

lab6_trending_topics

February 19, 2024

1 Lab 6: Trending Topics

In this lab, you'll use what you know about dictionary counters to create trending topics feature for Reddit!

Imagine you're in charge of creating a trending topics feature. Given a subreddit, you want to search through the x hottest posts and keep track of which important terms are showing up the most often.

First, let's try to do it naively. Go through the normal praw steps:

```
[90]: import praw
```

```
[91]: %run reddit_keys.py
```

```
[92]: reddit = praw.Reddit(
        username=username, password=password,
        client_id=client_id, client_secret=client_secret,
        user_agent="a custom python script for user /" + str(username)
    )
```

Now, create a function that accepts a subreddit, and returns the top 5 “trending topics.” You can do this by using a dictionary counter to keep track of every unique word in a submission's title, counting how many times they appear, and putting the top 5 into a ranked list.

To do that, it will be helpful to be able to keep track of individual words! You can use the split function to do this easily, and don't forget to make sure everything is lowercase:

```
[93]: "Let's split up a string!".lower().split()
```

```
[93]: ["let's", 'split', 'up', 'a', 'string!']
```

```
[94]: def trending(sub):
        submissions = reddit.subreddit(sub).hot(limit=25)
        submissions_list = list(submissions)
        topics = []
        word_counts = {}

        for submission in submissions_list:
            title_words = submission.title
```

```

    for word in title_words.lower().split():
        if word not in word_counts:
            word_counts[word] = 1
        else:
            word_counts[word] += 1

    sorted_topics = sorted(word_counts.items(), key=lambda x: -x[1])
    topics = sorted_topics[:50]

    return topics

#test your function here!
trending('Movies')

```

```

[94]: [('the', 10),
      ('in', 9),
      ('a', 8),
      ('and', 7),
      ('to', 5),
      ('movie', 5),
      ('i', 5),
      ('of', 4),
      ('what', 4),
      ('you', 4),
      ('best', 3),
      ('movies', 3),
      ('with', 3),
      ('films', 2),
      ('official', 2),
      ('discussion', 2),
      ('one', 2),
      ('me', 2),
      ('awards:', 2),
      ('film', 2),
      ('how', 2),
      ('ever', 2),
      ('most', 2),
      ('for', 2),
      ('was', 2),
      ('do', 2),
      ('be', 2),
      ('channels', 2),
      ('like', 2),
      ('are', 2),
      ('is', 2),
      ('actually', 2),

```

```

('full', 2),
('2023', 1),
('ranked', 1),
('according', 1),
('/r/movies', 1),
('polls.', 1),
('megathread', 1),
('madame', 1),
('web', 1),
('/', 1),
('bob', 1),
('marley:', 1),
('love)', 1),
('jordan', 1),
('peelee's', 1),
('next', 1),
('set', 1),
('release', 1)]

```

After you get it working, you might notice some issues. For example, words like “the”, “and”, “it”, etc aren’t really topics, but they probably keep appearing as your top topics! How can we get rid of them?

For that, we can turn to information theory! There’s a term called “inverse-document frequency,” that basically tries to quantify how much information a word provides (i.e. how rare it is). Basically, we divide the number of total number of documents, by the number of documents a word appears in.

For example, if we have three sentences: 1. “The best ice cream flavor is chocolate.” 2. “Only idiots think the best ice cream flavor is chocolate.” 3. “I hate ice cream”

...the word “the” appears in two documents (1 and 2), and there are three documents total. So, the idf for the word “the” is $3/2 = 1.5$.

To try and get better topics, your next step is to define a function to calculate the idf (inverse-document frequency) of a word. This will later help you to decide whether some words are too common to consider “topics”!

First, I asked ChatGPT to come up with a random set of 100 made-up Reddit titles to serve as a corpus of documents:

```

[95]: corpus = [
    "What's the weirdest thing you've found in a parallel universe?",
    "My cat learned to speak English and now won't stop giving me advice",
    "Time traveler here, AMA about the future!",
    "I accidentally summoned a demon while baking cookies, help!",
    "Just discovered a hidden society of talking frogs in my backyard, what do_
↪I do?",
    "Ask me anything, I'm a professional dragon tamer",

```

"Invented a device that lets me visit fictional worlds, it's not as great,
↳as it sounds",

"Help! My pet rock is missing, last seen orbiting Jupiter",

"Astronauts of Reddit, what's the scariest thing you've encountered in,
↳space?",

"I woke up with wings this morning, now what?",

"I built a robot that writes Shakespearean poetry, and it's now entered a,
↳poetry slam",

"Just got hired as a ghost therapist, any advice?",

"I accidentally turned my cat into a wizard, and now he won't stop casting,
↳spells",

"Visited a haunted library and now fictional characters are haunting my,
↳house, need advice",

"Found a map leading to the Fountain of Unlimited Coffee, who's in?",

"My refrigerator has become a portal to another dimension, what snacks,
↳should I bring?",

"I cloned myself, and now we're both avoiding doing chores",

"I can communicate with plants, and they're gossiping about us",

"I made a deal with a leprechaun for unlimited pizza, was it worth it?",

"Just became the mayor of a town populated by talking animals, any tips for,
↳governance?",

"Accidentally enrolled in a wizard school, how do I survive the first day?",

"I accidentally started a dance-off with aliens, and now they won't leave",

"Bought a cursed mirror on eBay, now my reflection has a mind of its own",

"My pet rock is evolving, should I be concerned?",

"I found a portal to the internet in my backyard, what subreddits should I,
↳visit?",

"I'm a time-traveling detective solving crimes in different centuries, AMA",

"My cat started a book club for other cats, and I'm the only human invited",

"Just adopted a ghost, any tips for a happy afterlife?",

"I turned my basement into a portal to the land of unicorns, now what?",

"I befriended a ghost, and now they're my roommate, any ghost etiquette,
↳tips?",

"I discovered a potion that turns everything into chocolate, chaos ensued",

"Aliens mistook me for their leader, and now I have to give a speech, help!
↳",

"I accidentally opened a portal to the land of sentient pizza, now they,
↳want to conquer Earth",

"I can communicate with dolphins, and they have strong opinions about human,
↳fashion",

"My cat hacked my computer and started an online business, should I be,
↳proud?",

"I found a genie in a bottle, but all he grants are really specific wishes",

"I'm a time-traveling barber, AMA about the weirdest hairstyles in history",

"Accidentally enrolled in a wizard cooking class, now my soup is casting,
↳spells",

"I discovered a secret society of talking plants, and they have a plan for
↳world domination",

"Just adopted a ghost dog, any tips on walking a transparent pet?",

"I became a superhero with the power of controlling spaghetti, it's messier
↳than you think",

"I'm a professional dragon therapist, AMA about helping dragons with their
↳emotional issues",

"My cat is attending Hogwarts, and I'm not sure how to feel about it",

"Found a map to the land of perpetual pizza, packing my bags now",

"I accidentally swapped bodies with my pet goldfish, AMA",

"Started a support group for people with haunted furniture, join us every
↳Tuesday",

"Accidentally became the king of a medieval kingdom, now I have to learn
↳jousting",

"I accidentally created a portal to a world without memes, how do I fix
↳this?",

"Just received a letter from my future self, and it's full of spoilers",

"I befriended a ghost, and now they're my personal stylist, any fashion
↳tips from the afterlife?",

"My cat joined a rock band, and they're about to release their first album",

"I'm a time-traveling chef, AMA about cooking in different eras",

"Accidentally became a wizard's apprentice, now my broom won't stop
↳misbehaving",

"Found a genie who grants wishes but only in the form of interpretative
↳dance",

"My pet robot learned to breakdance, and now he's challenging other robots
↳to dance-offs",

"I discovered a secret society of time-traveling librarians, they have
↳overdue books from every era",

"I can talk to inanimate objects, and they have strong opinions about
↳interior design",

"Accidentally became a detective in a world where everyone communicates
↳through interpretive dance",

"I adopted a ghost cat, and now my allergies are haunting me",

"My cat started a podcast for other cats, and it's surprisingly popular",

"I discovered a portal to a world where cats are the rulers and humans are
↳the pets",

"Accidentally started a revolution in a parallel universe, any tips for
↳diplomatic negotiations?",

"I found a treasure map, but the treasure is just a really good sandwich",

"My pet turtle learned to play the piano, and now he's composing
↳symphonies",

"I became a wizard's personal chef, now I have to cook magical meals every
↳day",

"I accidentally joined a secret society of time-traveling bakers, and now
↳I'm stuck in the Renaissance with a loaf of bread",

"My cat learned to use a smartphone and won't stop taking selfies",
 "I discovered a parallel universe where socks are the dominant species, and
 ↳they have a complex society",
 "Accidentally started a game of hide and seek with interdimensional beings,
 ↳and now they're really good at it",
 "My pet parrot learned to speak Klingon, and now he's giving me orders in
 ↳an alien language",
 "I found a portal to a world where puns are the currency, and now I'm a
 ↳billionaire",
 "Accidentally joined a secret society of time-traveling stand-up comedians,
 ↳and now I'm performing in ancient Rome",
 "My cat joined a circus, and he's the star of the tightrope-walking act",
 "I discovered a parallel universe where everyone communicates through
 ↳interpretive dance, and now I can't stop dancing",
 "Accidentally became a superhero with the power of extreme puns, villains
 ↳beware",
 "My pet hamster learned to drive a tiny car, and now he's chauffeuring his
 ↳rodent friends",
 "I found a portal to a world where dogs are the rulers, and humans are the
 ↳loyal companions",
 "Accidentally became a wizard's apprentice, and now my homework is casting
 ↳spells",
 "My cat became a detective and solved the mystery of the missing catnip",
 "I discovered a parallel universe where gravity works in reverse, and now I
 ↳can't stay on the ground",
 "Accidentally started a band with interdimensional beings, our first single
 ↳is a hit in multiple universes",
 "My pet snake learned to play the guitar, and now he's forming a reptilian
 ↳rock band",
 "I found a portal to a world where llamas are the dominant species, and
 ↳they have a llama monarchy",
 "Accidentally enrolled in a wizard's improv class, and now my jokes are
 ↳casting spells",
 "My cat started a fashion line for other cats, and it's taking the catwalk
 ↳by storm",
 "I discovered a parallel universe where humans communicate through meows,
 ↳and now I'm fluent in cat",
 "Accidentally became a superhero with the power of telekinetic marshmallow
 ↳control, it's stickier than it sounds",
 "My pet chameleon learned to mimic human speech, and now he's prank calling
 ↳my neighbors",
 "I found a portal to a world where squirrels are the rulers, and humans are
 ↳the ones hoarding nuts",
 "Accidentally joined a secret society of time-traveling synchronized
 ↳swimmers, and now I'm performing in the Victorian era",

```
]

```

Now, define a function that takes a string, and calculates the idf of that string, using that corpus:

```
[96]: def idf(string):
      #your code here
      count = 0

      for item in corpus:
          if string in item.lower().split():
              count += 1

      if count > 0:
          return 100 / count
      else:
          return None

#testing
idf('cat')
```

```
[96]: 8.333333333333334
```

Finally, modify your trending function, to only consider trending topics above a certain idf threshold (you can play around with this a bit to decide what would be a good cutoff point). Does this seem to help?

```
[97]: def trending(sub):
      submissions = reddit.subreddit(sub).hot(limit=25)
      submissions_list = list(submissions)
      topics = []
      word_counts = {}

      for submission in submissions_list:
          title_words = submission.title
          corpus.append(title_words)
          for word in title_words.lower().split():
              if idf(word) > 5:
                  if word not in word_counts:
                      word_counts[word] = 1
                  else:
                      word_counts[word] += 1

      sorted_topics = dict(sorted(word_counts.items(), key=lambda x: -x[1]))
      topics = list(word_counts.keys())[:5]

      return topics
```

```
#test your function here!  
trending('Movies')
```

```
[97]: ['2023', 'films', 'ranked', 'according', '/r/movies']
```