



Star Wars WordClouds

Rafael Schultze-Kraft, Data Engineer @ WATTx

text mining

data wrangling

natural language processing

data visualisation

wordclouds

request, beautifulsoup

NLTK, re, string

matplotlib, scikit-image, wordcloud

 amueller / word_cloud

 Watch ▾

85

 Star

1,343

 Fork

393

 Code

 Issues 22

 Pull requests 4

 Wiki

 Pulse

 Graphs

A little word cloud generator in Python

 181 commits

 19 branches

 1 release

 15 contributors

Branch: master ▾

New pull request

Create new file

Upload files

Find file

Clone or download ▾

 amueller add example images

Latest commit 16797f9 19 days ago

 conda.recipe	Add conda recipe	4 months ago
 continuous_integration	adapt tests for cli to run with travis	6 months ago
 doc	update examples, docs.	a year ago
 examples	add example images	19 days ago
 test	Add test for input argument of generate_from_frequencies().	3 months ago
 wordcloud	Document backward incompatible change.	3 months ago
 .gitignore	add C file for distribution, bump minor version	a year ago
 .travis.yml	use newer numpy, as 1.8.1 has a broken dependency on conda.	8 months ago
 LICENSE	Cleaned it up a little bit.	3 years ago



The web's largest
movie script resource!

www.imsdb.com

INT. MILLENNIUM FALCON - COCKPIT

Inside the cockpit, Chewie lets out a loud howl. Han checks as the ship is buffeted by exploding flak. He appears to be doing six things at once.

HAN

(harried)

I saw them! I saw them!

LEIA

Saw what?

HAN

Star Destroyers, two of them,
coming right at us.

Threepio bumps and bangs his way into the cockpit.

THREEPIO

Sir, sir! Might I suggest...

HAN

(to Leia)

Shut him up or shut him down!

(to Chewie)

Check the deflector shield!

Chewie barks a reply as he readjusts an overhead switch.

HAN

Oh, great. Well, we can still
outmaneuver them.

```
In [48]: # get request for episode V
res = requests.get(episodes.get('Episode V'))
# create soup object
soup = bs4.BeautifulSoup(res.content)

In [49]: # get all bold tags from the HTML
bolds = soup.findAll('b')

In [50]: # define a dictionary to store all characters' dialogues from this episode
characters_5 = defaultdict(str)

In [51]: # iterate over all bold tags
for b in bolds:
    # check if number of indentation tabs equals 4
    if len(b.text) - len(b.text.lstrip()) != 4:
        continue
    # iterate over each line
    for t in b.nextSibling.split('\n'):
        # if line is blank, break
        if not t:
            break
        stripped = t.strip()
        # append to dictionary if line does not start with character '('
        if not stripped.startswith('('):
            characters_5[b.text.strip()] += ' ' + stripped
```

```
In [19]: characters_4.get('VADER')
```

```
Out[19]: u" Where are those transmissions you intercepted? What have you done with those plans? If this is a consular ship...  
were is the Ambassador? Commander, tear this ship apart until you've found those plans and bring me the Ambassador.  
I want her alive! Don't play games with me, Your Highness. You weren't on any mercy mission this time. You passed di-  
rectly through a restricted system. Several transmissions were beamed to this ship by Rebel spies. I want to know wha-  
t happened to the plans they sent you. You're a part of the Rebel Alliance... and a traitor. Take her away! I have tr-  
aced the Rebel spies to her. Now she is my only link to find their secret base! Leave that to me. Send a distress sig-  
nal and then inform the senate that all aboard were killed! She must have hidden the plans in the escape pod. Send a  
detachment down to retrieve them. See to it personally, Commander. There'll be no one to stop us this time. The plan-  
s you refer to will soon be back in our hands. Don't be too proud of this technological terror you've constructed. Th-  
e ability to destroy a planet is insignificant next to the power of the Force. I find your lack of faith disturbing.  
As you wish. And, now Your Highness, we will discuss the location of your hidden Rebel base. Her resistance to the m-  
ind probe is considerable. It will be some time before we can extract any information from her. What do you mean? Com-  
mence primary ignition. I told you she would never consciously betray the Rebellion. They must be trying to return th-  
e stolen plans to the princess. She may yet be of some use to us. Did you find any droids? Send a scanning crew on bo-  
ard. I want every part of this ship checked. I sense something... a presence I haven't felt since... He is here... A  
tremor in the Force. The last time I felt it was in the presence of my old master. Don't underestimate the power of  
the Force. Obi-Wan is here. The Force is with him. Escape is not his plan. I must face him alone. I've been waiting  
for you, Obi-Wan. We meet again, at last. The circle is now complete. When I left you, I was but the learner; now I  
am the master. Your powers are weak, old man. They have just made the jump into hyperspace. This will be a day long  
remembered. It has seen the end of Kenobi and it will soon see the end of the Rebellion. We'll have to destroy them  
ship to ship. Get the crews to their fighters. Several fighters have broken off from the main group. Come with me! S-  
tay in attack formation! I'll take them myself! Cover me! Close up formation. Let him go! Stay on the leader! I'm on  
the leader. The Force is strong with this one! I have you now. What?"
```

```
In [64]: ' '.join(cleaned.get('DARTH VADER'))
```

```
Out[64]: 'separatists care master thank master lord yes master padme safe alive alive impossible transmissions intercepted pla  
ns consular ship ambassador commander tear ship apart found plans bring ambassador alive play games highness mercy mi  
ssion time passed directly restricted system transmissions beamed ship rebel spies happened plans sent rebel alliance  
traitor traced rebel spies link secret base leave send distress signal inform senate aboard killed hidden plans esca  
pe pod send detachment retrieve personally commander stop time plans refer soon hands proud technological terror cons  
tructed ability destroy planet insignificant power force lack faith disturbing wish highness discuss location hidden  
rebel base resistance mind probe considerable time extract information mean commence primary ignition told conscious  
ly betray rebellion trying return stolen plans princess droids send scanning crew board ship checked sense presence t  
remor force time presence master underestimate power force obi-wan force escape plan waiting obi-wan meet circle comp  
lete left learner master powers weak jump hyperspace day remembered seen kenobi soon rebellion destroy ship ship crew  
s fighters fighters broken main stay attack formation cover close formation stay leader leader force strong found con  
sole screen rebels system skywalker set course hoth system veers prepare rebels alerted presence admiral ozzel light  
speed close system clumsy stupid prepare troops surface attack failed time admiral captain piett ready land troops b  
eyond energy shield deploy fleet system command admiral piett yes admiral asteroids concern admiral ship excuses capt  
ain alive ship available sweep asteroid field found yes admiral move ship asteroid field send transmission thy biddin  
g master yes master boy obi-wan help powerful ally join die master substantial reward millennium falcon free methods  
alive disintegrations apology accepted captain needa alert commands calculate destination trajectory fail admiral ho  
nored join captain solo jabba hut skywalker permanently damaged leave city treated unfairly unfortunate leave garriso  
n facility crude adequate freeze skywalker journey emperor monitor skywalker allow land emperor prize damaged test ca  
ptain solo empire compensate dies calrissian survive bounty hunter reset chamber skywalker calrissian princess wookie  
e ship altering deal pray alter force skywalker jedi learned destiny lies skywalker obi-wan true easy strong emperor  
impressive impressive obi-wan taught controlled fear release anger hatred destroy beaten useless resist destroyed ob  
i-wan escape destroy realize importance begun discover power join complete training combined strength destructive con  
flict bring galaxy power dark obi-wan told happened father father search feelings true luke destroy emperor foreseen  
destiny join rule galaxy father son bring shuttle deactivate hyperdrive millennium falcon prepare boarding party set  
weapons stun luke son luke destiny dispense pleasantries commander schedule motivate emperor share optimistic apprai  
sal situation tell arrives correct commander displeased apparent lack progress hope commander sake emperor forgiving  
death star completed schedule yes master wish thy bidding master reports rebel fleet massing near sullust yes master  
shuttle code clearance leave deal force emperor rebel force penetrated shield landed endor master master commander l  
eave conduct search bring companions emperor expecting accepted truth skills complete indeed powerful emperor foresee  
n obi-wan power dark obey master destiny late son emperor true nature force master lightsaber pointless resist son ob  
i-wan taught unwise lower defenses conflict underestimate power dark fight meet destiny hide forever luke dark save f  
riends yes betray feelings strong especially sister twin sister feelings betrayed obi-wan wise hide failure complete  
dark stop look eyes'
```

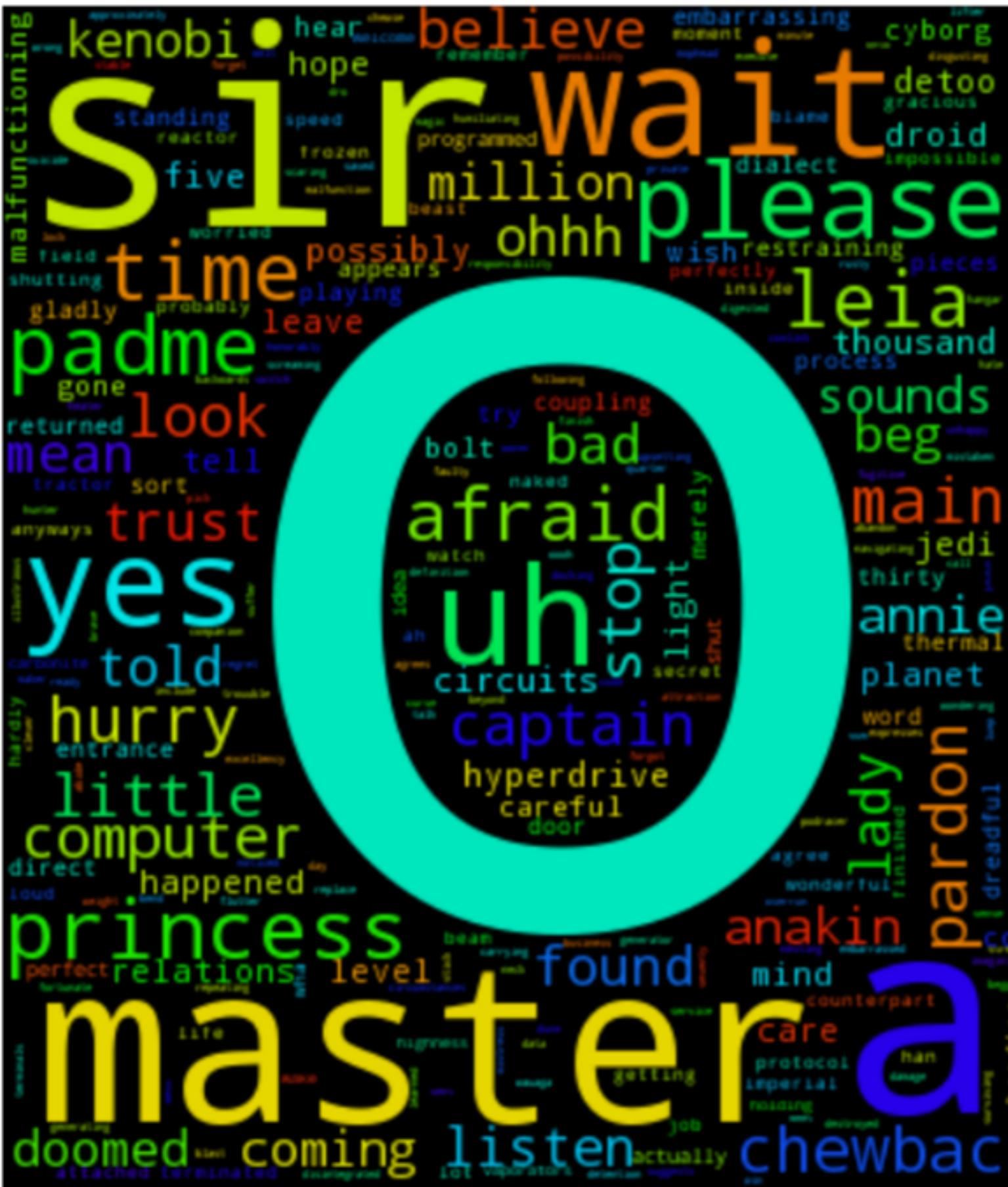
```
In [56]: combined_episodes = merge_entries(combined_episodes, [#('OBI-WAN', 'BEN'),
    ('ANAKIN', 'ANAKN'),
    ('ANAKIN', 'ANAKINN'),
    ('BEN', 'BEN\'S VOICE'),
    ('COUNT DOOKU', 'DOOKU'),
    ('CAPT. PANAKA', 'PANAKA'),
    ('DARTH VADER', 'VADER'),
    ('DARTH SIDIOUS', 'DABTH SIDIOUS'),
    ('HAN', 'HAN\'S VOICE'),
    ('LUKE', 'LUKE\'S VOICE'),
    ('MACE WINDU', 'MACE-WINDU'),
    ('MACE WINDU', 'WINDU'),
    ('OBI-WAN', 'OBI-WAM'),
    ('OBI-WAN', 'OBI-WAN (V.O.)'),
    ('PADME', 'AMIDALA'),
    ('PADME', u'PADM\xc9'),
    ('PADME', 'PAMDE'),
    ('QUI-GON', 'QUI -GON'),
    ('THREEPIO', 'C-3PO')])
```

```
In [57]: sorted(combined_episodes, key=lambda k: len(combined_episodes[k]), reverse=True)[:20]
```

```
Out[57]: [u'ANAKIN',
 u'OBI-WAN',
 u'PADME',
 u'THREEPIO',
 u'LUKE',
 u'HAN',
 u'PALPATINE',
 u'YODA',
 u'QUI-GON',
 u'LEIA',
 u'BEN',
 u'DARTH VADER',
 u'JAR JAR',
 u'LANDO',
 u'CAPT. PANAKA',
 u'COUNT DOOKU',
 u'WATTO',
 u'BAIL ORGANA',
 u'NUTE',
 u'EMPEROR']
```

```
    frequencies = Counter(text.split(' '))
    wordcloud = WordCloud(relative_scaling=.5, width=400, height=300,
                          max_words=500).generate_from_frequencies(frequencies.items())
    pl.figure(figsize=(15, 15))
    pl.imshow(wordcloud, cmap=pl.cm.gray)
    pl.axis('off')
    pl.title(k, fontsize=20)
    pl.show()
```


A dense word cloud collage centered around Star Wars characters and themes. The words are arranged in large, bold letters that form the names of various characters: Obi-Wan, dack, little, Luke, Leia, Vader, Han, droid, art, tell, wedge, hang, shut, Jedi, yoda, comin, oh, chewie, looking, course, kenobi, talking, system, broken, and ground. These central words are surrounded by numerous smaller words in various colors (black, white, blue, green, red, orange) representing different concepts and terms from the Star Wars universe. In the top right corner, there is a small rectangular image of a young Mark Hamill as Luke Skywalker, holding a lightsaber.

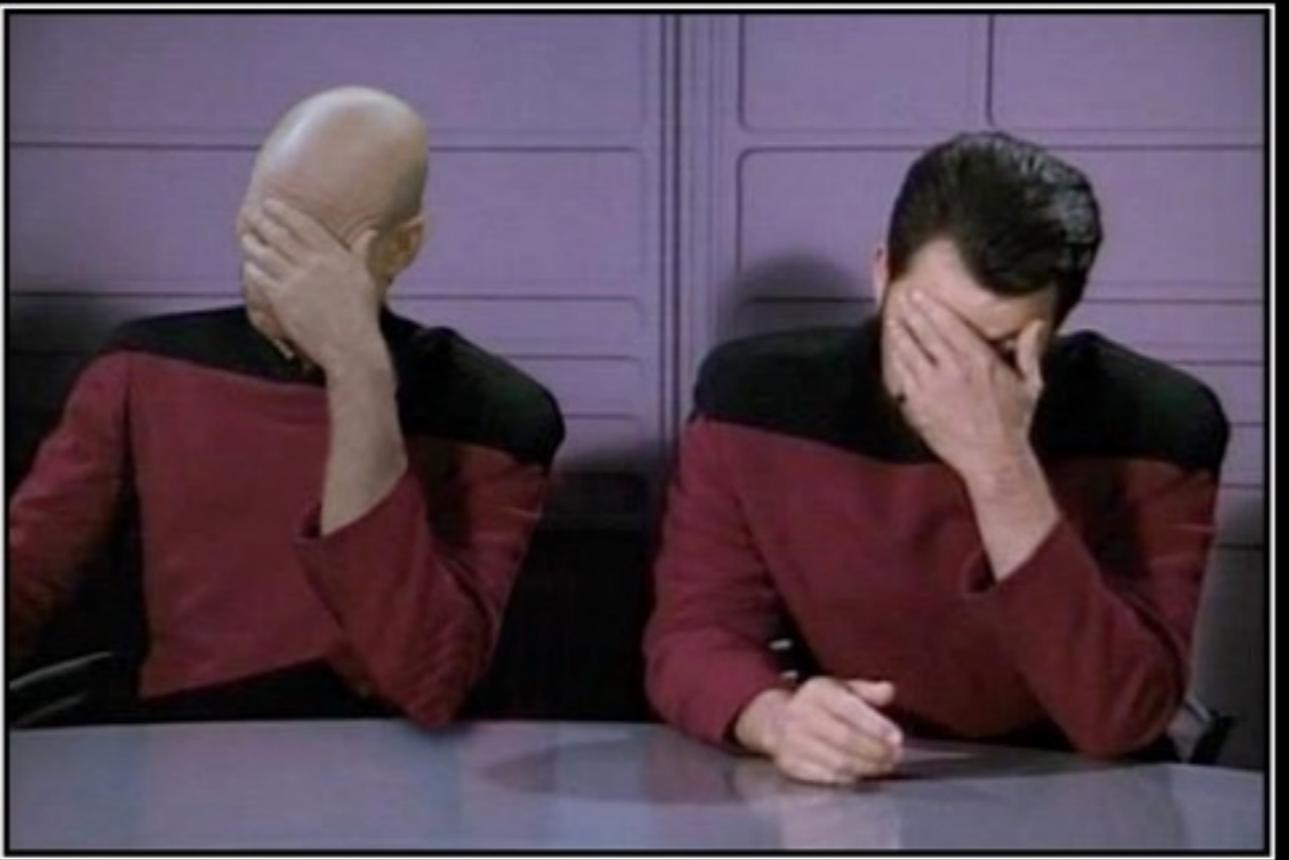




A dense word cloud collage centered around the Star Wars universe. The words are arranged in large, bold letters, with smaller related words floating around them. Key characters include Han Solo, Chewie, Leia, Luke, Obi-Wan Kenobi, and Yoda. Other words represent locations like Alderaan, Tatooine, and Bespin, as well as concepts like hyperspace,光劍 (lightsaber), and原力 (the Force). The overall theme is the rich vocabulary of the Star Wars franchise.

A dense collage of words and phrases from the Star Wars universe, arranged in a grid-like pattern. The words are in various colors and sizes, some appearing multiple times. In the top right corner, there is a black and white photograph of Han Solo from the original Star Wars film, standing with his arms outstretched and holding a blaster pistol.





DOUBLE FACEPALM

FOR WHEN ONE FACEPALM DOESN'T CUT IT









companions
importance
journey
city
intercepted
clearance
look
traced
pray
veers
plan
highness
pleasantries
crude
release
traitor
sake
dies
nature
energy
stun
failure
separatists
substantial
checked
permanently
retrieve
primary
signal
endor
defenses
luke
command
join
begin solo
bidding
excuses
course
safe
survive
bring
extract
save return
rebels
schedule
princess
displeased
soon
commands
close
hidden
thy
jedi
refer
killed

Shi
captain
plan
force
city
release
sister
transmission
conduct
waiting
beyond
weapons
jump
hunter
link
optimistic
learner
garrison
landed
treated
clumsy
treated
discuss
arrives
board
situation
chamber
detachment
discover
reports
cover
millennium
astroids
correct
day
late
jabba
failed
field
expecting
freeze
impressive
faith
true
apart
senate
mercy
boy
twin
light
calculate
appraisal
powerful
star
fight
screen
hope
circle
main
thank
reward
progress
code

plans
meet
highness
cress
obey
sister
conduct
wait
beyond
nature
separatists
substantial
checked
useless
directly
test
mind
massing
care
mission
attack
leader
help
mean
planet
fleet
stop
location
pointless
facility
restricted
damaged
speed
prize
learning
accepted
apparent
probe
tear
completed
tear
fear
empire
anger
indeed
fear
calrissian
stolen

admiral
terr
share
adequate
seen
hide
troops
ally
beaten
realize
altering
information
sent
alliance
piett
base
penetrated
especially
taught
beam
father
lack
hut
time
conflict
droids
ability
destination
shield
destroyed
passed
set
hyperspace
alter
boarding
apart
senate
mercy
boy
twin
light
calculate
appraisal
powerful
star
fight
screen
hope
circle
main
thank
reward
progress
code

master
emperor
commander
complete
resist
asteroid
skywalker
admiral
leave
captain
Shi
force
city
release
sister
transmission
conduct
waiting
beyond
weapons
jump
hunter
link
optimistic
learner
garrison
landed
treated
clumsy
treated
discuss
arrives
board
situation
chamber
detachment
discover
reports
cover
millennium
astroids
correct
day
late
jabba
failed
field
expecting
freeze
impressive
faith
true
apart
senate
mercy
boy
twin
light
calculate
appraisal
powerful
star
fight
screen
hope
circle
main
thank
reward
progress
code

underestimate
formation
hyperdrive
strong
hot
found
monitor

what's next?

- use TF-IDF
- apply machine learning
- network graph analysis
-

code:

<https://github.com/neocortex/mining-star-wars>

“Gracias!”



<https://github.com/neocortex/mining-star-wars>