



AI Newsroom Tool: Automated FOIL Response Organizer

Cullen Fagan



The Problem

FOIL Requests are one of our really exciting tools as journalists – it's how we get a ton of vital information from the government **BUT** they're the worst! They take forever and are hard to keep track of.

Use Case: I am a reporter who is wants to cast a wide net in terms of FOILs – I'm talking filing multiple a day, just in case something sticks! But I don't want to clog my email with responses AND I anticipate most of them won't pan out.

The Solution

A dummy email account and an automated FOIL response categorizer!

How does the tool work?

The code is pretty straightforward: It uses imaplib to log into the account, read the bodies of new emails, and then creates a model based on training data that sorts the emails into types of responses. It then produces a data frame of these responses.

```
        if content_type == "text/plain":
            return part.get_payload(decode=True).decode()
        elif content_type == "text/html":
            return part.get_payload(decode=True).decode()
    else:
        return msg.get_payload(decode=True).decode()
```

```
In [130... results = []

for email_id in email_ids:
    status, msg_data = mail.fetch(email_id, "(RFC822)")

    for response_part in msg_data:
        if isinstance(response_part, tuple):
            msg = email.message_from_bytes(response_part[1])
            body = extract_email_body(msg)

            label = classify_email(body)

            results.append((body, label))
```

```
In [131... df = pd.DataFrame(results, columns=["email_content", "predicted_label"])
df
```

```
Out[131... 

|   | email_content                                     | predicted_label |
|---|---------------------------------------------------|-----------------|
| 0 | Please see the attached documents regarding yo... | records         |


```

Training Data

Content from real FOIA response letters, manually sorted into 50 "extensions", "denials", "records", and "other".

I sourced the training from muckrock.com, where they have a HUGE database of FOIA requests and responses. I used 200 responses for the training set.

		A:		B:
1	response_content	Our office has no responsive records [jcd4804050-0280-0b6c-9b6c-80b16e47a9f]		response
		Yamuna (Yam) Manion General Counsel/Assistant State Comptroller		
2		Office of the State Comptroller 165 Capitol Ave Hartford, CT 06106 860-732-3312		
		This is information and any of its attachments are for the sole use of the intended recipient(s) and it may contain information that is proprietary, denials, Good morning, This is Jefferson County Wisconsin not Alabama. Thanks, Megan Wagner		
		Megan Wagner Administrative Secretary Jefferson County Sheriff's Office 411 S. Center Ave Jail, WI 53589 Direct line: 800-474-7329 Fax: 920-474-7126		
		December 4, 2024Alan KananRock NewsDEPT MR 176992930 Huntington AvenueBoston, Massachusetts 02115Dear Mr. Kanan: p.2 FOIA Kana Page 2December 4, 2024Response to the appeal within 10 business days after its receipt. Under unusual circumstances, the denial Good afternoon:		denial
		Kansas State Department of Education staff conducted a search of all KSDOE email accounts for the use of the word "bills" since Jan I have confirmed with agency leadership that no discussions have taken place within the last two years, nor is there any movement to Kansas State Department of Education logohttps://www.ksde.org/ R. Scott Gordon General Counsel Office of General Counsel 785-296-3254 reg@ksde.org www.ksde.orghttp://www.ksde.org/		
1		Kansas leads the world in the success of each student. This message, the information contained therein and any attachments thereto may be protected by attorney client privilege and/or we Be notified of all e-mail communication is not a secure method of communication; (2) any e-mail that is sent between you and this		
		The Kansas State Department of Education does not discriminate on the basis of race, color, national origin, sex, disability or age in a		denial
		Dear P. B. Sai, This is in final response to your 23-02-24 Freedom of Information Act (FOIA) request for Dear Air Force.		
		"On behalf of myself and Fiat Flendrum, I hereby request the following records: 1. The one email in your chief FOIA officer's email in Your request as submitted did not sufficiently describe a Department of the Air Force (DAF) record. The request is not "perfected" b As such, after careful review of your request, we are administratively closing it as it does not meet the requirements under the FOIA. I Should you decide to appeal this decision, you must write to the Secretary of the Air Force, through: SA/IAAF (FOIA), 1000 Air For		
6		If you are dissatisfied with the servicing of this request or this final responses from this office, you may contact the Air Force FOIA Pub Please contact this office at the address above, via email at SAFAA-HAF FOIA.Work@us.af.mil, should you have any questions i		
		Sincerely: HAF FOIA TEAM SAFAA (FOIA) Office of the Secretary of the Air Force 1000 Air Force Pentagon Washington, DC 20330-1000		denial
		Dear Sir, This is in final response to your 24-02-24 Freedom of Information Act (FOIA) request for: - 1. The one email in your chief FOIA officer's Your request as submitted does not sufficiently describe a Department of the Air Force (DAF) record, specifically a United States Spa As such, after careful review of your request, we are administratively closing it as it does not meet the requirements under the FOIA. I Should you decide to appeal this decision, you must write to the Secretary of the Air Force, through: SAFAA (FOIA), 1000 Air For If you are dissatisfied with the servicing of this request or this final responses from this office, you may contact the Air Force FOIA Pub Please contact this office at the address above, via email at SAFAA-HAF FOIA.Work@us.af.mil, should you have any questions i		
7		Sincerely: HAF FOIA TEAM SAFAA (FOIA) Office of the Secretary of the Air Force 1000 Air Force Pentagon Washington, DC 20330-1000 Virginia Military Institute		denial
		***** A message was sent to you regarding record request #24-08- *****		
8		Your request is very broad and involves several different departments or posts. It will take a substantial amount of time to access, sea What was provided to you was an estimate of staff time involved in accessing, searching, reviewing, and redacting (if necessary) any Whether responsive records exist in electronic format is irrelevant. Each record is reviewed to determine if the record is responsive, a ***** *****		
		400-Questions about your request?***** Reply to this email or sign in to contact staff at Virginia Military Institute.*****400-Technical denia		

Demo

Next Steps

For Records

- Download pdf attachments through imaplib
- Forward them to my email.

For Extensions

- Use Regex to pull the date indicated in the extension notification.
- Put dates into df
- Reminder on date?

For Denials

- Put all denials in a spreadsheet with their content and forward to me in an email to review for appeals.

Other?

For emails that can't be sorted into the above categories, we can either dismiss them by getting really good at IDing the first three OR do what we probably should do, which is break them down further – categories such as “requests for payment” and “receipt.”

What libraries would I need for the next steps?

- re
- smtplib
- datetime, probably

Limitations + Improvements

Missing Categories

Problem

There are way more kinds of responses than the ones I've indicated – the 'other' category is way too broad which is confusing the data and making it hard to identify the other categories.

Solution

More categories in the training data.

Better Indicators

Problem

Right now, this tells me broadly about what's happening in my inbox, and can point me to check if there are new records, etc. But it's not super specific if I want to check individual FOIAs in the spreadsheet, and it's not historical – it only checks new emails.

Solution

Regex to pull FOIA ID, sender, other identifiers.

Accuracy

Problem

The model isn't very accurate right now.

Solution

It needs WAY more training data.



Thank you!