# Homework 11

**Name: Catherine Squillante**

**PID: cat1997**

# What are the trending topics on twitter by week

I found what people were talking about each week by doing a TD_IDF analysis of all the words in tweets for each week and analyzing and drawing conclusions based on the top weighted ones.

In [195]:

```
import pandas
data = pandas.read_csv('VT_tweets_2019_geo.csv')
```

Create bags and remove unnecessary words:

In [196]:

```
data['bag'] = data.tweet.map(lambda t: t.replace(',', '').lower().split())
```

Format date-time so it's a tupule containing month, day. Create intervals of one week long and put all words from tweets from that week into a single bag

In [197]:

```
data.datetime = data.datetime.apply(lambda x: (x.split(' ')[0].split('-')[1], x.spl
it(' ')[0].split('-')[2]))
```

In [198]:

```
row_1 = ('02', '11')
bag1 = []
bags = []
dates = [('02', '11')]
for i in range(len(data)):
    date = data.datetime[i]
    bag = data.bag[i]
    if date[0] == row_1[0] and int(date[1])< int(row_1[1]) + 7:
        bag1.append(bag)
    else:
        dates.append(date)
        bags.append(bag1)
        bag1 = []
        row_1 = date
#process the last week missed by the loop
for i in range(len(data)):
    date = data.datetime[i]
    bag = data.bag[i]
    if date == row_1:
        bag1.append(bag)
bags.append(bag1)

#put all of the lists of words into one bnig list for each week
final_bags = []
for bag in bags:
    flat_bag = []
    for listt in bag:
        for word in listt:
            flat_bag.append(word)
    final_bags.append(flat_bag)
```

Create dataframe of the data aggregated by week

In [199]:

```
days = list(set(list(data.datetime)))
days_frame = pandas.DataFrame(index = dates)
days_frame['bag'] = final_bags
days_frame = days_frame.sort_index()
```

Calculate frequncy and IDF to create TF_IDF table

In [200]:

```
pandas.Series(days_frame.bag[0]).value_counts()
TF=days_frame.bag.apply(lambda x: pandas.Series(x).value_counts())
```

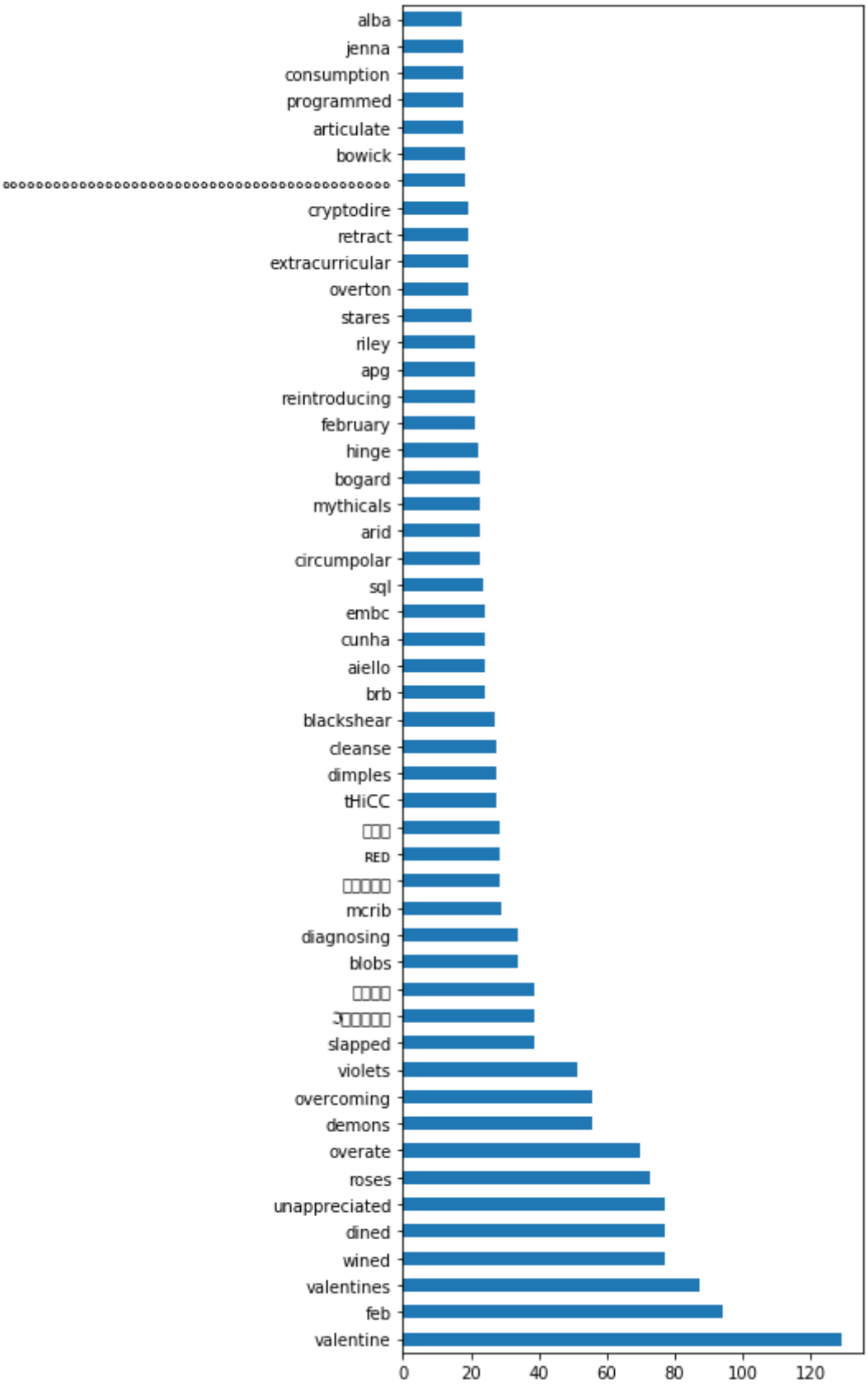
In [201]:

```
import numpy
IDF = numpy.log(len(TF)/TF.count())
TF_IDF = TF * IDF
```

Analyze the TF_IDF results for each week. Filter out words with http or @names or emojis

In [266]:

```
a = TF_IDF.iloc[0]
one = a.where(a>0).dropna().sort_values(ascending = False)
filt = one.index.str.contains('http|@', regex = True)#filter out links and @names
one = one[~filt]
filt1 = one.index.str.isalpha()
one = one[filt1]
d = one[:50]
chart = d.plot.barh(figsize=(5,15), sharex=True)
```


alba
jenna
consumption
programmed
articulate
bowick
cryptodire
retract
extracurricular
overton
stares
riley
apg
reintroducing
february
hinge
bogard
mythicals
arid
circumpolar
sql
embc
cunha
aiello
brb
blackshear
cleanse
dimples
tHiCC
RED
mcrib
diagnosing
blobs
slapped
violets
overcoming
demons
overate
roses
unappreciated
dined
wined
valentines
feb
valentine
0
20
40
60
80
100
120

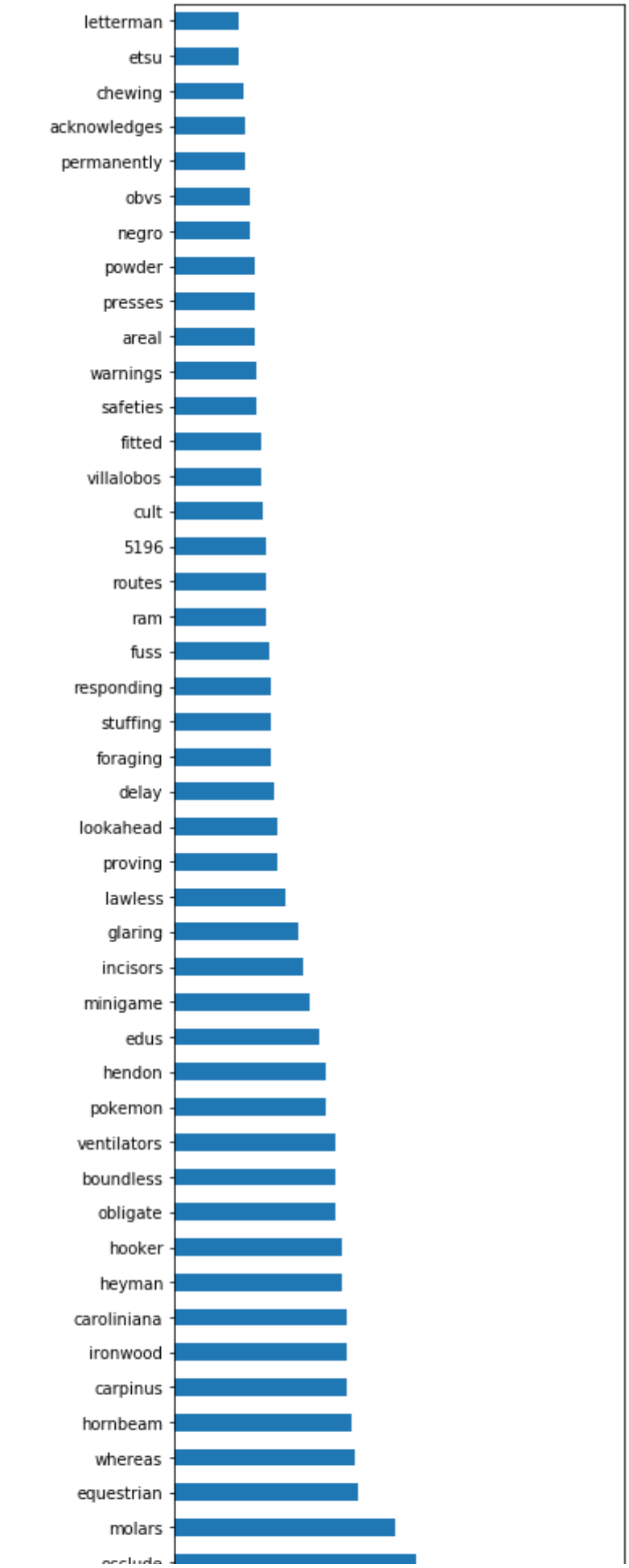
The popular words from the first week indicate that valentines day occured during this time from words such as 'feb, valuentines, roses' etc. Mc rib was also popular so that might indicate that that's when McDonald's brought back the mcRib for a limited time.

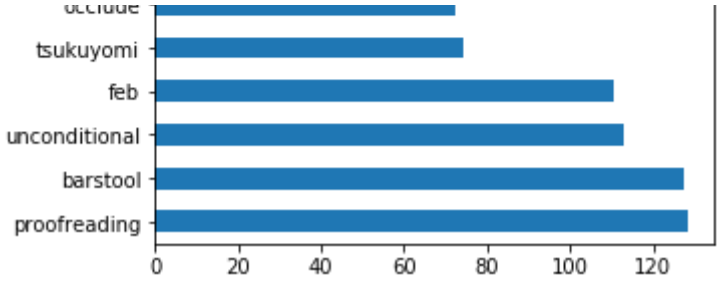
In [267]:

```
a = TF_IDF.iloc[1]
one = a.where(a>0).dropna().sort_values(ascending = False)
filt = one.index.str.contains('http|@', regex = True)#filter out links and @names
one = one[~filt]
filt1 = one.index.str.isalnum()
one = one[filt1]
d = one[:50]
chart = d.plot.barh(figsize=(5,20), sharex=True)
d
```

Out[267]:

```
proofreading     128.280702
barstool         127.364412
unconditional    112.703760
feb              110.338335
tsukuyomi         74.034144
occlude           72.386968
molars            65.986954
equestrian        54.720889
whereas           54.061153
hornbeam          53.111451
carpinus          51.502013
ironwood          51.502013
caroliniana       51.502013
heyman            49.892575
hooker            49.892575
obligate          48.283137
boundless         48.283137
ventilators       48.283137
pokemon           45.521284
hendon            45.064262
edus              43.454824
minigame          40.235948
incisors          38.484211
glaring           37.017072
lawless           32.986466
proving           31.016954
lookahead         30.579320
delay             29.901236
foraging          28.969882
stuffing          28.969882
responding        28.785518
fuss              28.405013
ram               27.488722
routes            27.488722
5196              27.360445
cult              26.562932
villalobos        25.751007
fitted            25.751007
safeties          24.739850
warnings          24.739850
areal             24.141569
presses           23.823559
powder            23.823559
negro             22.532131
obvs              22.532131
permanently       20.922693
acknowledges      20.922693
chewing           20.433025
etsu              19.313255
letterman         19.313255
Name: (02, 18), dtype: float64
```


letterman
etsu
chewing
acknowledges
permanently
obvs
negro
powder
presses
areal
warnings
safeties
fitted
villalobos
cult
5196
routes
ram
fuss
responding
stuffing
foraging
delay
lookahead
proving
lawless
glaring
incisors
minigame
edus
hendon
pokemon
ventilators
boundless
obligate
hooker
heyman
caroliniana
ironwood
carpinus
hornbeam
whereas
equestrian
molars



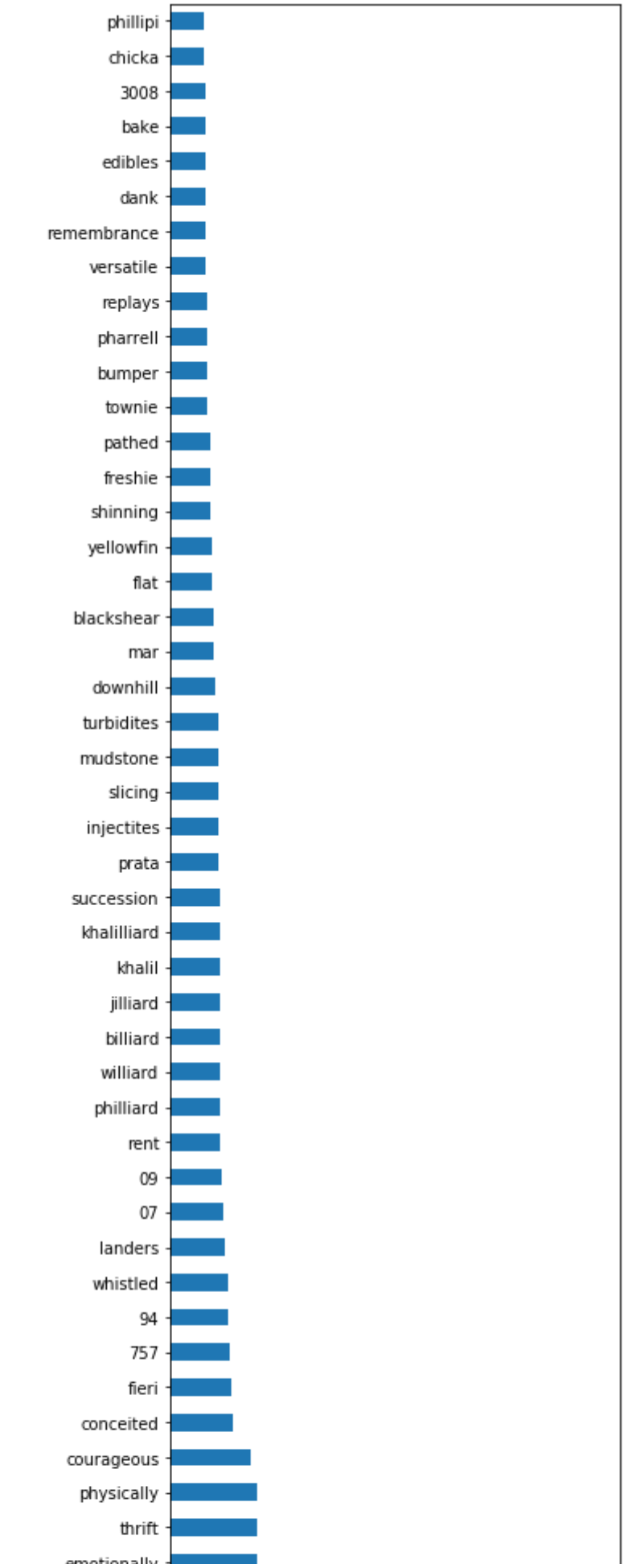

Week two popular words contain a lot of keywords that have to do with college life such as 'barstool' and 'proofread'. There were no holidays so it makes sense students were just talking about regular school happenings. There are also a lot of words that have to do with plants such as 'hornbeam' and 'carpinus' so there might have been something to do with that going on.

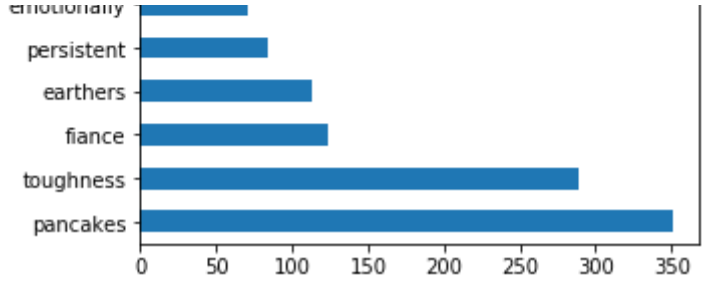
In [269]:

```
a = TF_IDF.iloc[2]
one = a.where(a>0).dropna().sort_values(ascending = False)
filt = one.index.str.contains('http|@', regex = True)#filter out links and @names
one = one[~filt]
filt1 = one.index.str.isalnum()
one = one[filt1]
d = one[:50]
chart = d.plot.barh(figsize=(5,20), sharex=True)
d
```

Out[269]:

```
pancakes       350.939350
toughness      288.631581
fiance         123.926719
earthers       112.660654
persistent      83.690771
emotionally     70.736506
thrift          70.554386
physically      70.513362
courageous      65.986954
conceited       51.593388
fieri           49.892575
757             48.283137
94              46.730827
whistled        46.730827
landers         43.981955
07              43.454824
09              41.845386
rent            40.866050
philliard       40.235948
williard        40.235948
billiard        40.235948
jilliard        40.235948
khalil          40.235948
khalilliard     40.235948
succession      40.235948
prata           39.400501
injectites      38.626510
slicing         38.626510
mudstone        38.626510
turbidites      38.626510
downhill        36.268619
mar             35.033538
blackshear      34.587250
flat            34.140963
yellowfin       33.798196
shinning        32.188758
freshie         32.188758
pathed          32.188758
townie          30.579320
bumper          30.579320
pharrell        30.579320
replays         29.321303
versatile       28.969882
remembrance     28.405013
dank            28.405013
edibles         28.405013
bake            28.405013
3008            28.405013
chicka          27.360445
phillipi        27.360445
Name: (03, 05), dtype: float64
```


phillipi
chicka
3008
bake
edibles
dank
remembrance
versatile
replays
pharrell
bumper
townie
pathed
freshie
shinning
yellowfin
flat
blackshear
mar
downhill
turbidites
mudstone
slicing
injectites
prata
succession
khalilliard
khalil
jilliard
billiard
williard
phillard
rent
09
07
landers
whistled
94
757
fieri
conceited
courageous
physically
thrift




This one is pretty random I didn't really see a trend.

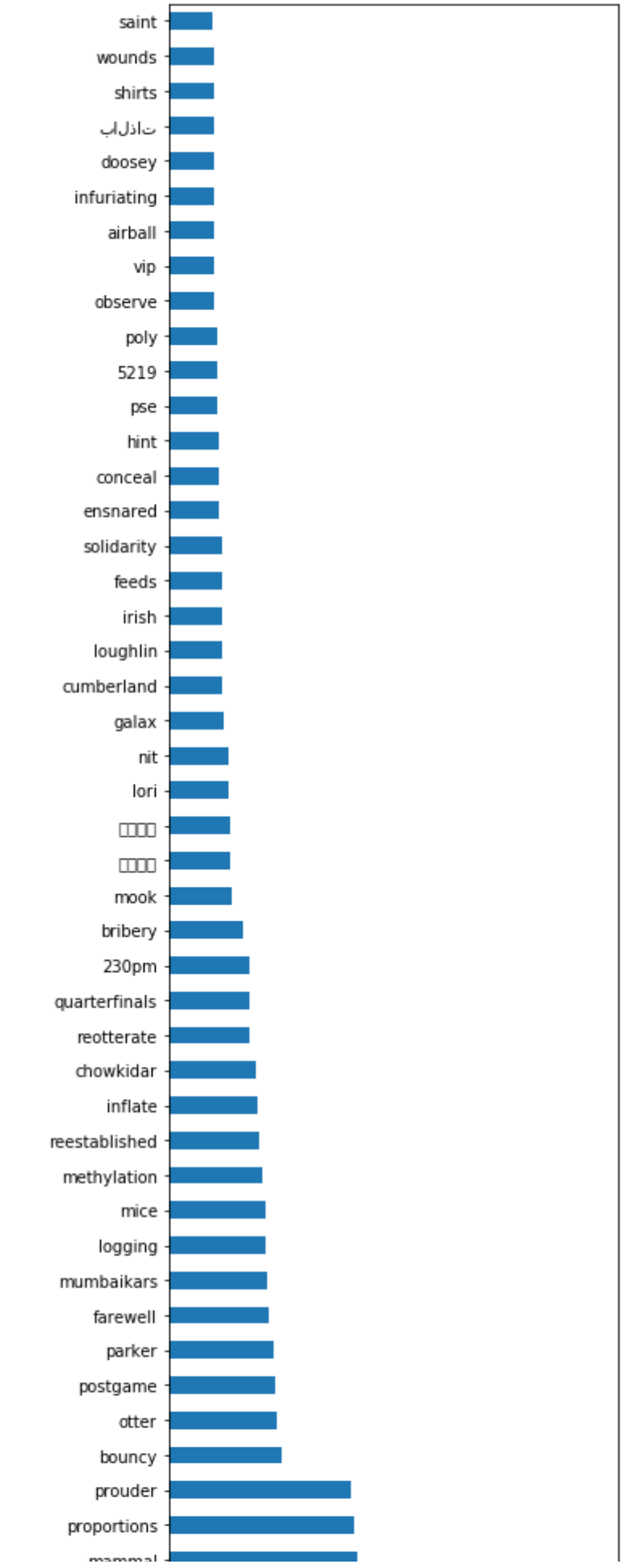
In [271]:

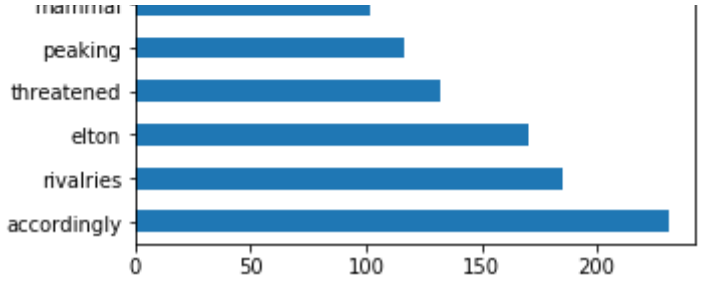
```
a = TF_IDF.iloc[3]
one = a.where(a>0).dropna().sort_values(ascending = False)
filt = one.index.str.contains('http|@', regex = True)#filter out links and @names
one = one[~filt]
filt1 = one.index.str.isalnum()
one = one[filt1]
d = one[:50]
chart = d.plot.barh(figsize=(5,20), sharex=True)
d
```

Out[271]:

```
accordingly      230.905264
rivalries        185.085360
elton            170.600419
threatened       131.945865
peaking          116.368923
mammal           101.394588
proportions       99.785151
prouder           98.043108
bouncy            61.158641
otter             58.017323
postgame          57.726316
parker            56.701644
farewell          54.147516
mumbaikars        53.111451
logging           52.228572
mice              52.228572
methylation       50.395990
reestablished     48.563409
inflate           47.647118
chowkidar         46.673699
reotterate        43.454824
quarterfinals     43.454824
230pm             43.454824
bribery           40.235948
mook              33.902757
TEST              32.986466
TEST              32.986466
lori              32.188758
nit               32.182014
galax             29.321303
cumberland        28.969882
loughlin          28.969882
irish             28.562375
feeds             28.405013
solidarity        28.405013
ensnared          27.360445
conceal           27.360445
hint              27.073758
pse               25.751007
5219              25.751007
poly              25.751007
observe           24.768934
vip               24.322647
airball           24.141569
infuriating       24.141569
doosey            24.141569
24.141569            بالذات
shirts            24.099504
wounds            24.099504
saint             23.497979
Name: (03, 12), dtype: float64
```

```
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120399 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120384 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120398 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120295 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120280 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120294 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120399 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120384 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120398 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120295 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120280 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120294 missing from current font.
  font.set_text(s, 0, flags=flags)
```


saint
wounds
shirts
بالذات
doosey
infuriating
airball
vip
observe
poly
5219
pse
hint
conceal
ensnared
solidarity
feeds
irish
loughlin
cumberland
galax
nit
lori
□□□□
□□□□
mook
bribery
230pm
quarterfinals
reotterate
chowkidar
inflate
reestablished
methylation
mice
logging
mumbaikars
farewell
parker
postgame
otter
bouncy
prouder
proportions




This one has a lot of words to do with basketball games like 'postgame', 'airball', 'bouncy'. Indicates it was basketball season and people were talking about that

In [272]:

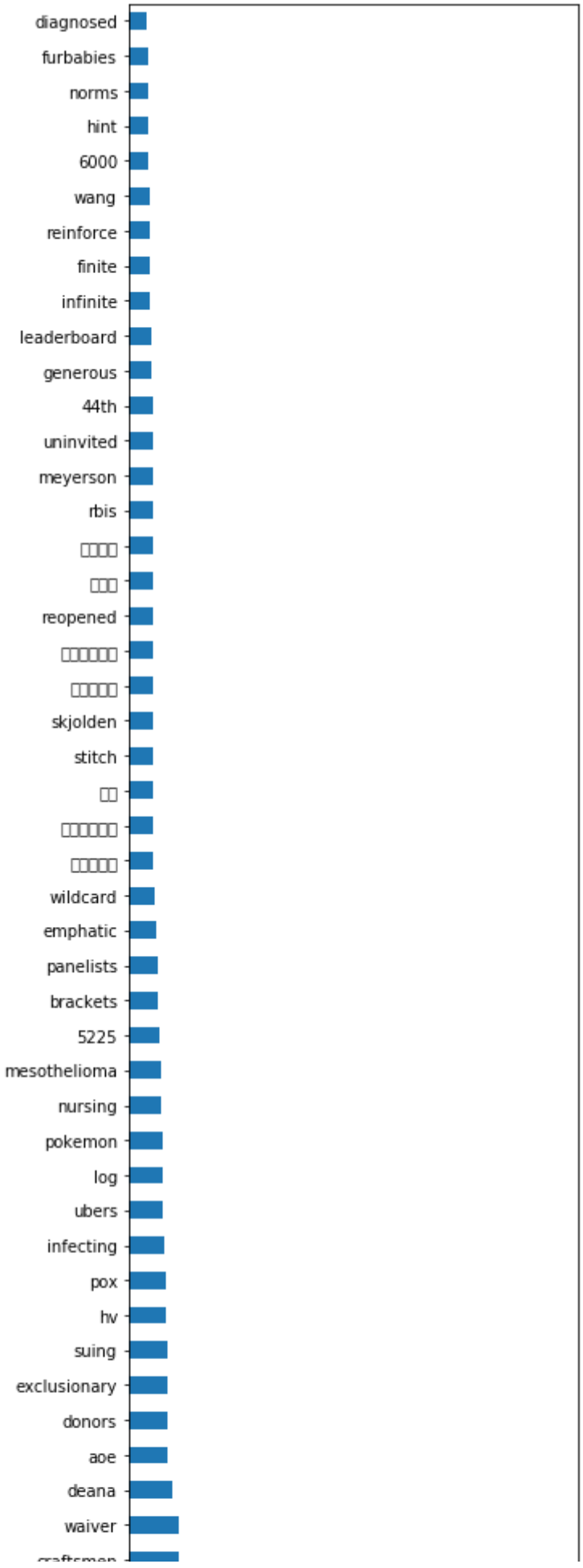
```
a = TF_IDF.iloc[4]
one = a.where(a>0).dropna().sort_values(ascending = False)
filt = one.index.str.contains('http|@', regex = True)#filter out links and @names
one = one[~filt]
filt1 = one.index.str.isalnum()
one = one[filt1]
d = one[:50]
chart = d.plot.barh(figsize=(5,20), sharex=True)
d
```

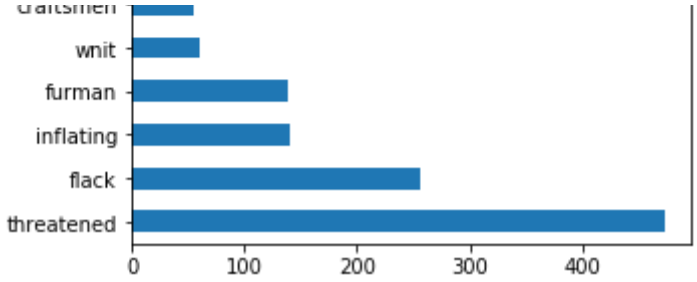
Out[272]:

```
threatened      472.806018
flack           255.900628
inflating       140.021098
furman          138.411660
wnit             59.549203
craftsmen        55.893735
waiver           54.720889
deana            48.283137
aoe              43.454824
donors           42.398527
exclusionary     41.845386
suing            41.845386
hv               40.235948
pox              40.235948
infecting        38.626510
ubers            37.017072
log              36.595542
pokemon          36.595542
nursing          36.149255
mesothelioma     35.407634
5225             33.798196
brackets         32.692840
panelists        32.182014
emphatic         30.579320
wildcard         28.969882
drink            27.360445
dunkin           27.360445
my               27.360445
stitch           27.360445
skjolden         27.360445
wrong            27.360445
pissed           27.360445
reopened         27.360445
and              27.360445
made             27.360445
rbis             26.572431
meyerson         25.751007
uninvited        25.751007
44th             25.751007
generous         25.030456
leaderboard      24.141569
infinite         22.907268
finite           22.907268
reinforce        22.907268
wang             22.532131
6000             21.965502
hint             21.454676
norms            20.943851
furbabies        20.922693
diagnosed        20.158396
Name: (03, 19), dtype: float64
```

```
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119837 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119851 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119842 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119847 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119844 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119854 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119846 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119858 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119856 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119848 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119840 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120005 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119998 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 120008 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119890 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119993 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119834 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:211: RuntimeWarning: Glyph 119838 missing from current font.
  font.set_text(s, 0.0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119837 missing from current font.
  font.set_text(s, 0, flags=flags)
```

```
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119851 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119842 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119847 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119844 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119854 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119846 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119858 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119856 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119848 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119840 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120005 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119998 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 120008 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119890 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119993 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119834 missing from current font.
  font.set_text(s, 0, flags=flags)
//anaconda3/lib/python3.7/site-packages/matplotlib/backends/backend_ag
g.py:180: RuntimeWarning: Glyph 119838 missing from current font.
  font.set_text(s, 0, flags=flags)
```


diagnosed
furbabies
norms
hint
6000
wang
reinforce
finite
infinite
leaderboard
generous
44th
uninvited
meyerson
rbis
□□□□
□□□
reopened
□□□□□□
□□□□□
skjolden
stitch
□□
□□□□□□
□□□□□
wildcard
emphatic
panelists
brackets
5225
mesothelioma
nursing
pokemon
log
ubers
infecting
pox
hv
suing
exclusionary
donors
aoe
deana
waiver




This one has a lot of medical words like 'infecting', 'mesothelioma', 'diagnosed'. Maybe a lot of people were getting sick during this time period.

In [ ]: