

Wall Street vs. Wall Street Bets

What Reddit has to say about investing and financial instruments.



••••••••••••••

Dataset

Subreddit data from r/wallstreetbets and more!

Submissions from r/wallstreetbets, r/investing, r/options, r/SecurityAnalysis

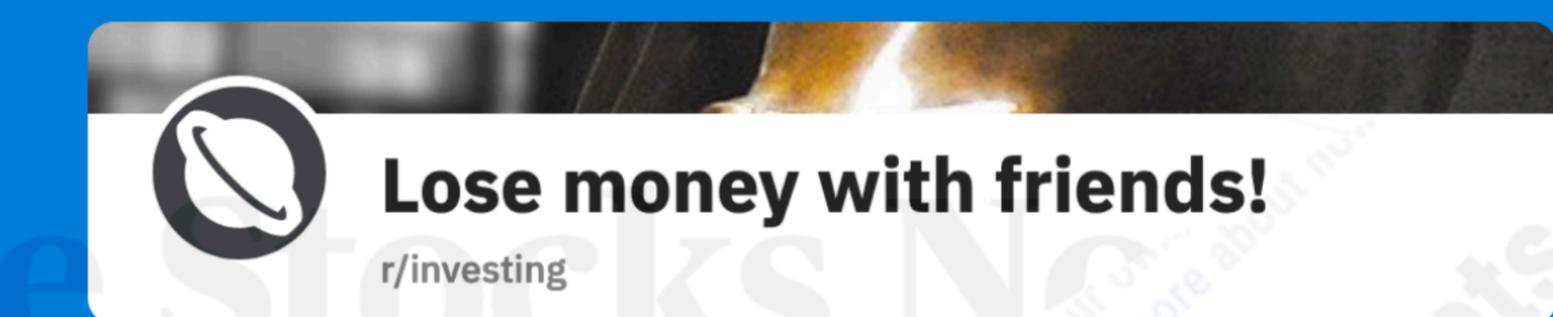
wallstreetbets JOINED

Sheridan Green • updated 3 months ago (Version 1)

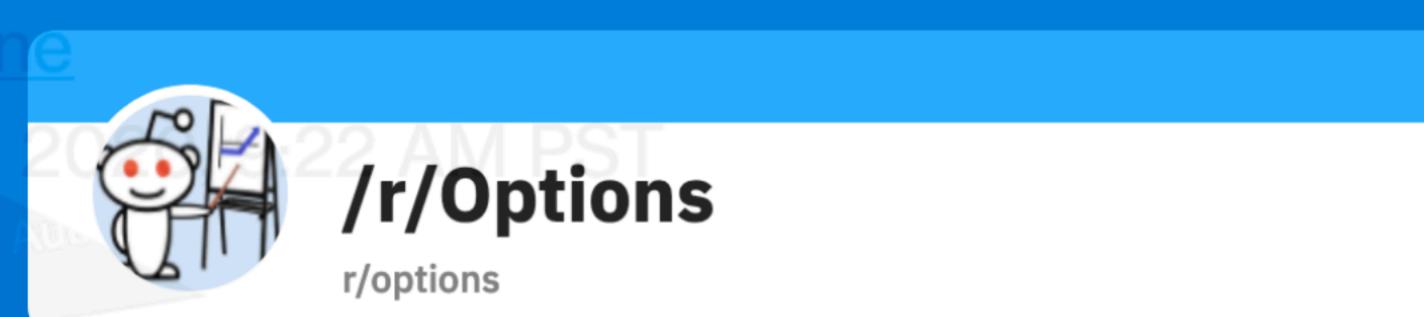
Posts FAQ WSB Discord Daily Discussion Filter By Flair No-Meme Mode Remote Login About Community



1.6M Degenerates
454,025 documents



1.2M Members
210,151 documents

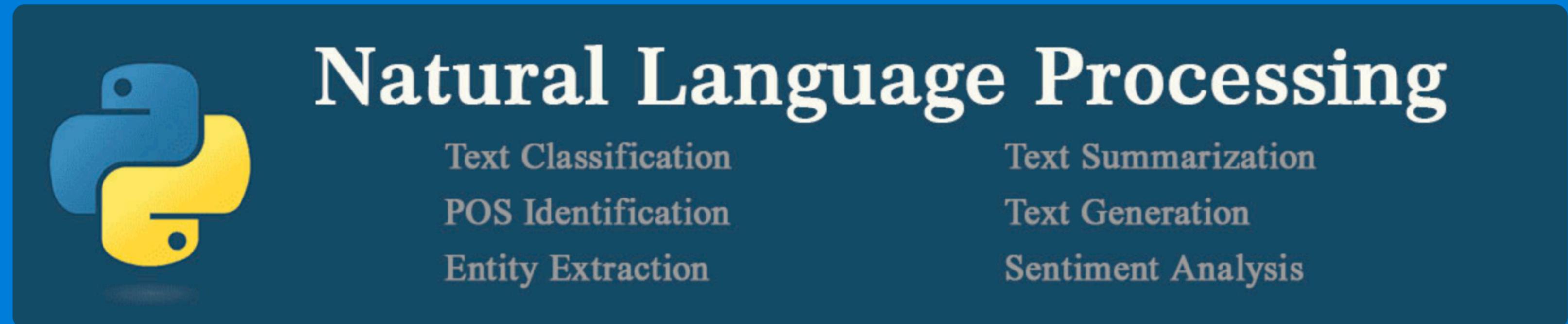


357k Members
42,385 documents



100k Security Analysts
22,939 documents





Natural Language Processing

Python logo

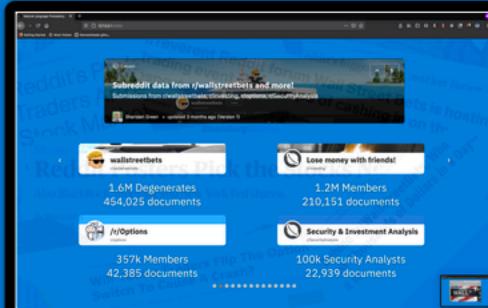
- Text Classification
- POS Identification
- Entity Extraction
- Text Summarization
- Text Generation
- Sentiment Analysis

Latent Semantic
Analysis

Latent Dirichlet
Allocation

Non-Negative Matrix
Factorization

Corex



LSA topics

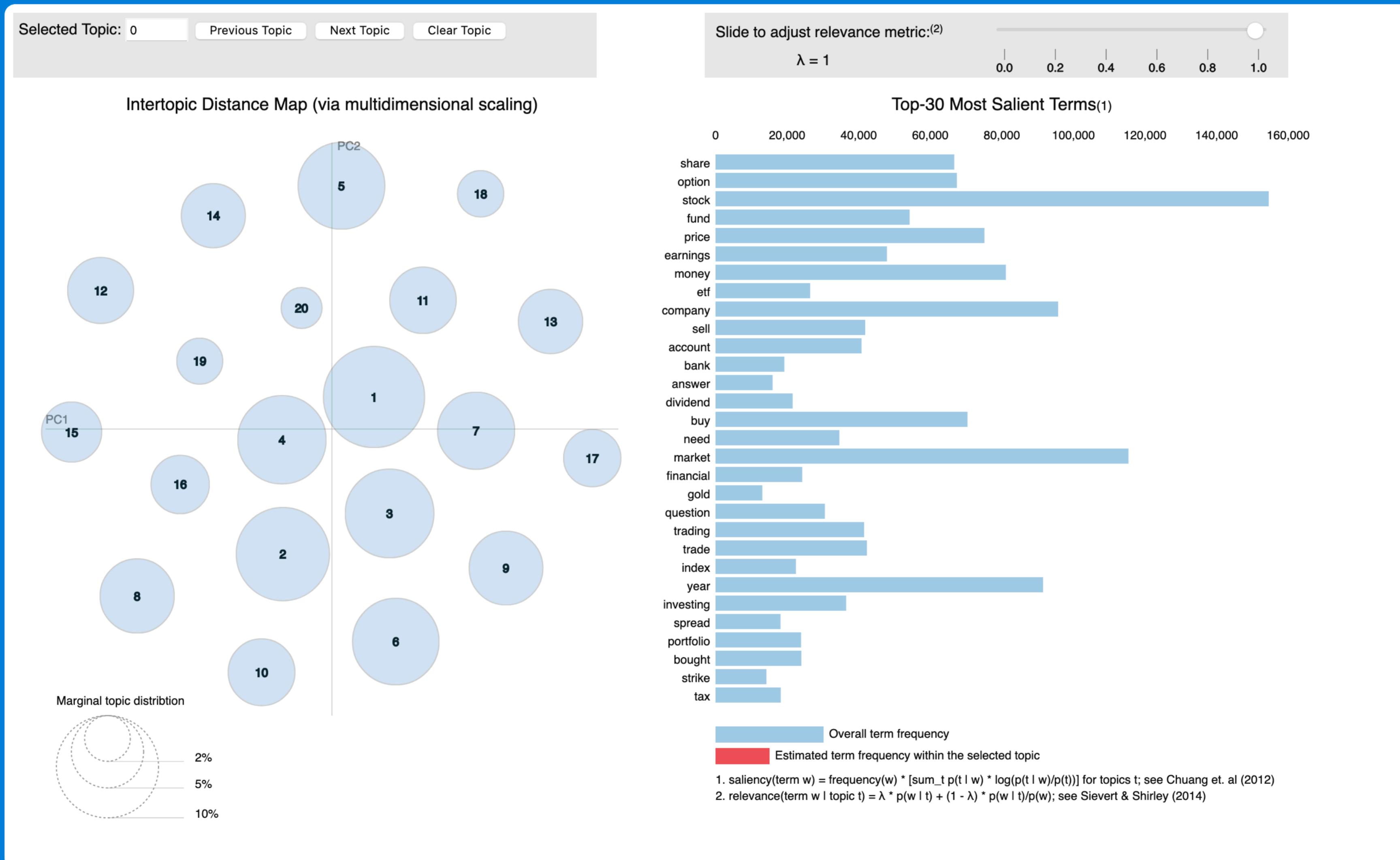
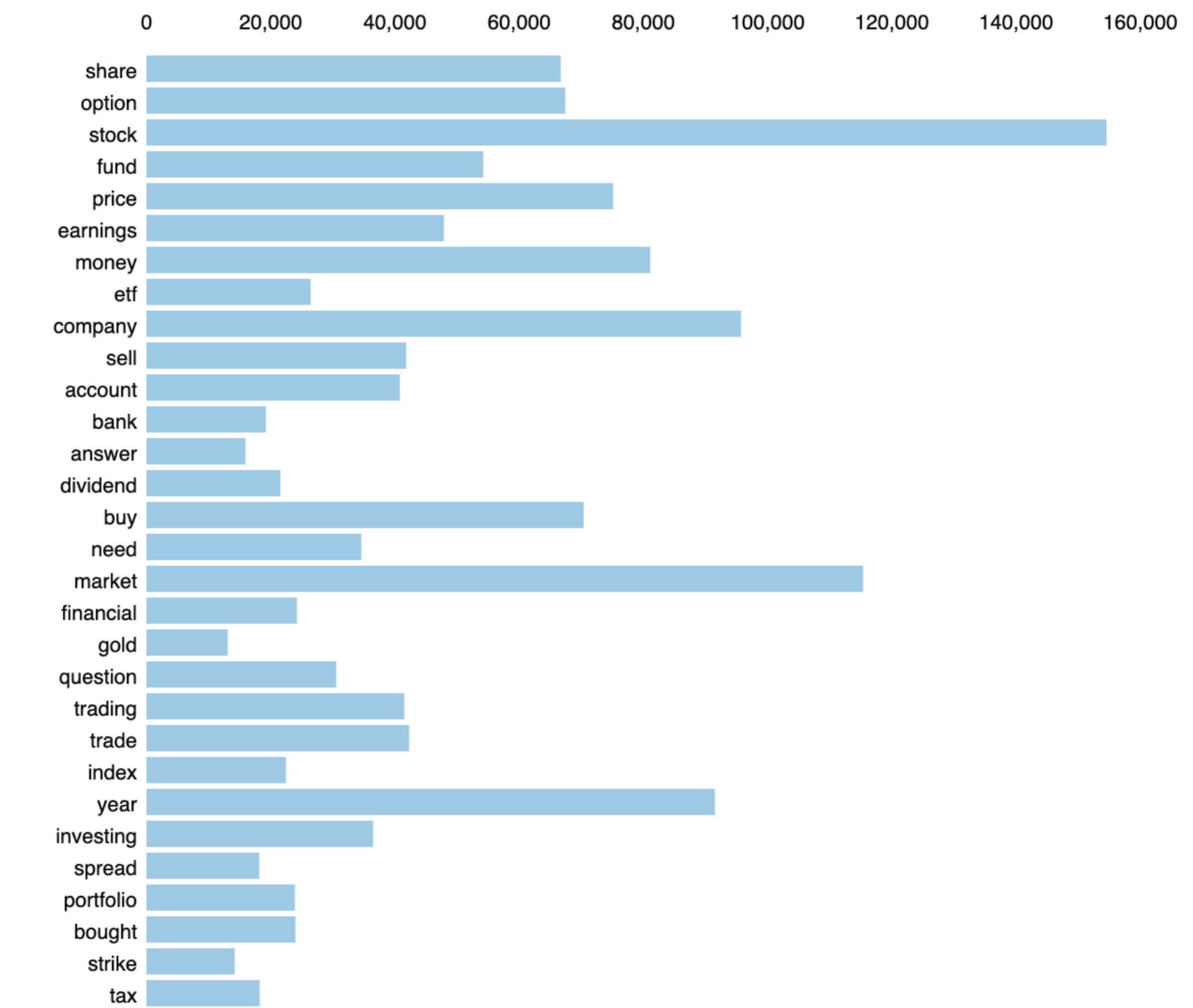
gt	earnings	stock	ha	market	company
***	america	nt	trade	post	make
gt	earnings	click	release	estimate	chart
company	usd	ha	bln	rev	eps
option	call	usd	rev	price	trade
stock	click	share	usd	gt	company
usd	exp	bln	gt	nt	vs
rev	beat	inline	expected	last	tgt
***	company	nt	share	rev	wa
***	amet	wa	wed	may	oct
market	***	stock	earnings	fund	usd
***	gt	option	price	fund	financial
***	earnings	stock	fund	money	amet
xb	answer	money	need	buy	financial
xb	nt	option	fund	gt	like

NMF topics

earnings	ha	release	estimate	company	stock
***	america	post	trade	make	***
nt	like	know	going	people	think
company	ha	business	million	product	revenue
stock	click	today	dividend	live	symbol
option	trade	trading	spread	price	value
usd	bln	exp	vs	revenue	eps
rev	beat	inline	expected	consensus	report
***	***	***	going	na	buy
amet	oct	wed	thurs	sept	earnings
market	ha	time	investor	fed	otc
gt	click	chart	said	share	today
answer	financial	need	question	house	information
wa	day	time	week	trade	today
fund	index	etf	bond	vanguard	investment



pyLDAvis - mds="tsne"

Top-30 Most Salient Terms⁽¹⁾

Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

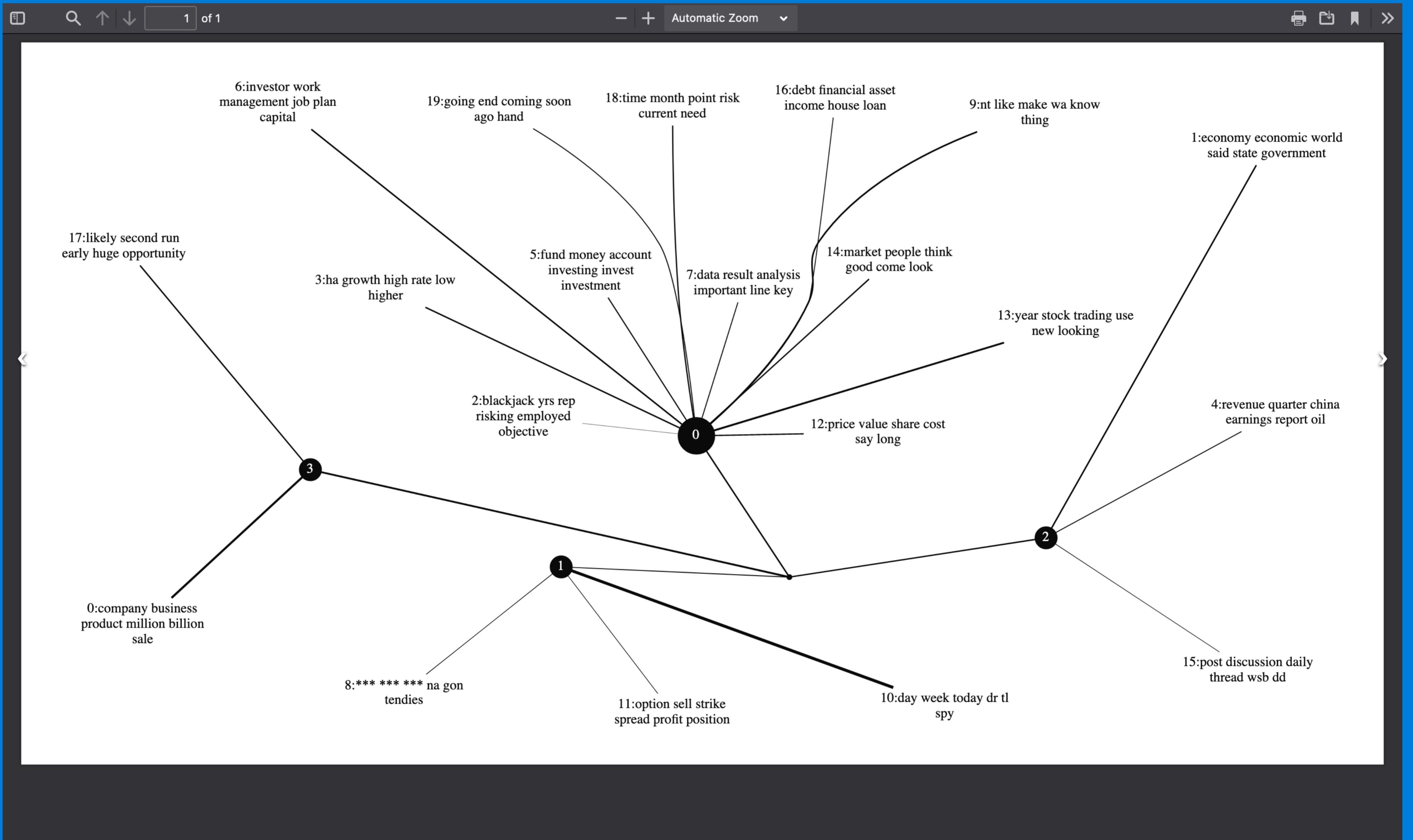
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

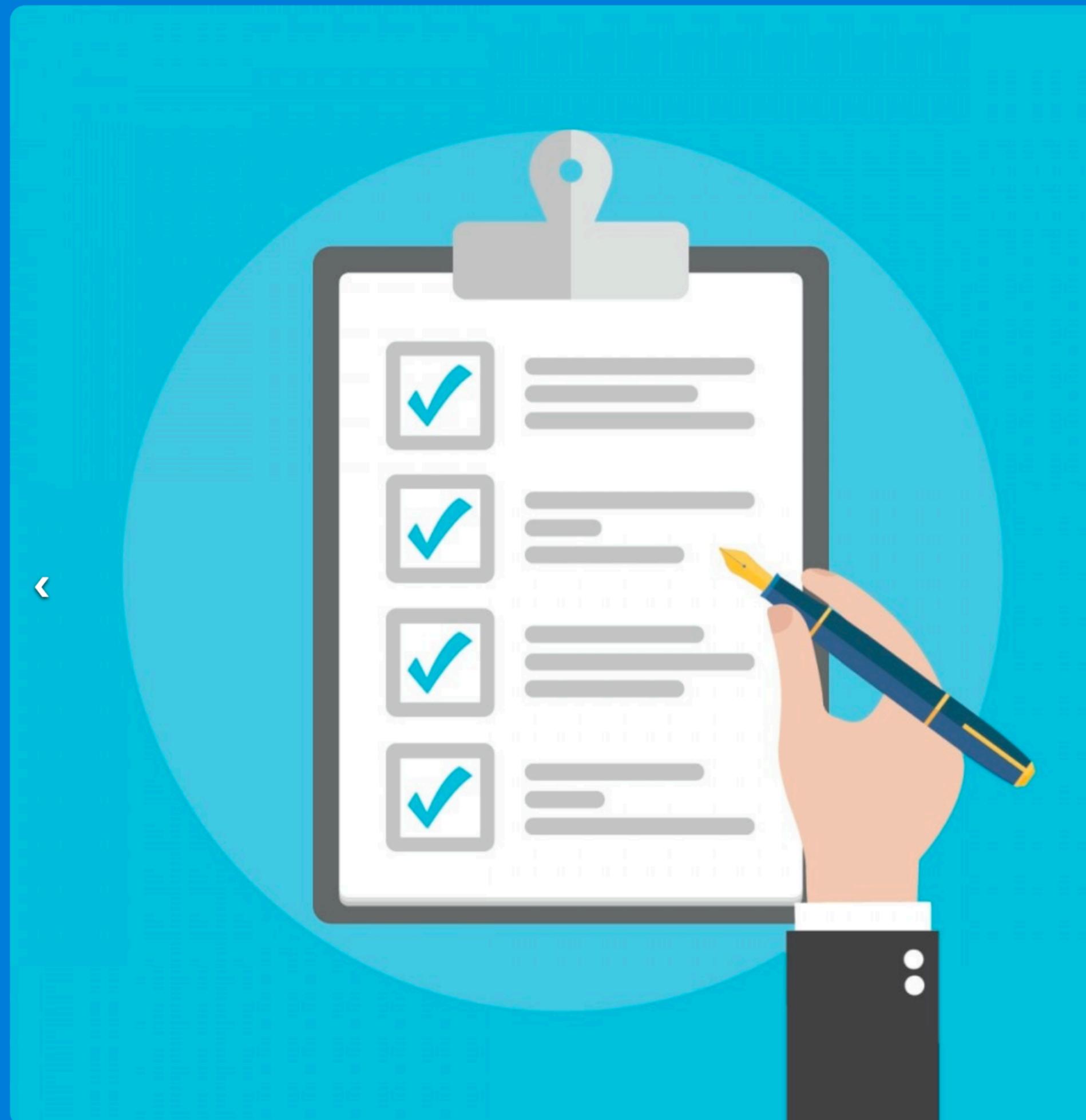
Corex using *options, tendies, yolo, fomo, stock, stonk, etf, bond, trump, hillary, god, politics*

trade	data	using	spread	based	example
tendies	***	***	***	na	gon
day	sell	week	position	loss	profit
price	rate	higher	increase	value	likely
stock	penny	picking	weed	picker	delisted
end	th	march	news	hit	coming
fund	etf	year	investment	income	retirement
blackjack	yrs	risking	rep	employed	objective
trump	china	economy	said	country	world
second	deal	control	power	hand	ability
people	come	fact	life	street	wall
company	business	investor	product	growth	million
cash	balance	bank	equity	loan	flow
real	estate	new	property	great	area
market	going	high	thing	long	right

Corex without anchors

company	business	product	million	billion	sale
economy	economic	world	said	state	government
blackjack	yrs	rep	risking	employed	objective
ha	growth	high	rate	low	higher
revenue	quarter	china	earnings	report	oil
fund	money	account	investing	invest	investment
investor	work	management	job	plan	capital
data	result	analysis	important	line	key
***	***	***	na	gon	tendies
nt	like	make	wa	know	thing
day	week	today	dr	tl	spy
option	sell	strike	spread	profit	position
price	value	share	cost	say	long
year	stock	trading	use	new	looking
market	people	think	good	come	look





Take aways:

- The different techniques drew similar conclusions.
- Seeing how much similarity there was between CorexTopic's unsupervised topic generation and that of my anchor selections gave me a lot of confidence in CorexTopic as a foundation for further exploration.

Next steps:

- Identify stocks.
- Add sentiment analysis.
- Expand resolution of tool by adding time component.
- Correlate to stock's performance.



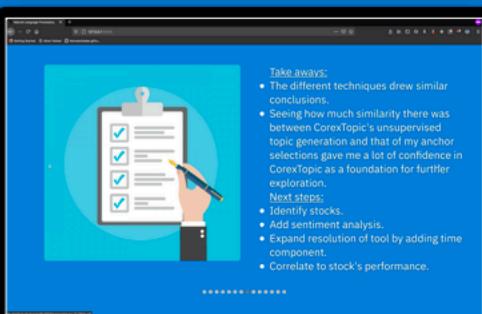


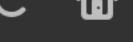
127.0.0.1:5000



Getting Started Most Visited Monsterblader.github...

Questions?



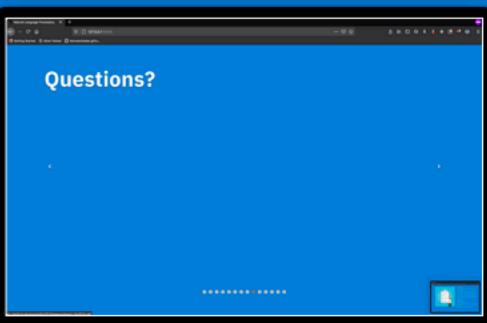


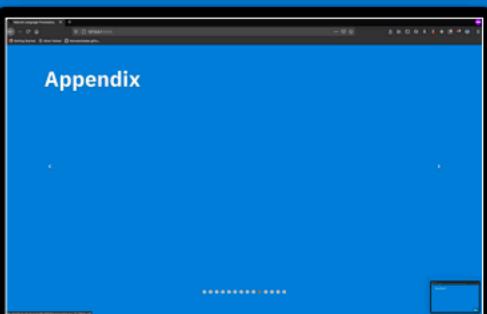
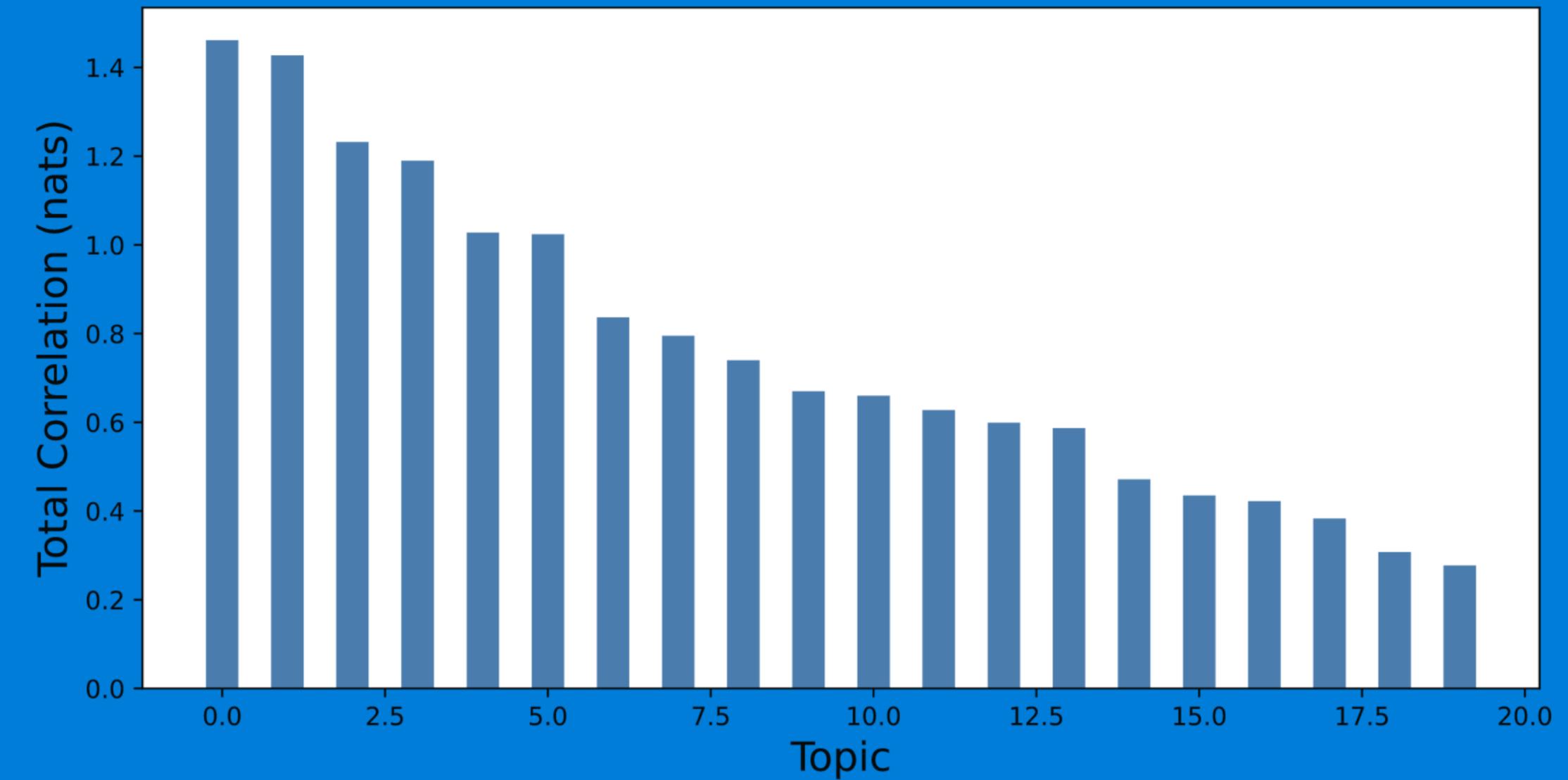
127.0.0.1:5000



Getting Started Most Visited Monsterblader.github...

Appendix

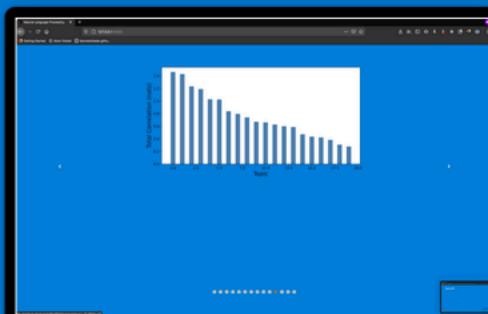




```
In [20]: n_hidden = 40
anchor_word_list = [['options'], ['tendies'], ['yolo'], ['fomo'], ['stock'], ['stonk'], ['etf'], ['bond'], ['tr

for topic in get_topic_words(anchor_word_list, cor_words, ct, cor_posts, sanitized_posts_df, n_hidden):
    print(topic)

0: data, value, based, analysis, important, security, model, provide, performance, link
1: tendies, shit, fucking, fuck, na, gon, dr, tl, bear, retard
2: day, week, today, position, close, sold, bought, loss, friday, play
3: blackjack, yrs, risking, rep, employed, objective, registered, redditors, tolerance, proper
4: stock, nan, penny, picking, picker, poll, vest, dividendpaying, nondivididend, twit
5: revenue, earnings, growth, quarter, sale, expected, report, increase, number, result
6: fund, etf, account, investment, retirement, ira, saving, risk, index, vanguard
7: bond, debt, asset, income, rate, financial, cash, equity, yield, loan
8: trump, china, economy, world, country, state, government, economic, global, demand
9: corp, nasdaq, ba, morgan, jpm, express, pharmaceutical, djia, jnj, therapeutic
10: people, come, fact, line, street, life, wall, seen, support, story
11: business, product, service, industry, technology, management, customer, including, ceo, group
12: post, discussion, daily, thread, weekly, wsb, dd, discord, chat, comment
13: nt, like, thing, think, good, right, ca, doe, feel, little
14: way, actually, different, able, problem, course, simply, basically, simple, trader
15: company, ha, share, billion, million, potential, shareholder, deal, major, tech
16: totaling, nickel, sufficiently, monitored, sustaining, coincide, freeing, stub, generously, isle
17: going, future, believe, past, coming, hit, reason, drop, news, bad
18: wa, ago, couple, dollar, thought, came, left, half, finally, knew
19: said, told, hand, seven, planned, spoke, dozen, fresh, potato, suit
20: point, current, big, note, significant, following, related, member, additional, roughly
21: yes, watch, party, disclaimer, changed, dry, notice, stopped, mm, leaf
22: money, make, time, know, want, need, really, question, month, buy
23: year, great, grow, set, main, considering, pick, adding, keeping, assumption
24: option, price, sell, strike, spread, long, profit, selling, short, trade
25: order, protect, cancel, purchasing, involve, exceeds, gtc, infusion, glaring, overlook
26: likely, given, period, event, lead, benefit, activity, concern, addition, larger
27: second, final, perfect, special, count, auto, covering, fifth, containing, spectrum
28: market, cap, small, large, emerging, overall, mid, relatively, developed, core
29: investor, opportunity, especially, fundamental, highly, extremely, environment, relative, view, mentioned
30: end, early, beginning, moving, xb, reach, added, greater, essentially, depending
31: case, issue, gt, power, reduce, protection, mention, affect, tested, creates
32: gold, silver, metal, mining, precious, commodity, miner, gld, physical, copper
33: rough, tradition, patch, occupied, fig, attainable, drafted, draining, eighteen, alarmed
34: investing, thanks, help, new, start, use, work, looking, trading, advice
35: cost, basis, margin, possible, cover, transaction, figure, profitable, condition, discount
36: rely, multiasset, longshot, migrated, monica, permabears, skewing, prejudice, assembling, threeday
37: lt, pan, dave, stamp, thirty, furious, buster, ramsey, src, blatantly
38: mean, le, look, change, far, wrong, instead, matter, turn, clear
39: real, estate, home, area, general, building, hope, add, trust, quite
time: 4min 20s
```



```
In [17]: anchor_word_list = [['aapl'], ['tsla'], ['amzn'], ['spy'], ['qqq'], ['bynd'], ['nio'], ['nkla']]  
  
for topic in get_topic_words(anchor_word_list, cor_words, ct, cor_posts, sanitized_posts_df):  
    print(topic)  
  
0: investor, data, number, based, management, analysis, important, change, security, issue  
1: world, street, wall, food, care, control, face, city, history, power  
2: company, business, revenue, product, sale, service, industry, technology, consumer, customer  
3: spy, na, gon, shit, day, fucking, week, fuck, tl, dr  
4: value, high, rate, growth, low, higher, cost, term, increase, average  
5: said, million, economy, billion, economic, result, quarter, state, gt, global  
6: year, debt, income, saving, financial, house, old, asset, need, pay  
7: investment, thanks, help, use, new, start, looking, work, advice, currently  
8: nt, like, make, know, want, thing, way, good, really, little  
9: time, point, month, risk, current, big, lot, note, following, gold  
10: chip, amd, blue, intel, nvda, nvidia, mu, intc, semiconductor, micron  
11: wa, post, look, got, ago, went, great, far, away, took  
12: likely, end, case, fact, continue, potential, line, major, second, early  
13: discord, weekly, discussion, daily, chat, thread, dd, wsb, shitposting, minimum  
14: market, right, come, believe, mean, reason, probably, drop, bad, crash  
15: price, sell, option, buy, share, stock, profit, strike, trade, position  
16: fund, account, money, retirement, ira, vanguard, index, invest, investing, etf  
17: blackjack, yrs, rep, risking, employed, objective, registered, redditors, proper, tolerance  
18: china, oil, country, production, energy, american, chinese, gas, america, europe  
19: ha, people, going, think, past, earnings, coming, huge, run, seen  
time: 2min 10s
```



