

# Text analysis for drug use categorization on social media

Adi Mukundan

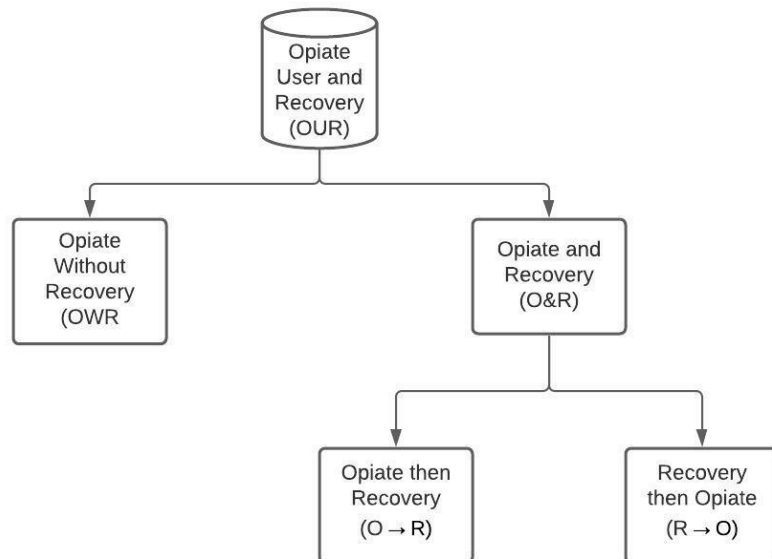
# The data

Reddit dump files from December 2009 to February 2022

# users	# posts	Opiate Subreddits	Recovery Subreddits
122,200	7,940,558	7	13

Recovery Subreddits
OpiatesRecovery
suboxone
Methadone
addiction
Opiatewithdrawal
REDDITORSINRECOVERY
NarcoticsAnonymous
recovery
Subutex
Naranon
SMARTrecovery
buddhistrecovery
drugrehabcenters

Opiate Subreddits
opiates
OpiateChurch
heroin
PoppyTea
glassine
opiatescirclejerk
HeroinHighway



Grouping	# users
Number of users posting to opiate and opiate recovery subreddits (OUR):	78,388
Number of users posting to opiate without recovery subreddits (OWR):	68,989
Number of users who posted in both opiate and a recovery subreddit (O&R):	9,399
Number of users who posted to opiate then a recovery subreddit (O → R):	6,016
Number of users who posted to recovery then posted to an opiate subreddit (R → O):	3,383

# DUI

- Parse a post and categorize terms/post using the DUI parser

Paraphrased post	Categories and terms identified using DUI
Okay so today is my 61 <sup>st</sup> day in <u>recovery</u> . I was deep into dealing and using <u>acid</u> and <u>DMT</u> . I was eventually put into <u>rehab</u> , which I quickly signed myself out of. Now Im in <u>IOP</u> ( <u>intensive out patient</u> ) and have been going to <u>NA meetings</u> regularly since.	<b>drug_recovery.recovery:</b> recovery; <b>drug_recovery.recovery_support.aa:</b> meetings, na, na_meetings; <b>drug_recovery.recovery_support.recovery programs:</b> out_patient, iop, iop_intensive; <b>drug_recovery.recovery_support.rehab:</b> rehab; <b>drug_terms.hallucinogens:</b> acid, dmt

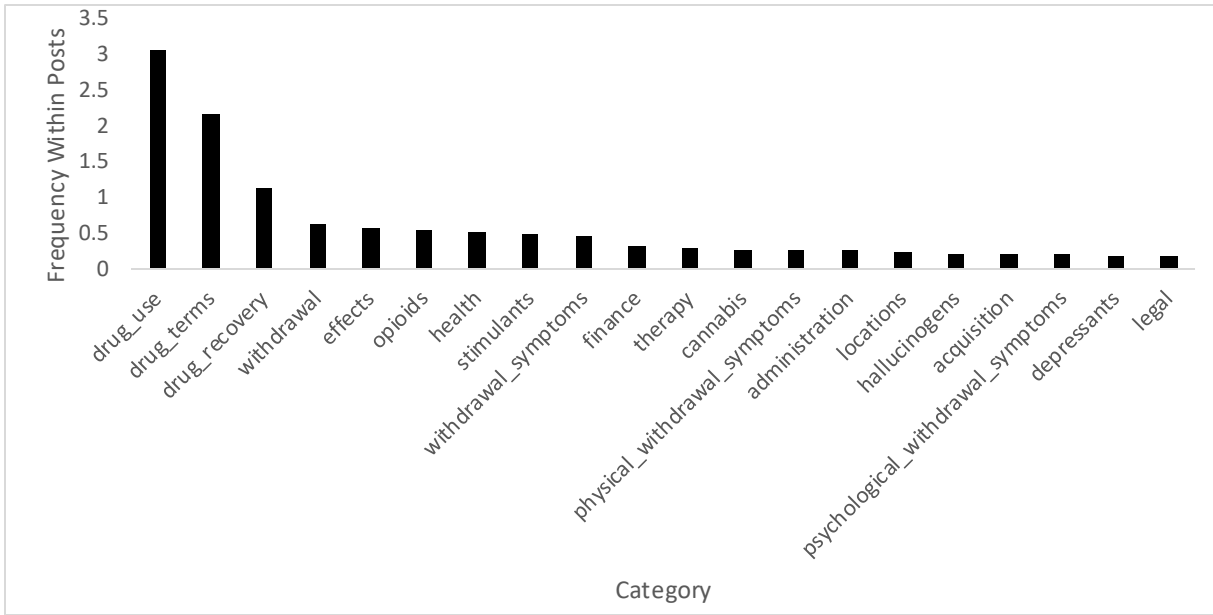


# Data: Subreddit monthly intervals

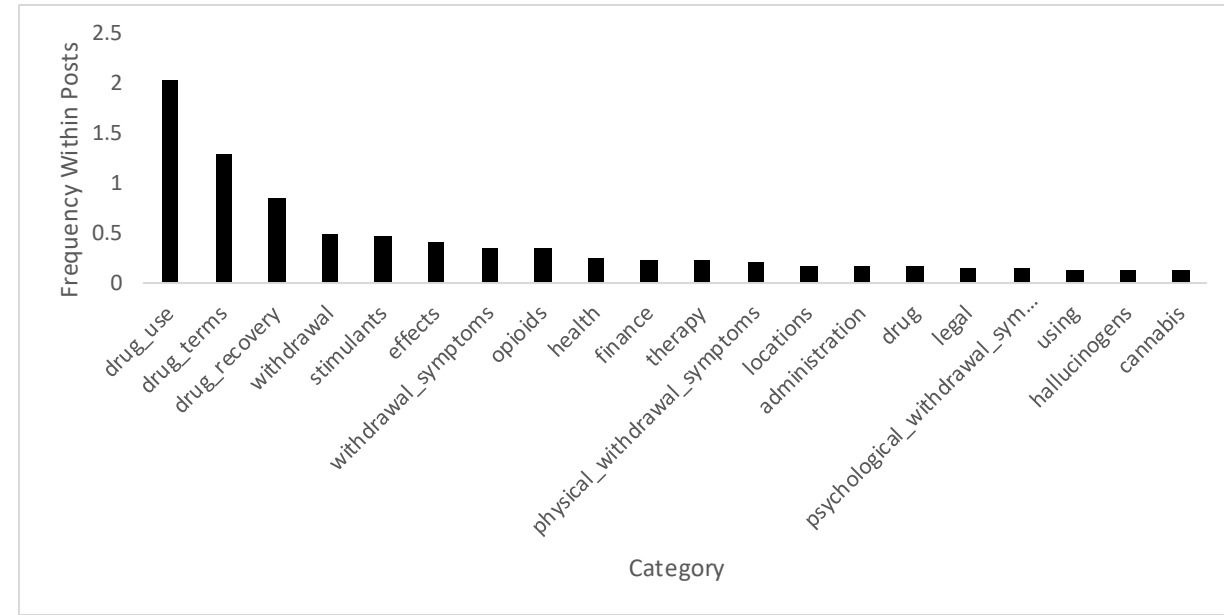
Trying to get similar insights into population levels of the data by looking at subreddits

subreddit	[2016, 10]	[2017, 4]	[2017, 10]	[2018, 4]	[2018, 10]	[2019, 4]	[2019, 10]	[2020, 4]	[2020, 10]	[2021, 4]	[2021, 9]	total posts
opiates	1	1	1	1	1	1	1	1	1	1	1	213134
AskReddit	1	1	1	1	1	1	1	1	1	1	1	83542
Drugs	1	1	1	1	1	1	1	1	1	1	1	81441
trees	1	1	1	1	1	1	1	1	1	1	1	44851
benzodiazep	1	1	1	1	1	1	1	1	1	1	1	41560
OpiateChurc	0	0	0	0	0	1	1	1	1	1	1	39398
politics	1	1	1	1	1	1	1	1	1	1	1	29544
heroin	0	0	1	1	1	1	1	1	1	1	1	25853
The_Donald	1	1	1	1	1	1	1	0	0	0	0	24374

# Text Analysis



Old



New

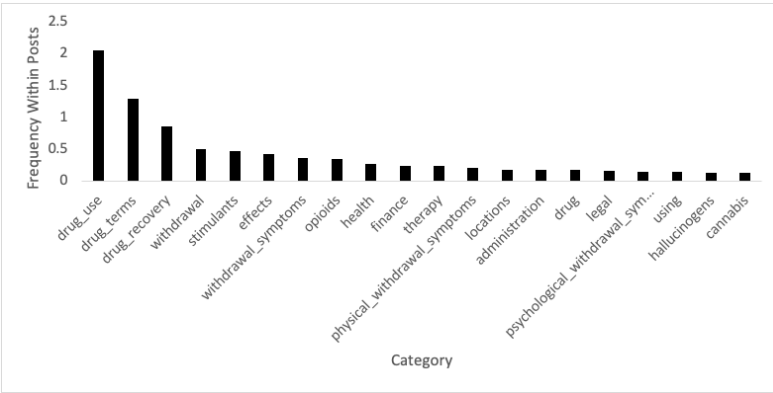
**Figure 14: Frequency of DUI categories (OWR).** This chart displays the twenty most frequent DUI categories appearing within all subreddits from users in the OWR grouping.

Category	Frequency
drug_use	3.07
drug_terms	2.18
drug_recovery	1.12
withdrawal	0.64
effects	0.57

Category	Frequency
drug_use	2.04
drug_terms	1.30
drug_recovery	0.86
withdrawal	0.50
stimulants	0.47

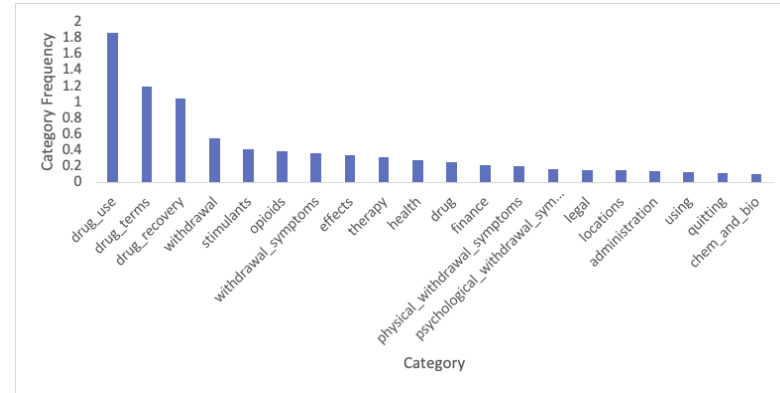
# DUI category frequency across user groupings

## OWR



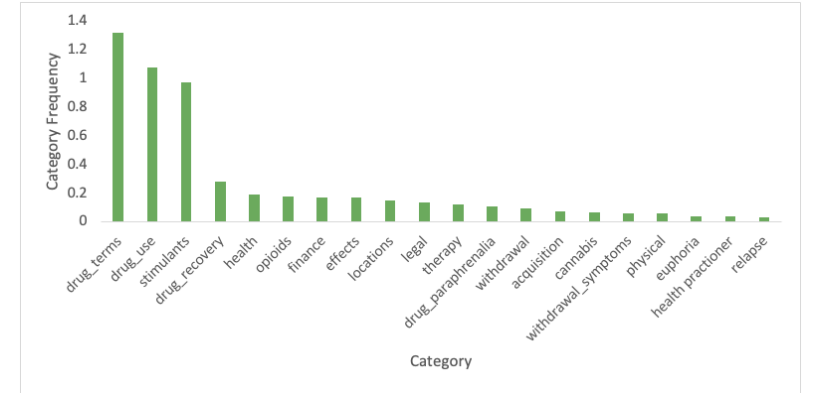
Category	Frequency
drug_use	2.04427045
drug_terms	1.29658801
drug_recovery	0.85523352
withdrawal	0.49551433
stimulants	0.46824628

## O→R



Categories	Frequency
drug_use	1.86513427
drug_terms	1.19191594
drug_recovery	1.04942377
withdrawal	0.54542169
stimulants	0.41288983

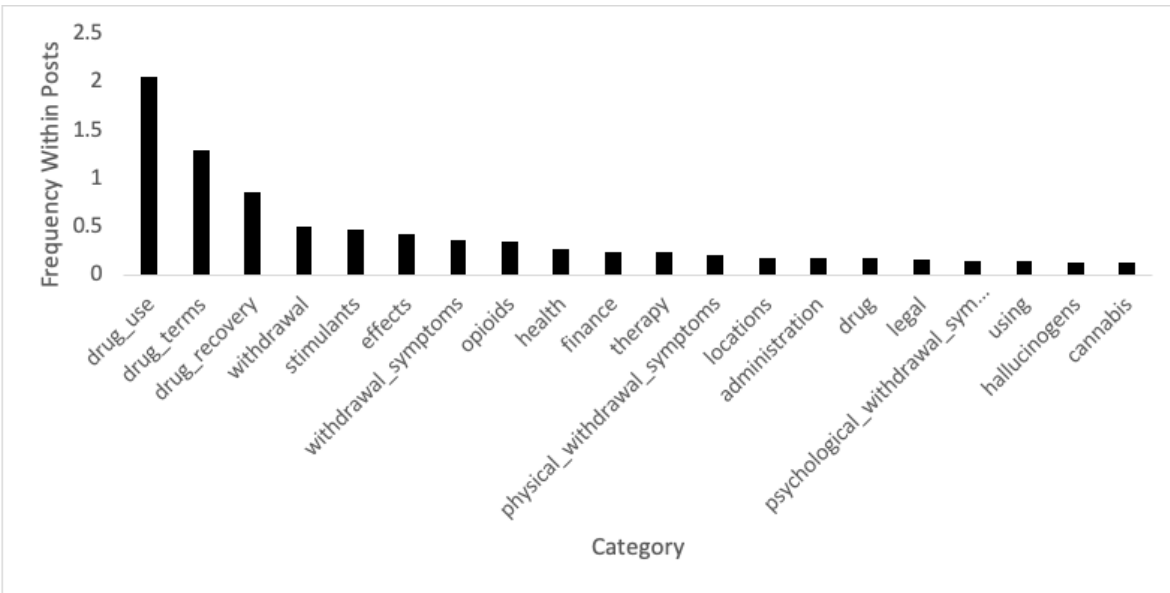
## R→O



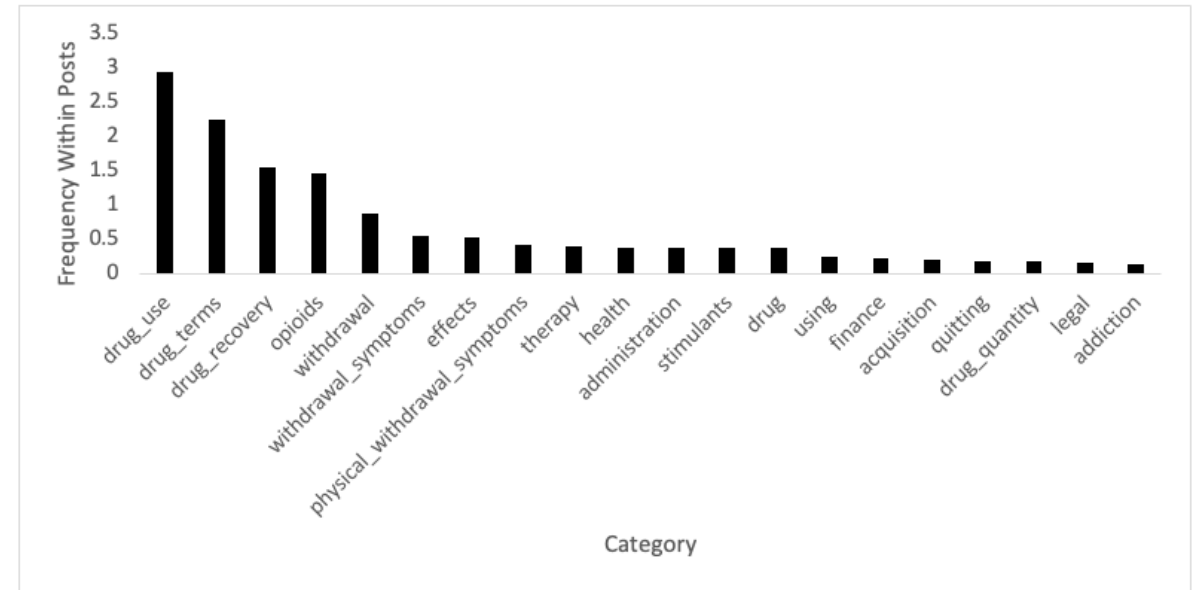
Category	Frequency
drug_terms	1.31339898
drug_use	1.07692653
stimulants	0.97272595
drug_recovery	0.28206021
health	0.18583186

# DUI category frequency changes within OWR

All subreddits



Opiate subreddits



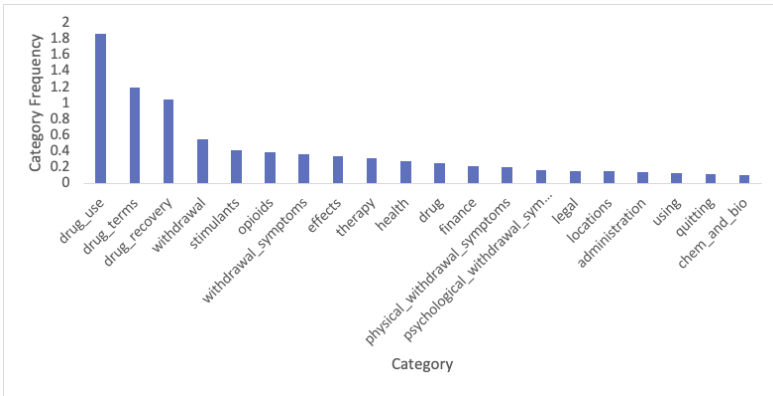
Category	Frequency
drug_use	2.04427045
drug_terms	1.29658801
drug_recovery	0.85523352
withdrawal	0.49551433
stimulants	0.46824628

Category	Frequency
drug_use	2.92489556
drug_terms	2.23138916
drug_recovery	1.54469474
opioids	1.46991774
withdrawal	0.87650106

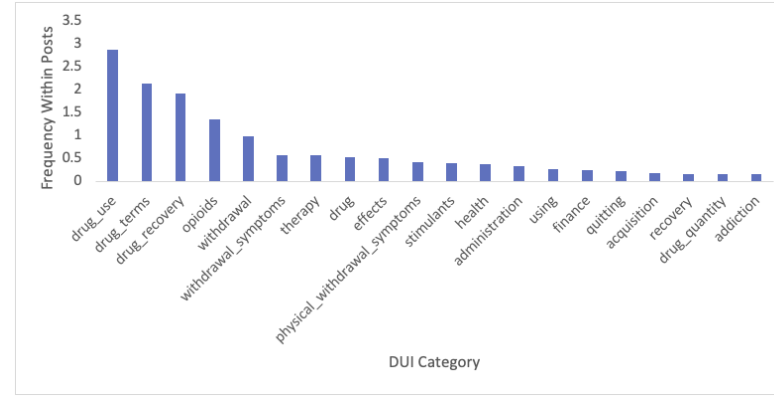


# DUI category frequency changes within O→R

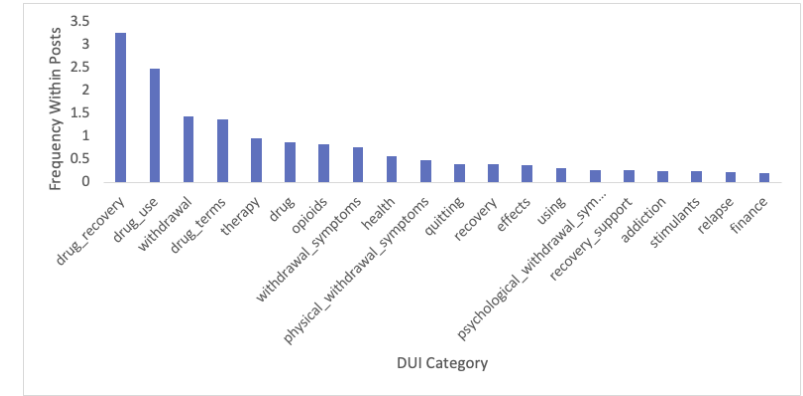
All subreddits



Opiate subreddits



Recovery subreddits



Categories	Frequency
drug_use	1.86513427
drug_terms	1.19191594
drug_recovery	1.04942377
withdrawal	0.54542169
stimulants	0.41288983

Category	Frequency
drug_use	2.87462875
drug_terms	2.12373932
drug_recovery	1.90691653
opioids	1.34813398
withdrawal	0.9899904

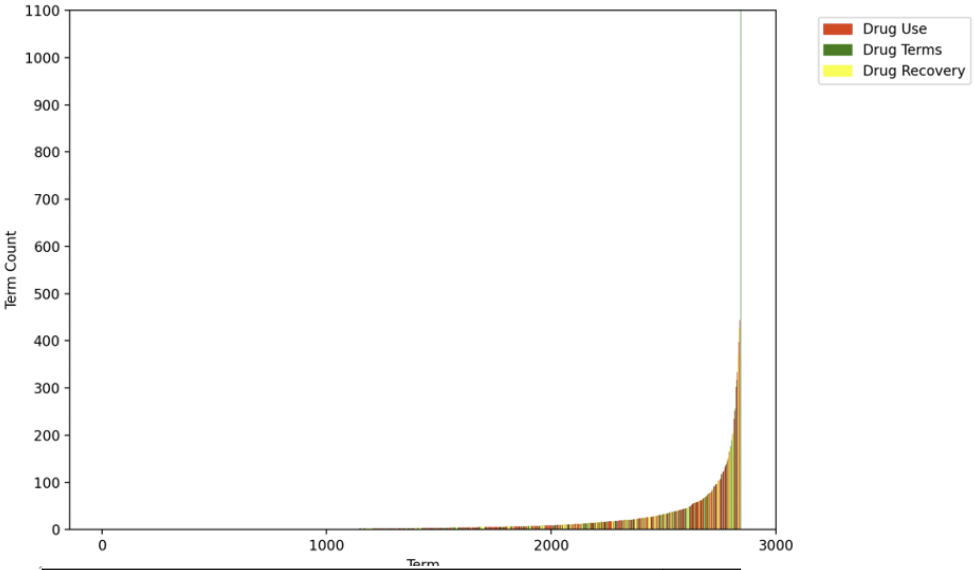
Category	Frequency
drug_recovery	3.26197153
drug_use	2.48593031
withdrawal	1.42873994
drug_terms	1.37258622
therapy	0.95483077

# Term categorizations in the OWR grouping

```

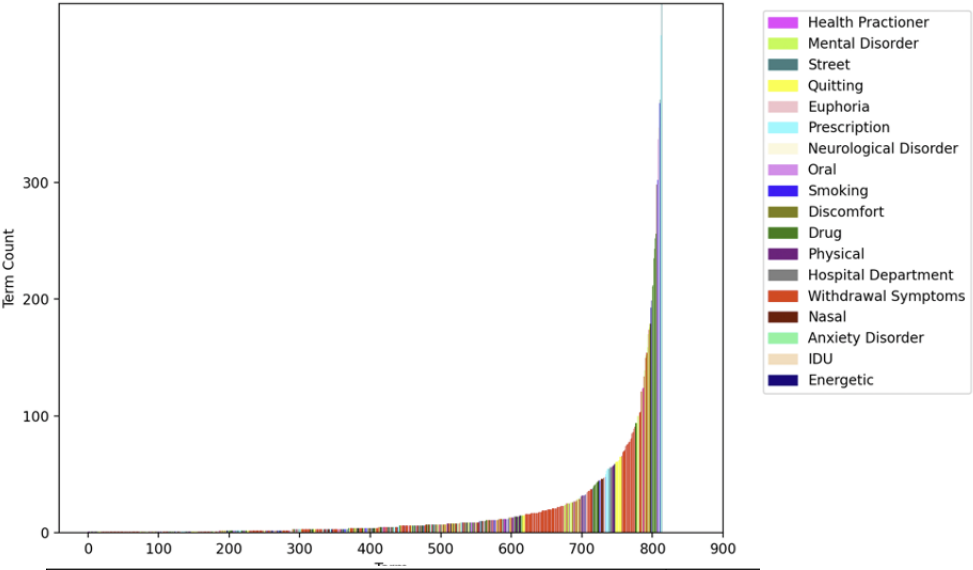
Term/phrase      Tier 1      Tier 2      Tier 3      Tier 4
cold_clammy,drug_recovery,withdrawal,withdrawal_symptoms,physical_withdrawal_symptoms
cold_flashes,drug_recovery,withdrawal,withdrawal_symptoms,physical_withdrawal_symptoms
  
```

Tier 1



Term:	Count:
amp	1486
high	1055
pain	653
money	629
dose	483

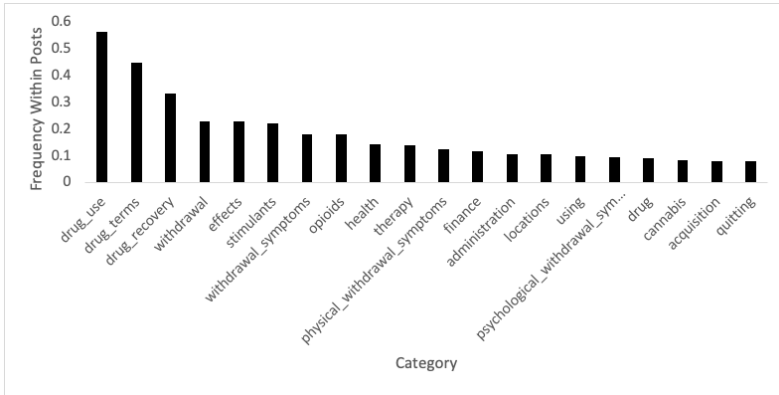
Tier 3



Term:	Count:
pain	653
stop	426
anxiety	371
kratom	368
happy	347

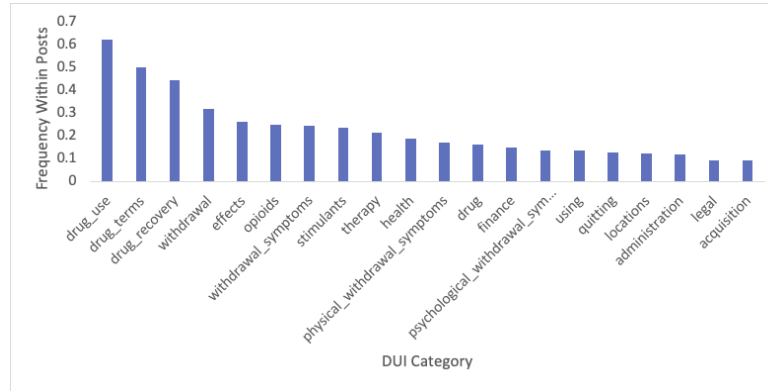
# Presence – absence of categories across groups

**OWR**



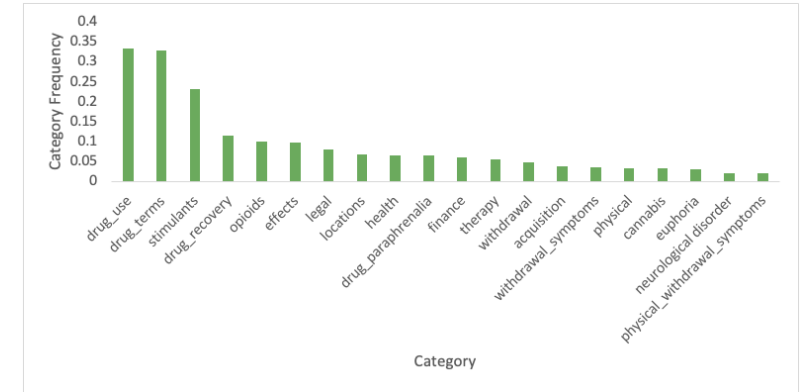
Category	Frequency
drug_use	0.5634
drug_terms	0.4489
drug_recovery	0.3303
withdrawal	0.2293
effects	0.2292

**O→R**



Categories	Frequency
drug_use	0.622
drug_terms	0.4985
drug_recovery	0.4435
withdrawal	0.316
effects	0.2617

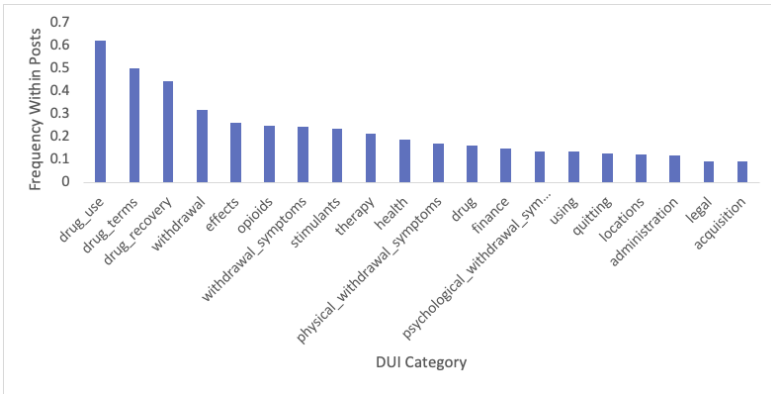
**R→O**



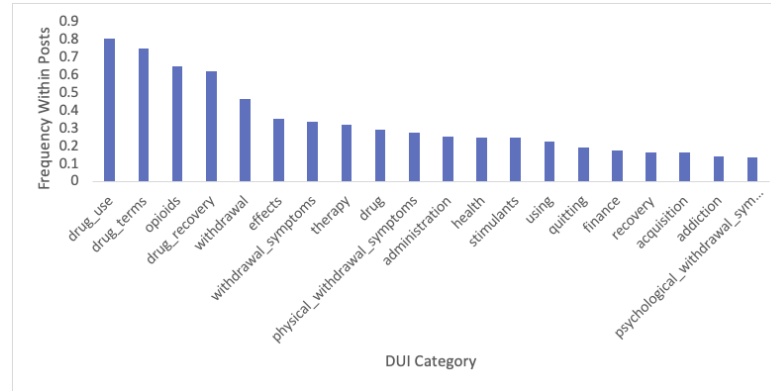
Category	Frequency
drug_use	0.3341
drug_terms	0.3282
stimulants	0.2307
drug_recovery	0.114
opioids	0.1003

# Presence – absence changes within $O \rightarrow R$

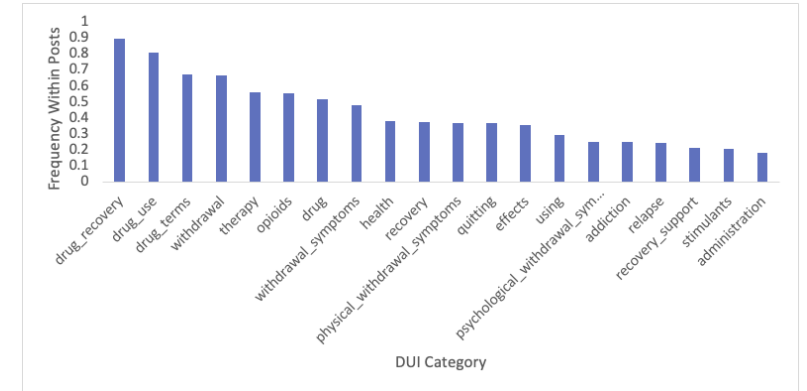
All subreddits



Opiate subreddits



Recovery subreddits



Categories	Frequency
drug_use	0.622
drug_terms	0.4985
drug_recovery	0.4435
withdrawal	0.316
effects	0.2617

Categories	Frequency
drug_use	0.8042
drug_terms	0.7492
opioids	0.6487
drug_recovery	0.6236
withdrawal	0.4632

Category	Frequency
drug_recovery	0.8965
drug_use	0.8098
drug_terms	0.6707
withdrawal	0.6647
therapy	0.5583

# Conclusions

- The current implementation of the DUI parser on my dataset shows promise to be useful to answer earlier questions
- More work to be done to verify my database and parser before drawing conclusions
- Questions?