DATA MINING

BY ADAM TAN - CRL-TURING 4.22 - 13 NOV 2020

OBJECTIVE

- TO ANALYSIS TURING 4.22 DISCORD CHANNEL
- TO OBTAIN THE WATCH LINK FROM DISCORD THE-WATCH CHANNEL

METHODOLOGY

- DISCORD CHAT EXPORTER
- NLTK
- WORDCLOUD
- PANDAS

DISCORD CHAT EXPORTER

- DISCORD CHAT EXPORTER CAN BE USED TO EXPORT MESSAGE HISTORY FROM A DISCORD CHANNEL TO A FILE. IT WORKS WITH DIRECT MESSAGES, GROUP MESSAGES, SERVER CHANNELS, SUPPORTS DISCORD'S DIALECT OF MARKDOWN AND ALL OTHER RICH MEDIA FEATURES.

https://github.com/Tyrrrz/DiscordChatExporter

NLTK - NATURAL LANGUAGE TOOLKIT

- NLTK IS A LEADING PLATFORM FOR BUILDING PYTHON PROGRAMS TO WORK WITH HUMAN LANGUAGE DATA. IT PROVIDES EASY-TO-USE INTERFACES TO OVER 50 CORPORA AND LEXICAL RESOURCES SUCH AS WORDNET, ALONG WITH A SUITE OF TEXT PROCESSING LIBRARIES FOR CLASSIFICATION, TOKENIZATION, STEMMING, TAGGING, PARSING, AND SEMANTIC REASONING, WRAPPERS FOR INDUSTRIAL-STRENGTH NLP LIBRARIES.

https://www.nltk.org/

WORDCLOUD

- A LITTLE WORD CLOUD GENERATOR IN PYTHON.

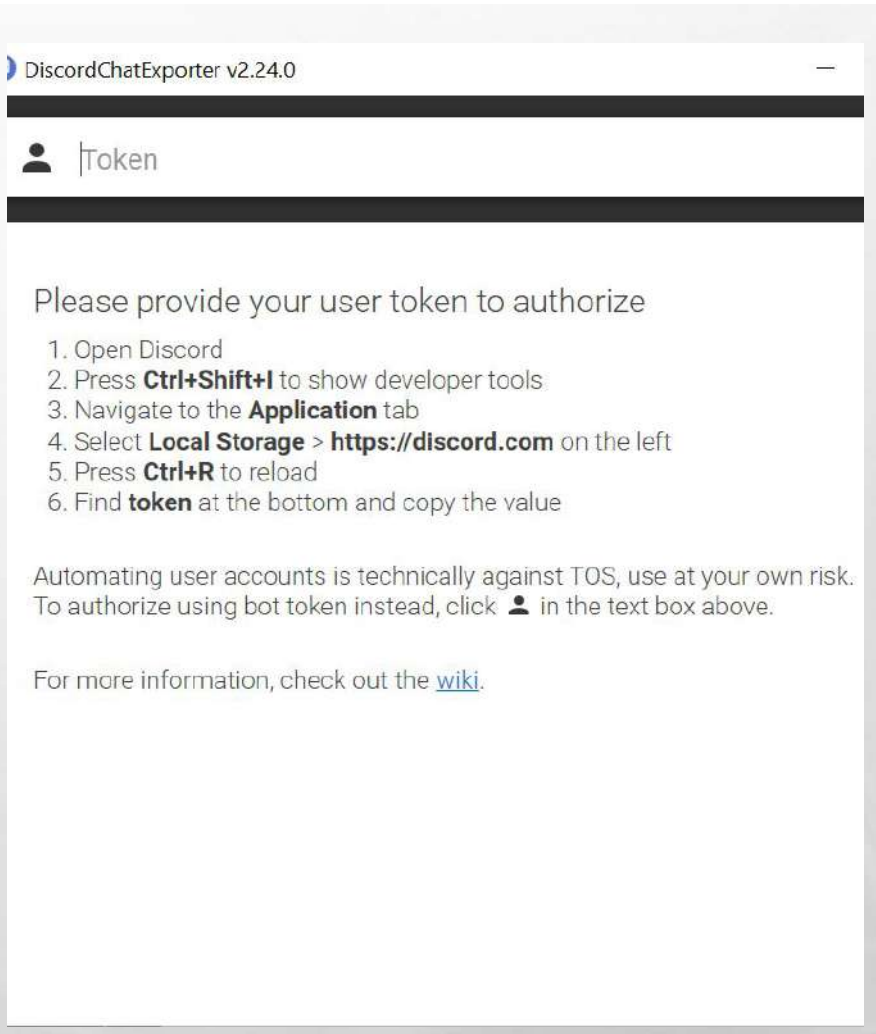
https://github.com/amueller/word_cloud

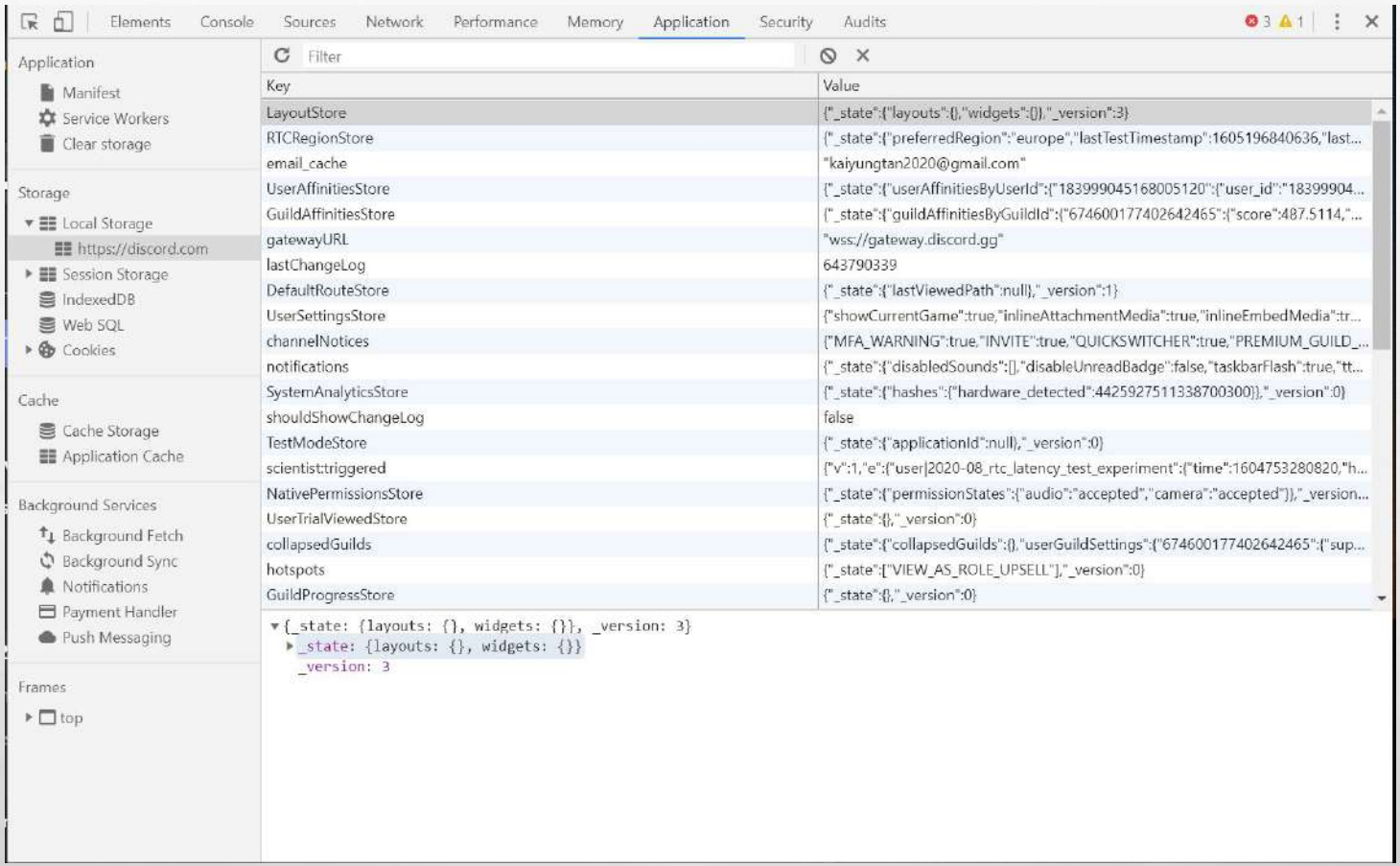
DISCORD CHAT EXPORTER

SCREENSHOTS

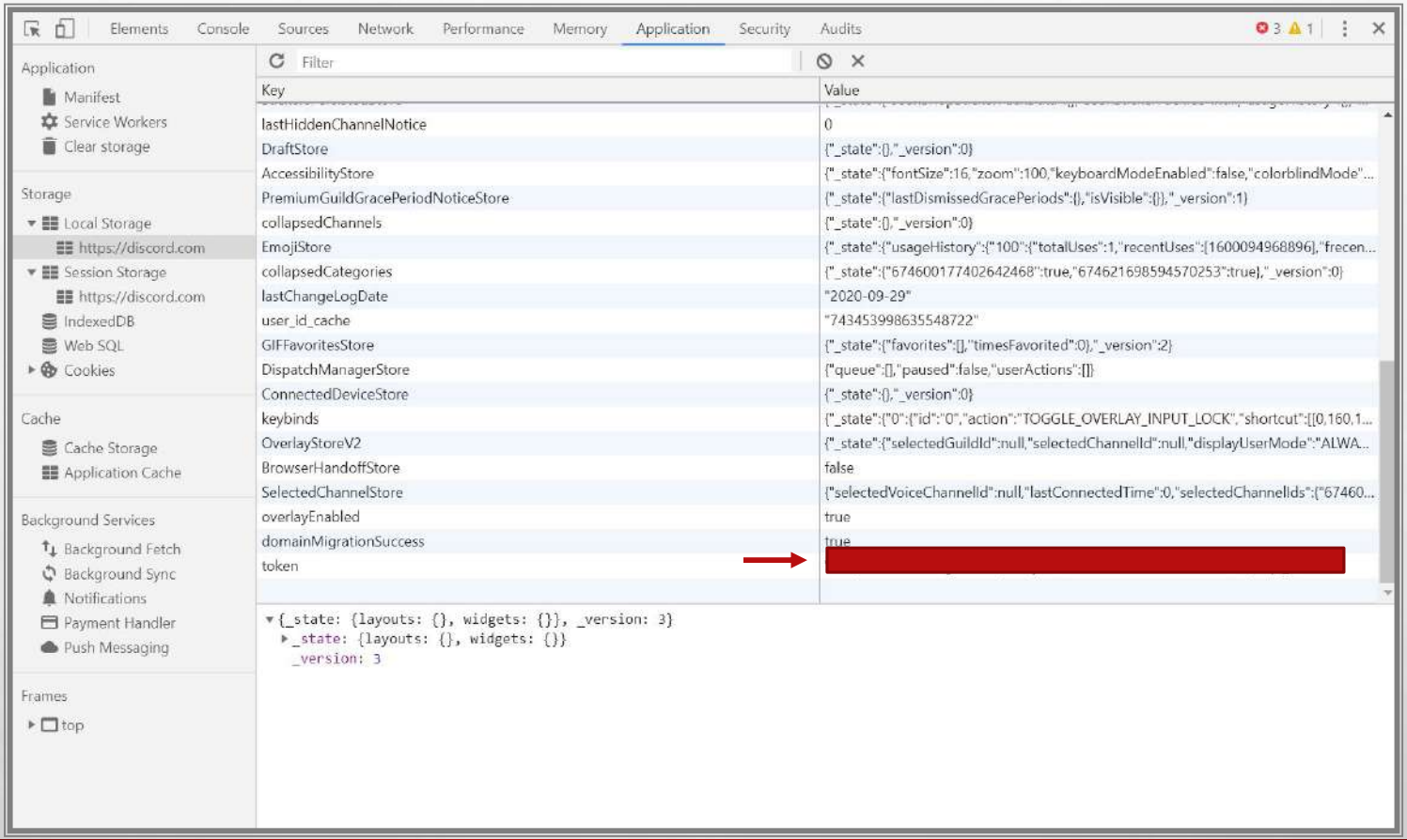
DOWNLOAD:

https://github.com/Tyrrrz/DiscordChatExporter/releases/tag/2.24

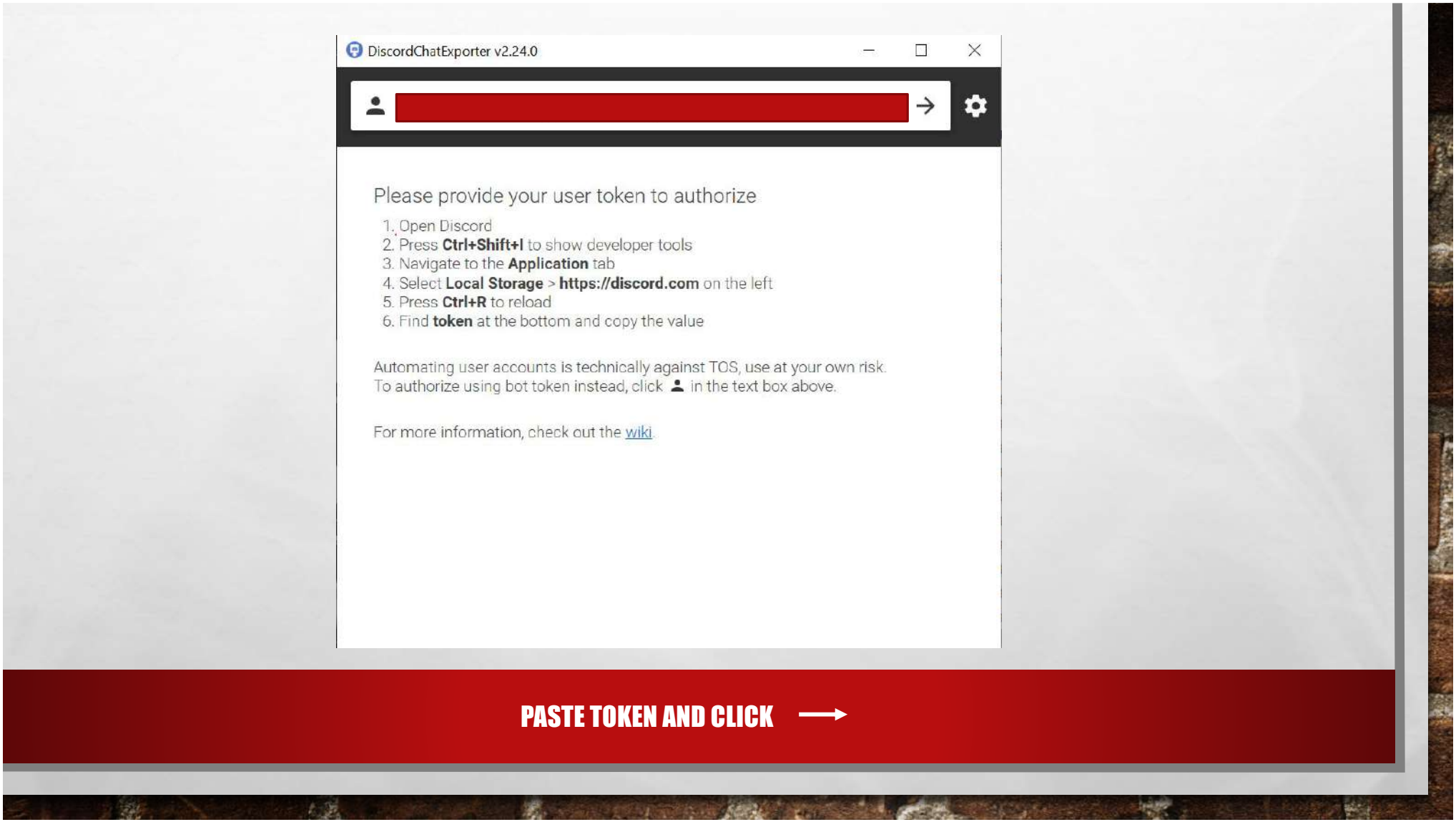


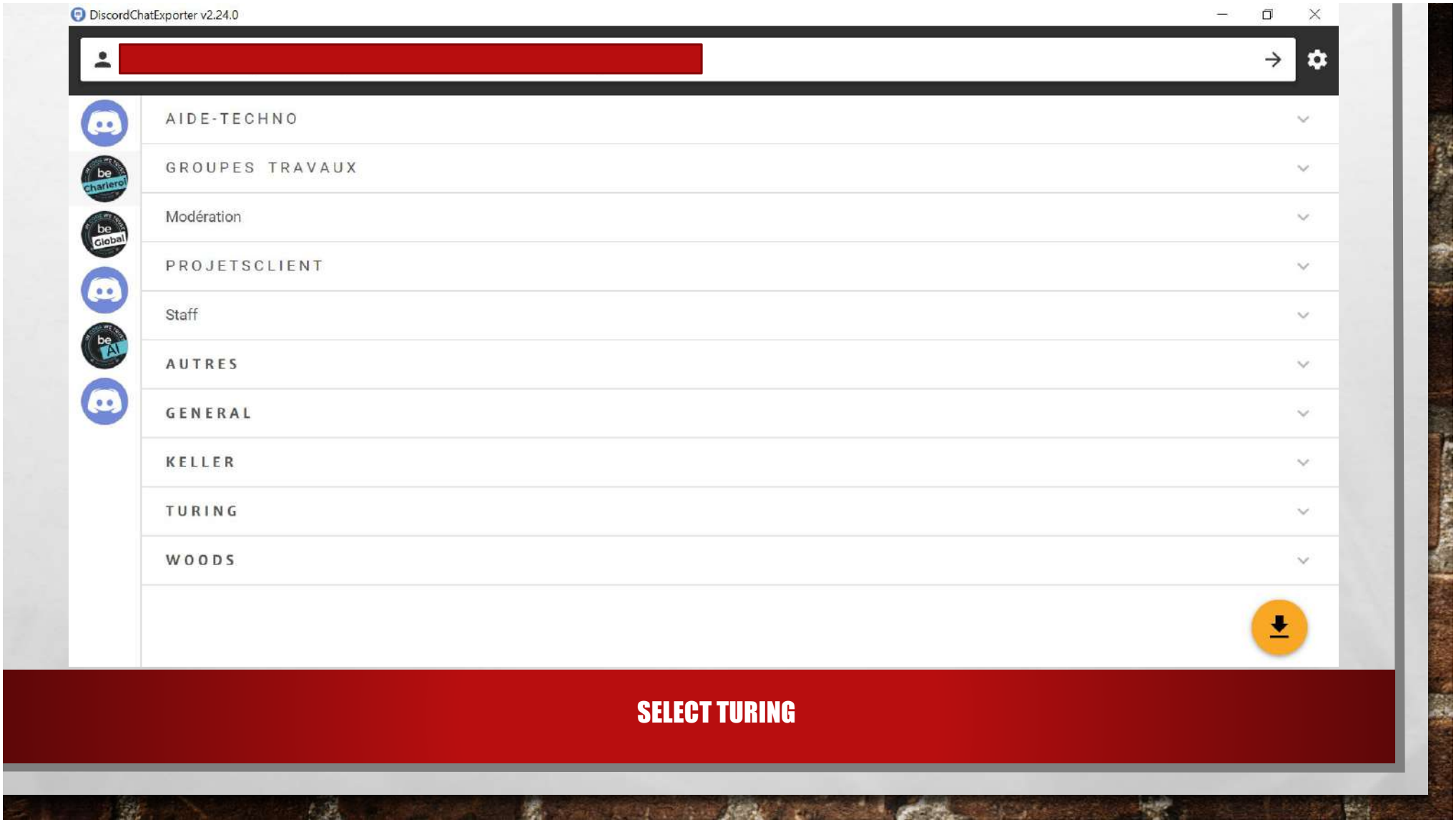

Elements
Console
Sources
Network
Performance
Memory
Application
Security
Audits
Application
Manifest
Service Workers
Clear storage
Storage
Local Storage
https://discord.com
Session Storage
IndexedDB
Web SQL
Cookies
Cache
Cache Storage
Application Cache
Background Services
Background Fetch
Background Sync
Notifications
Payment Handler
Push Messaging
Frames
top
Filter
Key
Value
LayoutStore
RTCRegionStore
email_cache
UserAffinitiesStore
GuildAffinitiesStore
gatewayURL
lastChangeLog
DefaultRouteStore
UserSettingsStore
channelNotices
notifications
SystemAnalyticsStore
shouldShowChangeLog
TestModeStore
scientist:triggered
NativePermissionsStore
UserTrialViewedStore
collapsedGuilds
hotspots
GuildProgressStore
{"_state":{"layouts":{},"widgets":{}},"_version":3}
"kaiyungtan2020@gmail.com"
"wss://gateway.discord.gg"
643790339
false
{_state: {layouts: {}, widgets: {}}, _version: 3}
_state: {layouts: {}, widgets: {}}
_version: 3

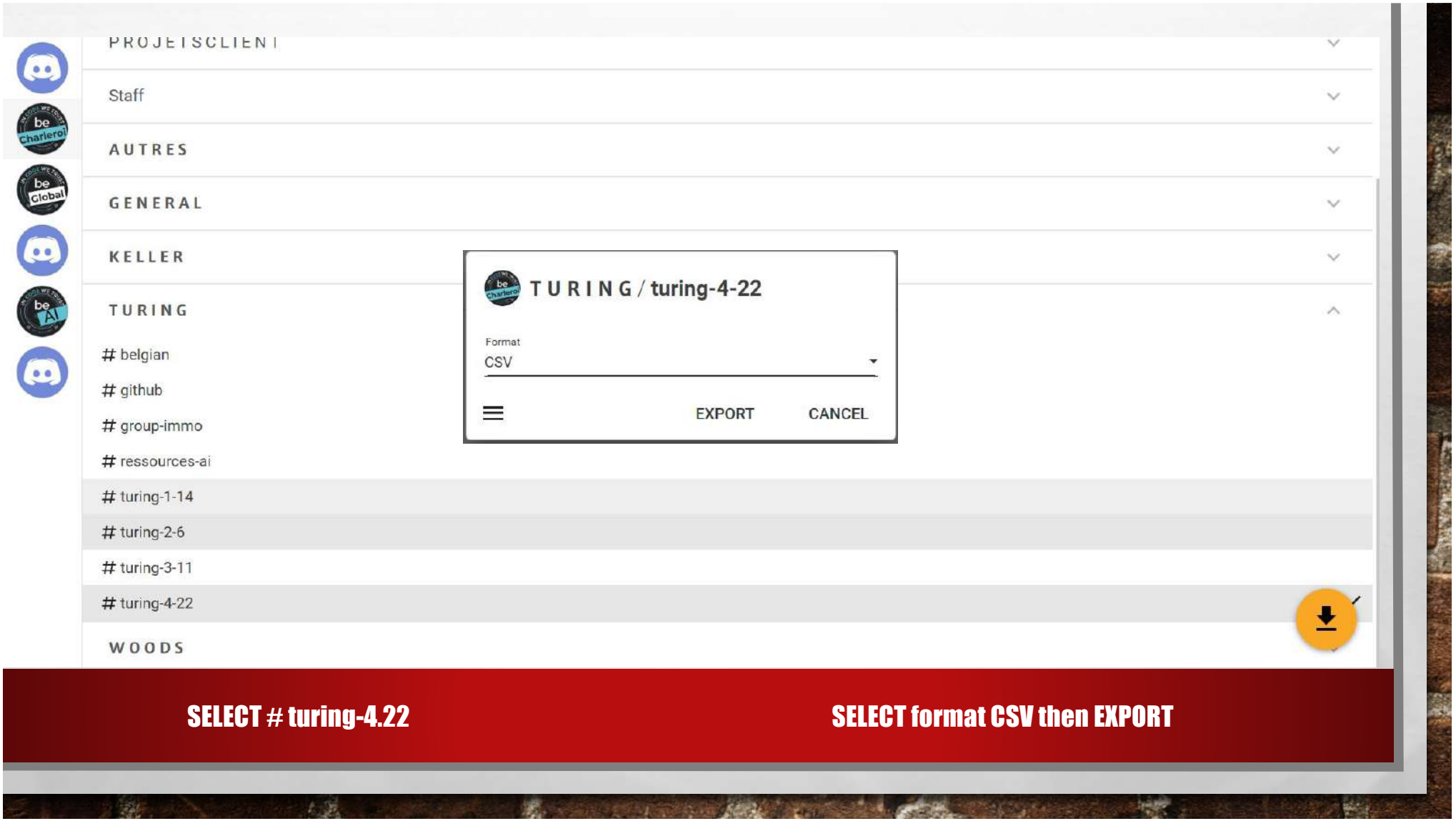
CTR + SHIFT + I

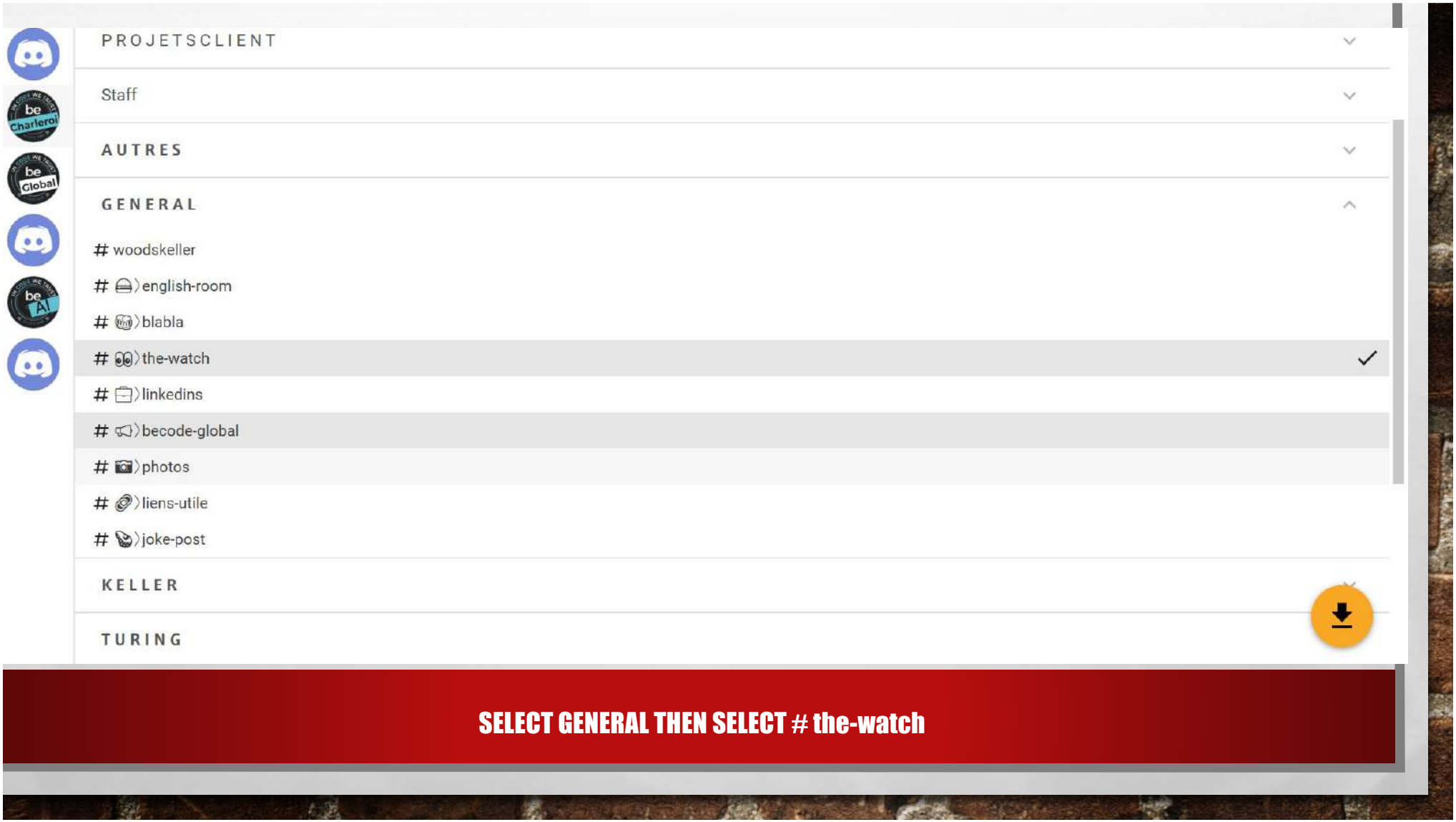


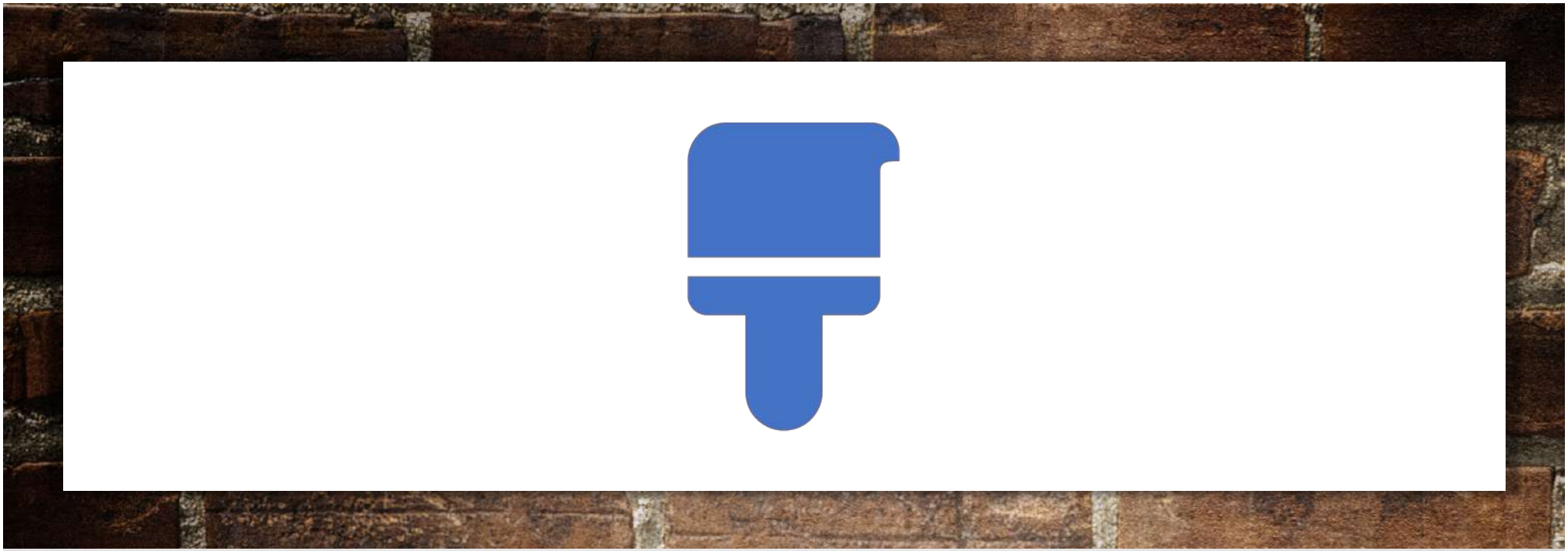
CRT + R **COPY TOKEN**


DiscordChatExporter v2.24.0
Please provide your user token to authorize
1. Open Discord
2. Press Ctrl+Shift+I to show developer tools
3. Navigate to the Application tab
4. Select Local Storage > https://discord.com on the left
5. Press Ctrl+R to reload
6. Find token at the bottom and copy the value
Automating user accounts is technically against TOS, use at your own risk.
To authorize using bot token instead, click in the text box above.
For more information, check out the wiki.
PASTE TOKEN AND CLICK


DiscordChatExporter v2.24.0
AIDE-TECHNO
GROUPES TRAVAUX
Modération
PROJETSCLIENT
Staff
AUTRES
GENERAL
KELLER
TURING
WOODS
SELECT TURING


PROJETSCLIENT
Staff
AUTRES
GENERAL
KELLER
TURING
belgian
github
group-immo
ressources-ai
turing-1-14
turing-2-6
turing-3-11
turing-4-22
WOODS
TURING / turing-4-22
Format
CSV
EXPORT
CANCEL
SELECT # turing-4.22
SELECT format CSV then EXPORT


PROJETSCLIENT
Staff
AUTRES
GENERAL
woodskeller
english-room
blabla
the-watch
linkedins
becode-global
photos
liens-utile
joke-post
KELLER
TURING
SELECT GENERAL THEN SELECT # the-watch



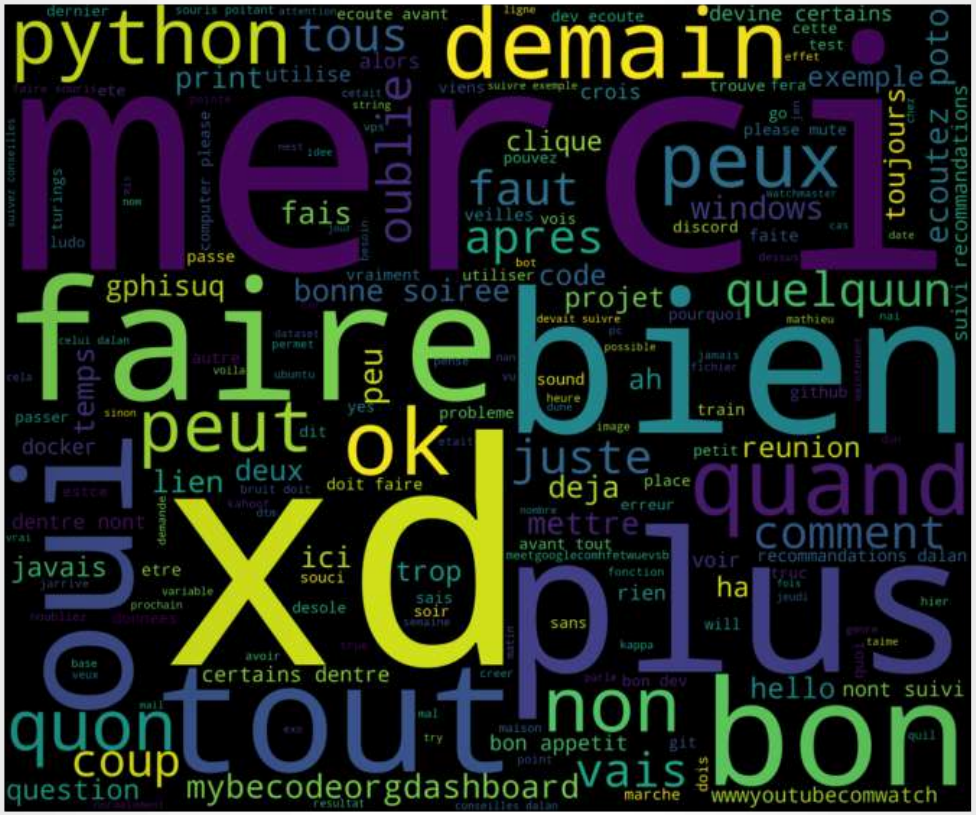
NLTK

DATA CLEANING BECODE_CHARLEROI_TURING_422.CSV

WORDCLOUD

https mybecodeorg
zabzab tamikara
mybecodeorg zabzab
joffrey bienvenu
pointer https
cest moment
lheure pause
mybecodeorg https
google meet
revoit minutes
pause revoit
demarre minutes
minutes google
moment pointer
vincent leurs
deadbool
minutes zabzab
jai
cest
cest lheure
bienvenu zabzab
becode https
veille
plus
faire
bien
turing
christophe meunier
meet cest
alan https
mybecodeorg joffrey
mybecodeorgdashboard
devine certains
docker veille demarre
dalan https
donc
demain
aussi
merci
ca
fait
si
xd
bon
oui
turing https
encore python
lheure
deja
faut
github
ecoutez poto
comment
quand
meet
peux
https gphisuq
bonne soiree
jean christophe
windows
print
ok
peu
temps
kaiyungtan
souris poitant
https wwwyoutubecomwatch
bienvenu tamikara
code clique cest
quelquun
recommandations dalan
meet https
coup
maxime dropsy
train test
souci
quon
avant tout
viens
sais
dalan turing
projetestce
voir
juste
apres
fonction
dev ecoute
minutes joffrey
tout alan
fichier
point
question
donnees
ludo
tamikara cest
javais
bruit doit
poto alan
desole
alan turing
deux
trouve
tamikara zabzab
suivi recommandations
oublie pointer
besoin
creer
doit faire
alors
permet
jarrive
probleme
maison
debrief
ah
ete
tous
dalan turing
jai oublie
utilisernont suivi
dentre nont
certains dentre
sans
sound
comme
oui
reunion demarre
utilise
dois
truc
effet
lien
faire souris
dessus
autre
heure
fera
hier
etait
pense
cetait
pourquoi
va
bon appetit
computer please
pointe toujours
noah alvarez
ha
ici
pon dev
crois
tot
soir
go
yes
nai
ubuntu
vps
quil
deadbool jai
cest bruit
passe
zabzab vincent
erreur
petit
vais
please mute
semaine
passer
true
will
discord alvarez gonzalez
ok
train test
souci
avoir
etre
mettre
quoi
trop
fais
fichier
tout alan
donnees
ludo
tamikara cest
javais
bruit doit
poto alan
desole
alan turing
deux
trouve
tamikara zabzab
suivi recommandations
oublie pointer
besoin
creer
doit faire
alors
permet
jarrive
probleme
maison
debrief
ah
ete
tous
dalan turing
jai oublie
utilisernont suivi
dentre nont
certains dentre
sans
sound
comme
oui
reunion demarre
utilise
dois
truc
effet
lien
faire souris
dessus
autre
heure
fera
hier
etait
pense
cetait
pourquoi
va
bon appetit
computer please
pointe toujours
noah alvarez
ha
ici
pon dev
crois
tot
soir
go
yes
nai
ubuntu
vps
quil
deadbool jai
cest bruit
passe
zabzab vincent
erreur
petit
vais
please mute
semaine
passer
true
will
discord alvarez gonzalez

revoit minutes
moment pointer
faire
plus
peut
vais
comme
veille
becode
bien
meme
non
fait
aussi
merci
tout
encore
oui
va
si
tous
google meet
debrief
pause
revoit
demain
minutes
google
comment
faut
python
quand
demarre
quon
bon
apres
heure
coup
donc
meet
ok
juste
xd
bonne soiree
veille demarre
erreur
utiliser
bon dev
alors
quelquun
certains dentre
devine certains
utilise
train
hello
mybecodeorgdashboard
bon appetit
fais
javais
probleme
dentre nont
sound
temps
lien
deux
questionnont suivi
souris poitant
place
github
test
petit
ici
wwwyoutubecomwatch
please mute
passe
pouvez
avant tout
desole
crois
gphisuq
docker
autre
souci
donnees
jarrive
windows
bruit doit mettre
vraiment
peu
truc
deja
permet
trop
oublie
creer
besoin
fonction
quand meme
image
dois
ete
quoi
viens
discord
projet
attention
ha
trouve
clique
ecoutez poto
etait
rien
point
pointer
code
sans
pense
print
pointez becode
ecoute avant
reunion demarre
toujours
recommandations dalan
suivi recommandations
computer please
yes
exemple
maison
mathieu
turings
pourquoi
vrai
mal
mis
quil
etre
possible
exo
nier
ubuntu
jeudi
car
soir
fichier
base
jen
sous
dessus
avoir
sinon
fera
vps
effet
mal
nest
marche
sound
vu
will
semaine
git
voila
cette
pause minutes
vois
pc
ludo
passer
heure
alors
vu


merci
python
demain
tous
print
utilise
alors
devine
certains
exemple
poto
ecoutez
toujours
suivi recommandations
peux
clique
faut
windows
fais
oublie
apres
faire
bien
gphisuq
bonne soiree
projet
quelquun
code
peu
temps
peut
ok
juste
reunion
quand
lien
deux
deja
comment
mettre
xd
javais
ici
trop
plus
ha
rien
sans
certains dentre
tout
non
bon
hello
nont suivi
quon
coup
question
mybecodeorgdashboard
vais
bon appetit
wwwyoutubecomwatch

THE WATCH

DATA CLEANING AND MANIPULATION WITH PANDAS

THE WATCH

```
df
```

	AuthorID	Author	Date	Content	Attachments	Reactions
0	263999882585571348	Winzar#9656	20-Feb-20 02:17 PM	Veille sur Redis\n\nhttps://docs.google.com/pr...	NaN	NaN
1	321013660707454976	Elenghart#3422	21-Feb-20 01:38 PM	Veille sur le Neuromorphism \nhttps://docs.goo...	NaN	😮 (1)
2	277097161294872576	YlvaAkUlven#1537	26-Feb-20 01:43 PM	https://docs.google.com/presentation/d/1Yt_xUV...	NaN	👍 (3)
3	277097161294872576	YlvaAkUlven#1537	26-Feb-20 01:43 PM	Je n'oublie pas notre cher discord bien sur x)	NaN	NaN
4	621387503491874816	Kapy#6504	27-Feb-20 01:45 PM	https://contattafiles.s3.us-west-1.amazonaws.c...	NaN	🎉 (1)
...	...	...	...	...	...	...
246	324445659228798978	John#7280	05-Nov-20 02:57 PM	https://docs.google.com/presentation/d/13TSyN9...	NaN	💯 (5),👍 (3),Python (3)
247	183999045168005120	Tamikara#4919	06-Nov-20 01:51 PM	https://prezi.com/p/jhsiv2m-le6q	NaN	👍 (3),Python (1)
248	716986685812113478	Stéphane#9140	06-Nov-20 02:06 PM	https://drive.google.com/file/d/15bg9EMF3U9S5P...	NaN	💯 (5),👍 (5)
249	387173536957202435	osimers#5852	10-Nov-20 01:50 PM	https://docs.google.com/presentation/d/16IsRoZ...	NaN	💯 (1)
250	119411980439257088	Karhell#2359	12-Nov-20 01:59 PM	https://docs.google.com/presentation/d/1xf-shr...	NaN	NaN

251 rows × 6 columns

THE WATCH

```
the_watch_link
```

	Author	Date	Content
0	Winzar#9656	20-Feb-20 02:17 PM	Veille sur Redis https://docs.google.com/presentation/d/1axXeuQ4wYlObIMYN0u_epguLL2t6q2unra6r8TV5SNc/edit?usp=sharing
1	Elenghart#3422	21-Feb-20 01:38 PM	Veille sur le Neuromorphism https://docs.google.com/presentation/d/1ZSLXUbMfzy6eR7DC9fs1AiPKhTUMozPKd7yJdqfv79c/edit?usp=sharing
2	YlvaAkUlven#1537	26-Feb-20 01:43 PM	https://docs.google.com/presentation/d/1Yt_xUVeemLv6ERDseyATf_cjWUb0VOhuGyzAcJ6_vRY/edit?usp=sharing
3	Kapy#6504	27-Feb-20 01:45 PM	https://contattafiles.s3.us-west-1.amazonaws.com/tnt14094/QzkY8CTyCrGDIS0/JHipster.pptx jHIPSTER
4	Erawen#2048	04-Mar-20 01:44 PM	https://docs.google.com/presentation/d/1uLZpWzEPd7qwovlaE4zKV0AsYoQTv5RlAUKEtD_7gnQ/edit?usp=sharing
...	...	...	...
150	John#7280	05-Nov-20 02:57 PM	https://docs.google.com/presentation/d/13TSyN9EJkQUTL9Yw5fSzyxDy6Z0uZkR0WpLQ6mE0OPo/edit?usp=sharing
151	Tamikara#4919	06-Nov-20 01:51 PM	https://prezi.com/p/jhsiv2m-le6q
152	Stéphane#9140	06-Nov-20 02:06 PM	https://drive.google.com/file/d/15bg9EMF3U9S5PU3ZqgfYUq7AlA1ds7Uz/view?usp=sharing
153	osimers#5852	10-Nov-20 01:50 PM	https://docs.google.com/presentation/d/16IsRoZIKhNHMjdofJKAeoXy-1G4idzgta9u6P2ffJx0/edit?usp=sharing
154	Karhell#2359	12-Nov-20 01:59 PM	https://docs.google.com/presentation/d/1xf-shrl0qOJRGnHCouw7elyAwfUBs2iqSD6QE2hD7vU/edit?usp=sharing

155 rows × 3 columns

```
turing422 = ['Deadbool#9700','Jean Christophe Meunier#8697', 'Noah Alvarez Gonzalez#7344', 'Tamikara#4919',
             'Zabzab#3333','Maxime Dropsy#8807', 'Joffrey Bienvenu#5302', 'Vincent Leurs#4535', 'kaiyungtan#4047']
```

```
filt = the_watch_link['Author'] == turing422[-1]
kaiyung = the_watch_link[filt]
kaiyung.head()
```

	Author	Date	Content
85	kaiyungtan#4047	01-Sep-20 01:47 PM	https://drive.google.com/file/d/1Iy7cOLhbLLXLRrZuYCs6eNTqieso89r2/view?ts=5f4e349f
97	kaiyungtan#4047	10-Sep-20 02:08 PM	https://drive.google.com/file/d/1A7GoiAe4XqccmX2TyPCsZYun21_JKHda/view?usp=sharing
110	kaiyungtan#4047	18-Sep-20 02:42 PM	https://drive.google.com/file/d/1Xvv3949l2IXwKMCIOQMHdGImgWBg5j-a/view?usp=sharing
123	kaiyungtan#4047	30-Sep-20 01:57 PM	https://drive.google.com/file/d/1SzutUHNYcCfdkDKEBxenTBf5WfsJayVh/view?usp=sharing
138	kaiyungtan#4047	16-Oct-20 03:43 PM	https://drive.google.com/file/d/1b975D6NUG52BBQBeT-40B6tW3lRCQeWE/view?usp=sharing

```
kaiyung.Content
```

```
85     https://drive.google.com/file/d/1Iy7cOLhbLLXLRrZuYCs6eNTqieso89r2/view?ts=5f4e349f
97     https://drive.google.com/file/d/1A7GoiAe4XqccmX2TyPCsZYun21_JKHda/view?usp=sharing
110    https://drive.google.com/file/d/1Xvv3949l2IXwKMCIOQMHdGImgWBg5j-a/view?usp=sharing
123    https://drive.google.com/file/d/1SzutUHNYcCfdkDKEBxenTBf5WfsJayVh/view?usp=sharing
138    https://drive.google.com/file/d/1b975D6NUG52BBQBeT-40B6tW3lRCQeWE/view?usp=sharing
Name: Content, dtype: string
```

THANK YOU