondName
Абраменко Иван Панфилович (1859) 1	Абраменко Федор Иванович (1882)
Агеев Матвей Васильевич (1870) 1	Агеев Иван Матвеевич (1889)
Андранинов Дмитрий Григорьевич (1880) 1	Андранинов Егор Дмитриевич (1901)
Антипин Николай Павлович (1873) 1	Антипин Иван Николаевич (1902)
Антонов Арсентий Михайлович (1880) 1	Антонов Михаил Арсентьевич (1907)
Белов Мартын Филиппович (1865) 1	Белов Александр Мартынович (1885)
Бердюков Тимофей Михайлович (1877) 1	Бердюков Михаил Тимофеевич (1900)
Беспалов Яков Егорович (1870) 1	Беспалов Петр Яковлевич (1898)
Бирбин Василий Анисимович (1852) 1	Бирбин Петр Васильевич (1892)
Бондарев Алексей Васильевич (1861) 1	Бондарев Александр Алексеевич (1901)
Бочаров Григорий Михайлович (1862) 1	Бочаров Владимир Григорьевич (1887)
Булгаков Степан Максимович (1858) 1	Булгаков Иван Степанович (1896)
Буряков Павел Федорович (1877) 1	Буряков Алексей Павлович (1897)
Варламов Фома Филатович (1873) 1	Варламов Иван Фомич (1902)
Власов Иван Афанасьевич (1863) 1	Власов Михаил Иванович (1894)
Востриков Георгий Григорьевич (1876) 1	Востриков Николай Георгиевич (1910)
Герасимов Дмитрий Харитонович (1875) 1	Герасимов Андрей Дмитриевич (1896)
Герасимов Григорий Иванович (1865) 1	Герасимов Герман Григорьевич (1892)

Ошибки в базе

Гейс Генрих Генрихович (1901) м

Гейс Генрих Генрихович (1901) м

Гейст Фридрих Генрихович (1902) м

Гельвер Александр Георгиевич (1911)

Гельвер Александр Георгиевич (1911)

Гельвер Александр Яковлевич (1911)

1901

1901

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

1902

sex N

NA 25

женщина 1673

мужчина 1988

М 1

Ж 4

женский 3

Гейс Мина Генриховна (1940)

Гейс Фрида Генриховна (1936)

Гейст Генрих Фридрихович (192

Гейст Александр Фридрихович (

Гейст Виктор Фридрихович (193

Гейст Готфрид Фридрихович (19

Гейст Давыд Фридрихович (1935

Гейст Роберт Фридрихович (194

Гейст Готфрид Фридрихович (19

Гейст Александр Фридрихович (

Гейст Виктор Фридрихович (193

Гейст Генрих Фридрихович (192

Гейст Давыд Фридрихович (1935

Гейст Роберт Фридрихович (194

Гейст Александр Фридрихович (

Гейст Андрей Фридрихович (192

Гейст Виктор Фридрихович (193

Гейст Готфрид Фридрихович (19

Гейст Давид Фридрихович (1935

Гейст Роберт Фридрихович (194

(right here) -----^
> items_memo <- fromJSON("~/Downloads/items_pilot2_1.json")
Error in parse_con(txt, bigint_as_char) : parse error: premature EOF

(right here) -----^

Сайт

localhost:4557/?id=4944

Lodz Иван Иосифович (1892)

Варианты имени: //Осипович; Осипович
д. Пустошь Стругокрасненского р-на

Родился: д. Пустошь Стругокрасненского р-на
Постановление: (5.1935) --

Приговор: раскулачен и выслан в Свердловскую обл.

Расстрел: нет
Умер:
Реабилитирован: (22.9.1995)

Lodz Рудольф Иванович (1930)

Родился:

Постоянное место жительства: д. Пустошь Стругокрасненского р-на

Постановление: (5.1935) --

Приговор: раскулачен и выслан в Свердловскую обл.

Расстрел: нет
Умер:
Реабилитирован: (22.9.1995)

[Показать ещё](#)

[Подтвердить](#)

[Это разные семьи](#)

Планы на будущее

- добавить гипотезы на “Открытый список”
- более совершенные модели родства
- Братья/сестры/жены
- подключить сторонние данных
- извлечь данных непосредственно из первичных текстов



Контакты в tlgrm:

Илья @VorontsovIE

Глеб @GlebSuvorov



ТЕПЛИЦА
СОЦИАЛЬНЫХ ТЕХНОЛОГИЙ
te-st.ru