Data Catalog

Office of Chief Data Officer

CIVIC DIGITAL FELLOWSHIP

U.S. Citizenship and Immigration Services

Onintze Contreras
Cornell University
Information Science

What's the Problem?

BBSS

BCS

CIS

CRIS

CFR

CMIS

C3

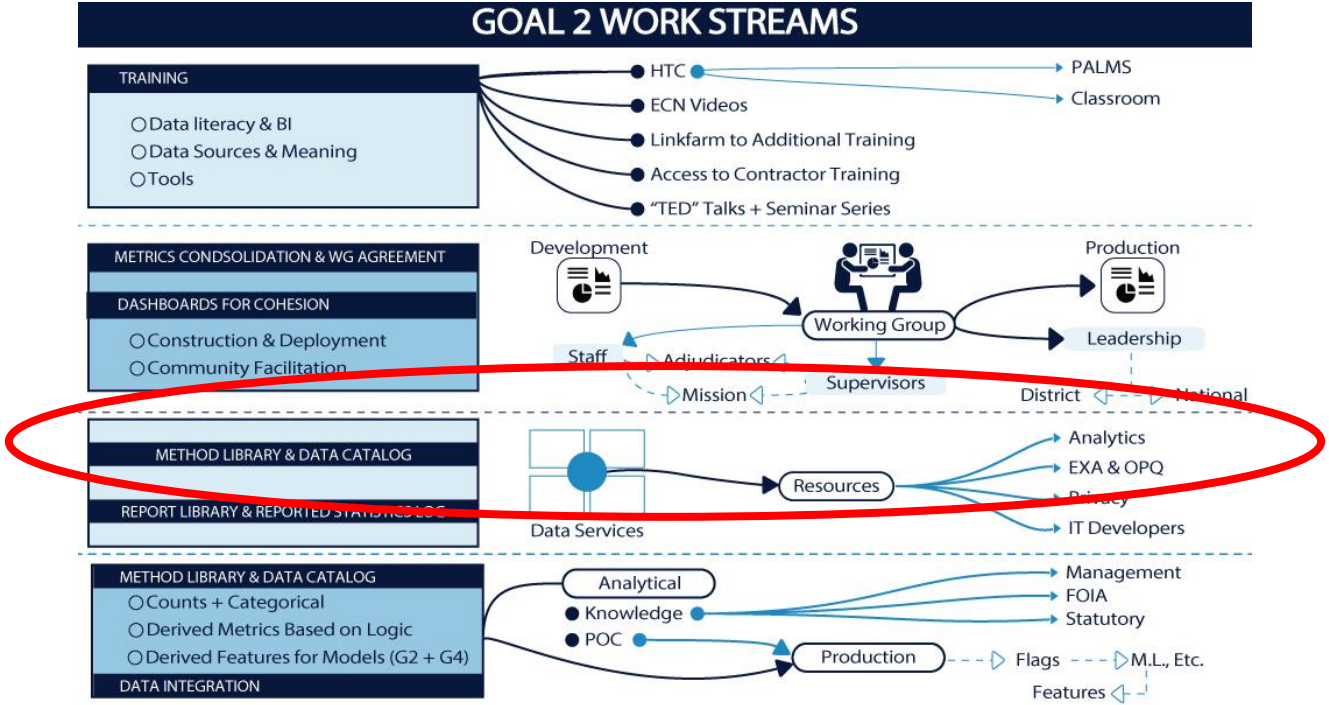
CMS

CIVIC DIGITAL FELLOWSHIP

APRSAFE-AP-CCO	Location code for the asylum office with jurisdiction over the credible fear case	APSS.csv	APRSAFEAPCCO Location code for the asylum office with jurisdiction over the credible fear case
APRSAFE-AP-LAST-NAME	last name.	APSS.csv	APRSAFEAPLASTNAME last name.
APRSAFE-AP-FIRST-NAME	first name.	APSS.csv	APRSAFEAPFIRSTNAME first name.
APRSAFE-AP-MIDDLE-NAME	middle name.	APSS.csv	APRSAFEAPMIDDLENAME middle name.

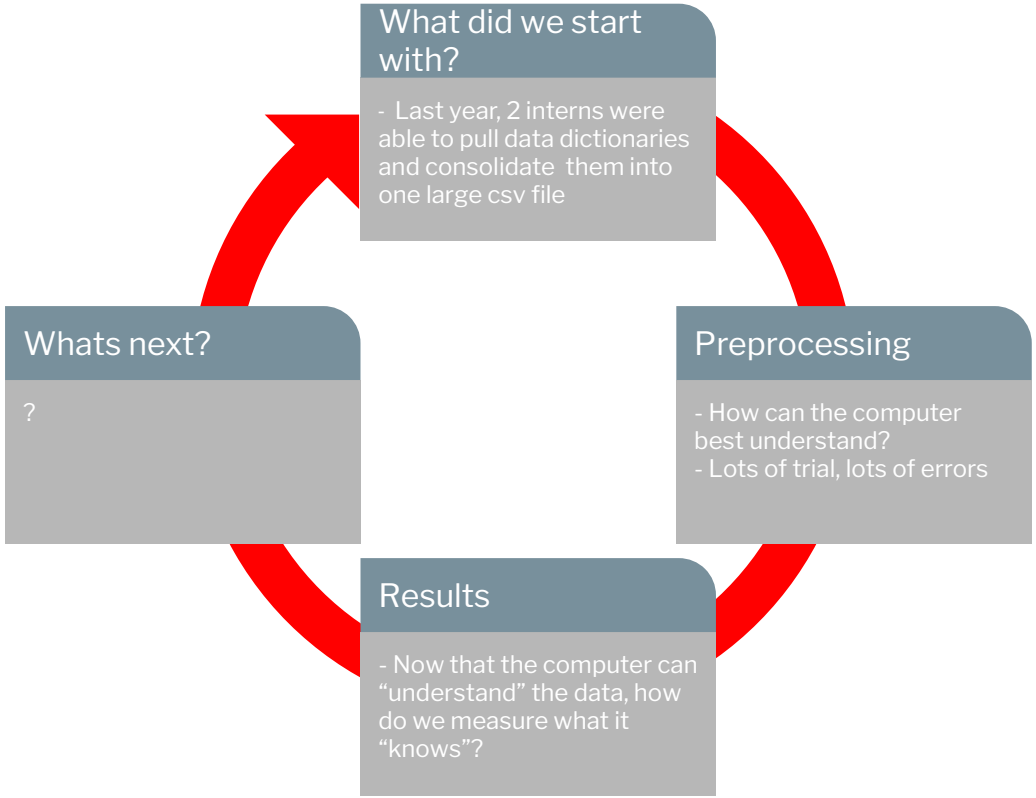
LAST_NAME	Last name (EFTS field 2.907).	BBSS.csv
FIRST_NAME	First name (EFTS field 2.908).	BBSS.csv

What is it? Why?



CIVIC DIGITAL FELLOWSHIP

The Process



CIVIC DIGITAL FELLOWSHIP

What did we Start With?

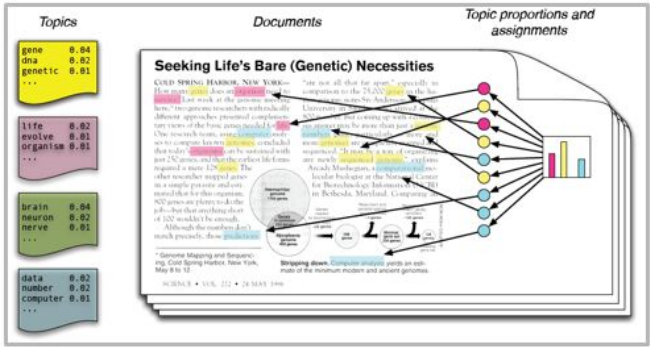
	A	B	C	D
1	Field	Field_Desc	Source	Combined
2	APRSAFE-AP-APPL-A-NUM	Applicant–ís A-number	APSS.csv	APRSAFEAPAPPLANUM Applicants Anumber
3	APRSAFE-AP-PRIN-A-NUM	Principal applicant–ís A-number	APSS.csv	APRSAFEAPPRINANUM Principal applicants Anumber
4	APRSAFE-AP-CCO	Location code for the asylum office with jurisdiction over the credible fear case	APSS.csv	APRSAFEAPCCO Location code for the asylum office with jurisdiction over the credible fear case
5	APRSAFE-AP-LAST-NAME	last name.	APSS.csv	APRSAFEAPLASTNAME last name.
6	APRSAFE-AP-FIRST-NAME	first name.	APSS.csv	APRSAFEAPFIRSTNAME first name.
7	APRSAFE-AP-MIDDLE-NAME	middle name.	APSS.csv	APRSAFEAPMIDDLENAME middle name.
8	APRSAFE-AP-AKA-LAST-NAME	Alias for last name	APSS.csv	APRSAFEAPAKALASTNAME Alias for last name
9	APRSAFE-AP-AKA-FIRST-NAME	Alias for first name	APSS.csv	APRSAFEAPAKAFIRSTNAME Alias for first name
10	APRSAFE-AP-BIRTH-DATE	Birthdate	APSS.csv	APRSAFEAPBIRTHDATE Birthdate
11	APRSAFE-AP-SDX-LNAME-CODE		APSS.csv	APRSAFEAPSDXLNAMECODE
12	APRSAFE-AP-SDX-FNAME-CODE		APSS.csv	APRSAFEAPSDXFNAMECODE
13	APRSAFE-AP-SEX-CODE	Sex.	APSS.csv	APRSAFEAPSEXCODE Sex.
14	APRSAFE-AP-COB-CODE	Country of Birth	APSS.csv	APRSAFEAPCOBCODE Country of Birth
15	APRSAFE-AP-CITIZEN1-CODE	Applicant–ís 1st citizenship code	APSS.csv	APRSAFEAPCITIZEN1CODE Applicants 1st citizenship code
16	APRSAFE-AP-CITIZEN2-CODE	Applicant–ís 2nd citizenship code	APSS.csv	APRSAFEAPCITIZEN2CODE Applicants 2nd citizenship code
17	APRSAFE-AP-APPL-LANG1-CODE	Applicant's 1st language code	APSS.csv	APRSAFEAPAPPLLANG1CODE Applicant's 1st language code
18	APRSAFE-AP-PORT-OF-ENTRY-CODE	Port of Entry Code	APSS.csv	APRSAFEAPPORTOFENTRYCODE Port of Entry Code
19	APRSAFE-AP-US-ARRIVAL-DATE	US Arrival Date	APSS.csv	APRSAFEAPUSARRIVALDATE US Arrival Date

APRSAFE-AP-AKA-FIRST-NAME	Alias for first name	APSS.csv	APRSAFEAPAKAFIRSTNAME Alias for first name

CIVIC DIGITAL FELLOWSHIP

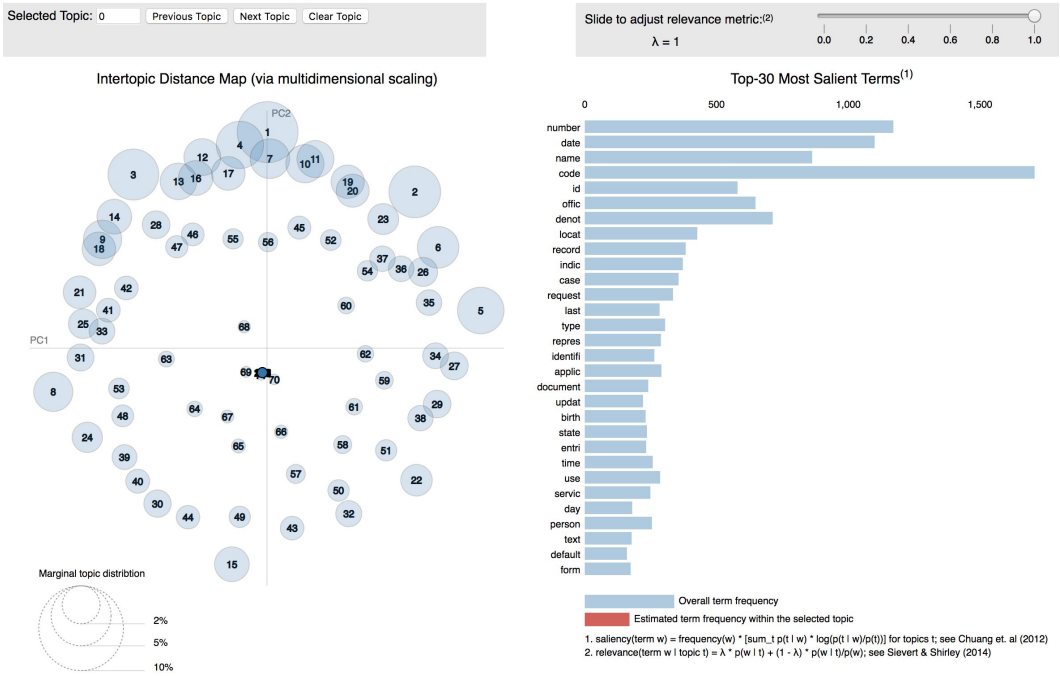
Preprocessing

- Load it into Python (*Applicant¬ís A-number*)
- Stripped punctucation (*applicantx96s a-number*)
- Removed Stop Words (*applicants a-number)*
- Stemmed (*applicant a-number)*
- Tokenized (“*applicant*” , ‘*a-number*”)



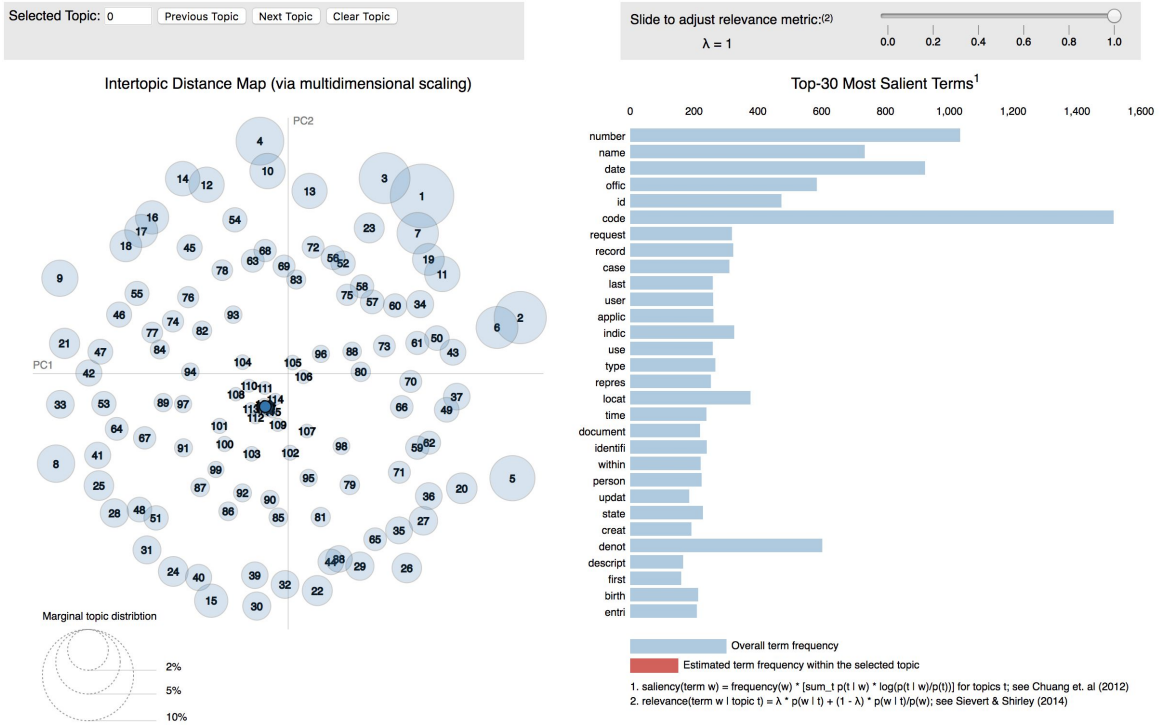
CIVIC DIGITAL FELLOWSHIP

Results



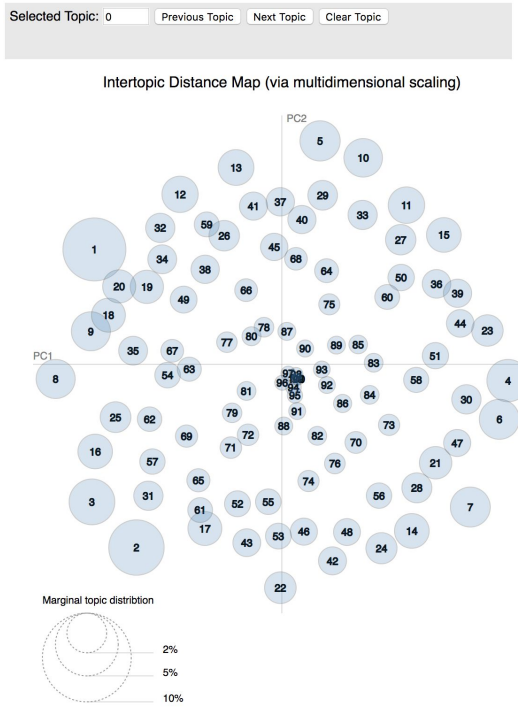
CIVIC DIGITAL FELLOWSHIP

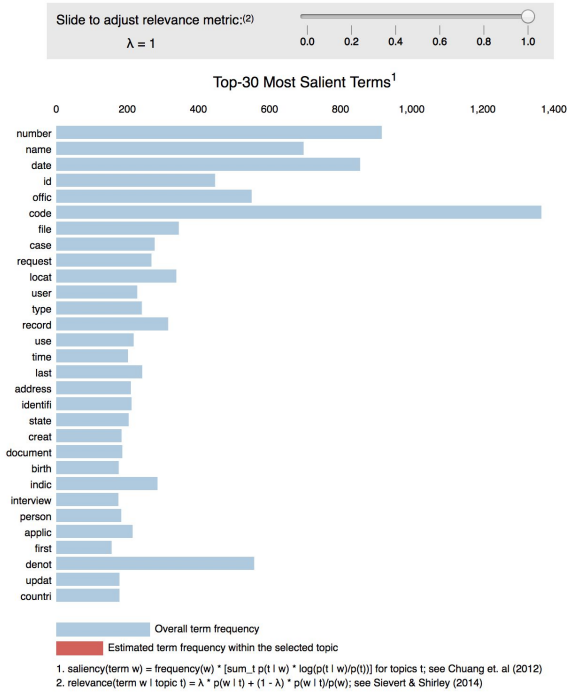
Results



CIVIC DIGITAL FELLOWSHIP

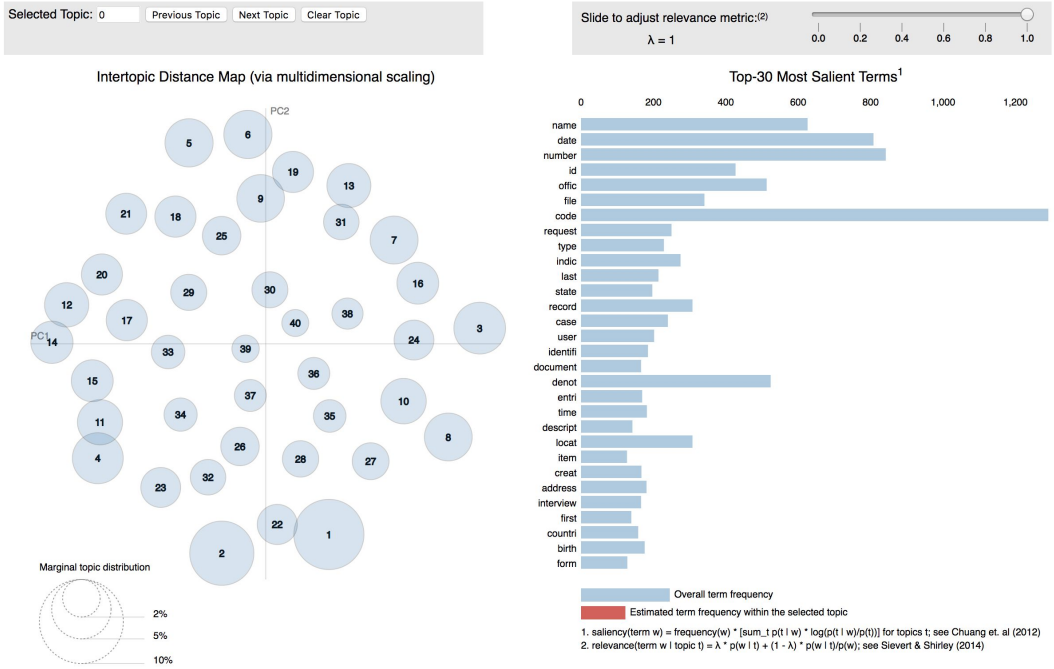
Results





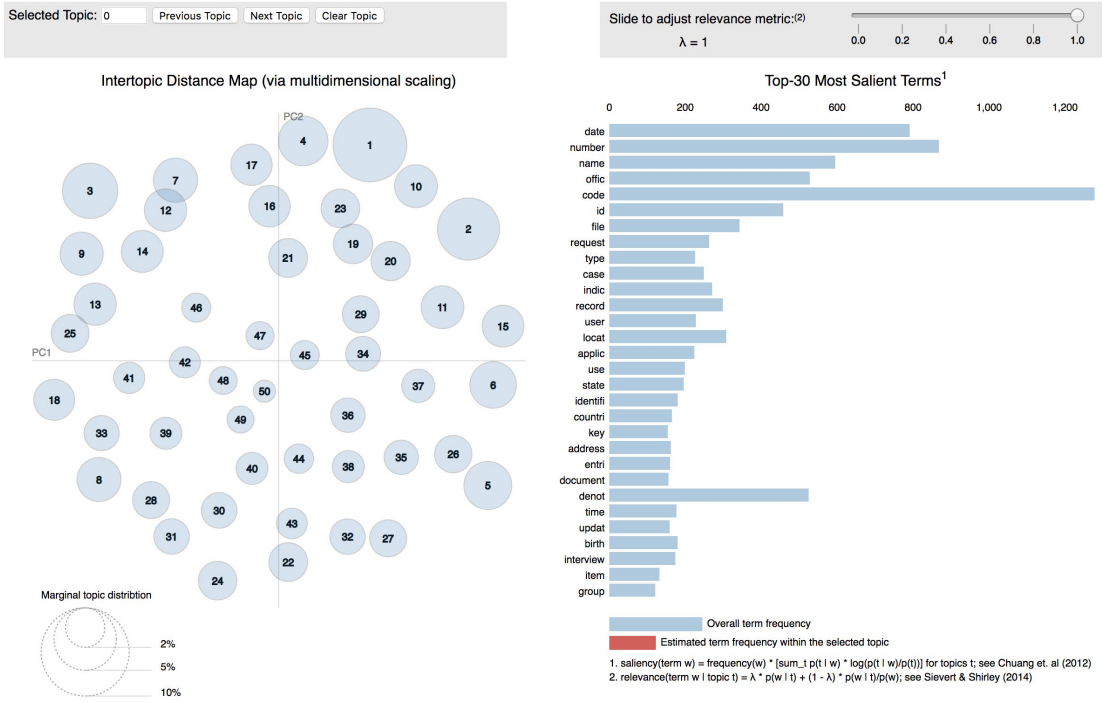
CIVIC DIGITAL FELLOWSHIP

Results



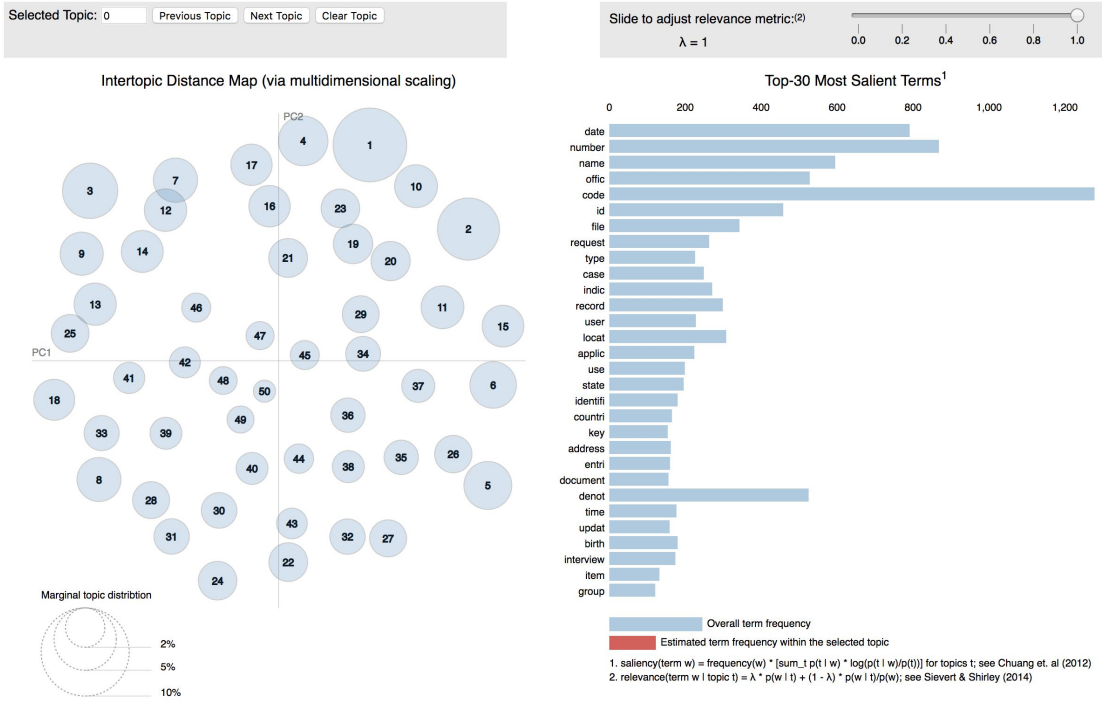
CIVIC DIGITAL FELLOWSHIP

Results



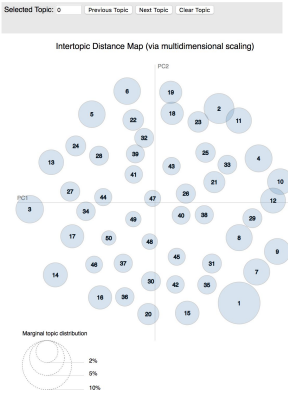
CIVIC DIGITAL FELLOWSHIP

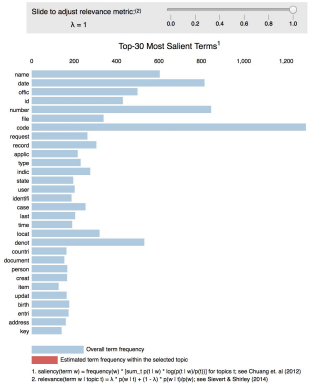
Results

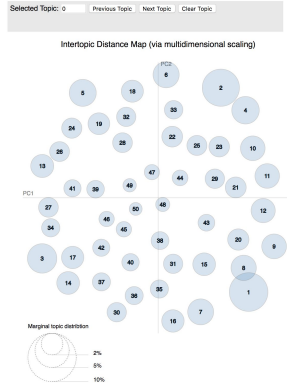


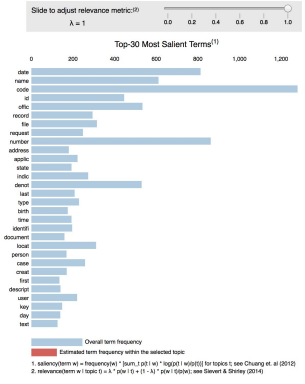
CIVIC DIGITAL FELLOWSHIP

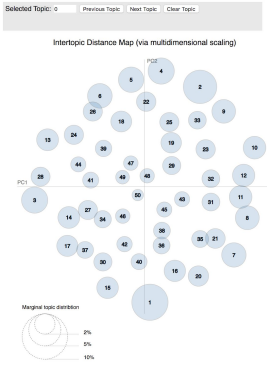
Results

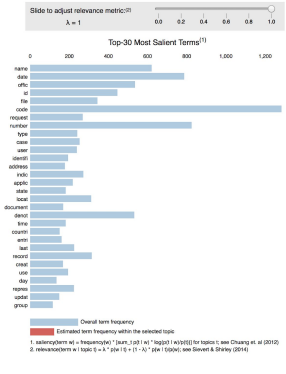






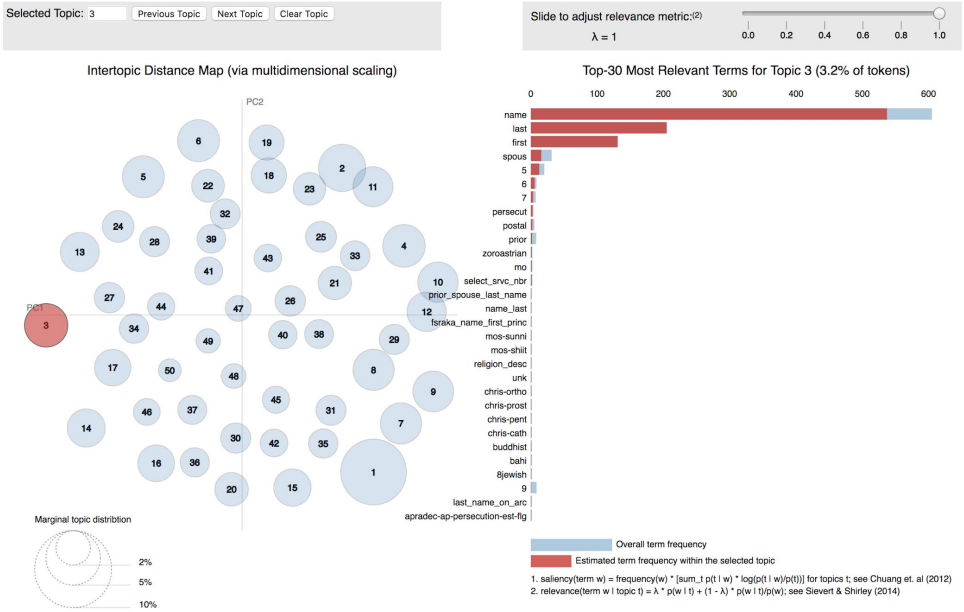






CIVIC DIGITAL FELLOWSHIP

Results



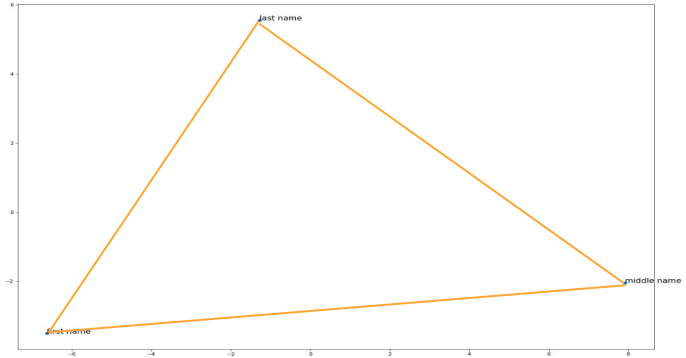
CIVIC DIGITAL FELLOWSHIP

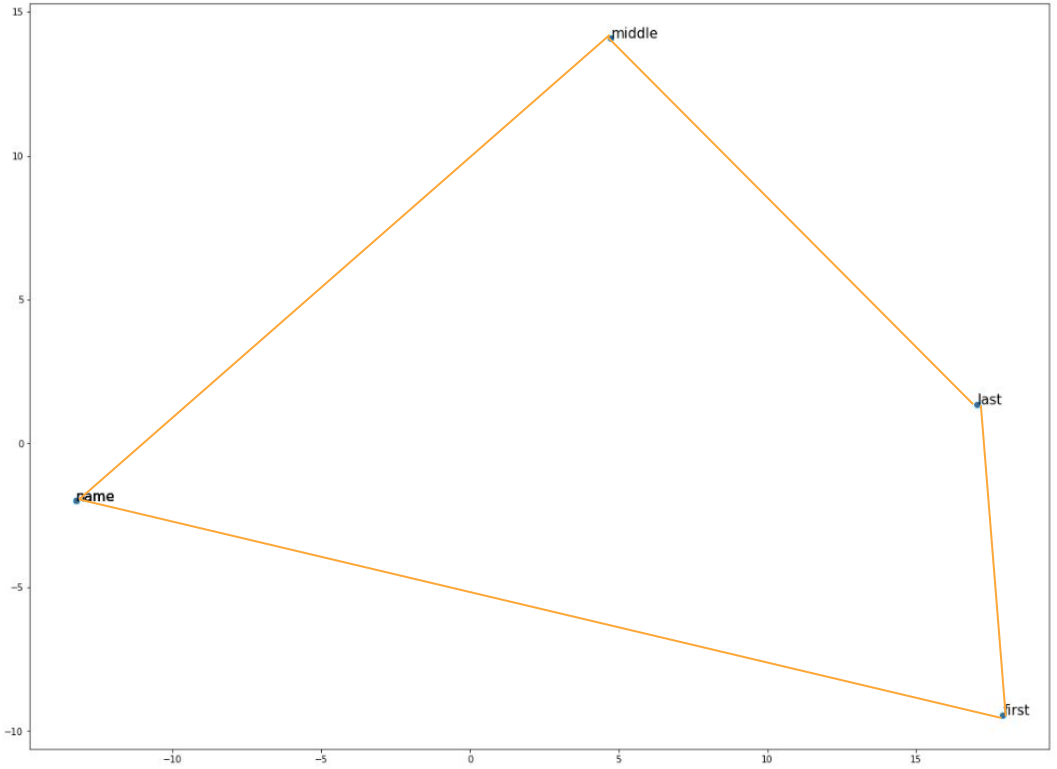
Takeaways

- Work with Data Standards Team (Data Governance)
- Try other word vectorizations:
    - Word2vec, GloVec
        - Different groupings
- Utilize named entity recognition - customize (Stanford Named Entity Recognizer (NER))
- Semantic Parsing (NLTK)

CIVIC DIGITAL FELLOWSHIP

Takeaways (continued)





CIVIC DIGITAL FELLOWSHIP

Thanks!

- OCDO (Damian, Justin, everyone)
- RAIO(Logan, Michael, Nevin, Christine, everyone)
- All of you!

CIVIC DIGITAL FELLOWSHIP