AUTOMATIC RECONFIGURING NEWS SITE

S5 CSB

Nikhill Shanmukhan
Sachit Anand
Siddharth Biju
Sidharth Sreekumar

Page of Contents

Introduction

- News is the information about an event happening right now or that is

- Online news reading has become a popular way to read news articles from a huge collection of news sources all around.
- In this work we present a method for identifying top news in the web environment[1] that consists of diversied news portals.

Problem Statement

- News websites are one of the most visited destinations on the web.
- As there are many news portals created on a daily basis, each having its own preference for which news are important. Detecting unbiased[2] important and popular news are the problematic part of the news portal development.

Approach 1

Input : nil

Output : website with trending news

server side algorithm:

step 1 : for each 24 hours :

step 1.1 : twurl /1.1/trends/place.json?id=1 ¿ copy.json

step 2: data = file.read()

step 3: from i=0 to i=10

step 3.1: temp = data[0]["trends"][i]["name"]

Step 3.2: temp = temp[1:]

Step 3.3: temp = temp + " latest news articles"

Step 3.4: keyarray.append(temp)

Approach 1 cont ...

step 4: crawl google for news links using keyarray values
step 5: seperate title and content from links obtained and append to an array with id
step 6: accept get request with news id
step 7: return array[id]

client side algorithm:

step 1 : render navbar .js
step 2: render contents .js

Approach 1 cont....

contents .js

1. for(i=0;i¡10;i++)
   1.1 get 127.0.0.1 with i as id
   1.2 store response in an array
2. render title and contents in array

Approach 2

Input: nil
Output: website with popular news articles

step 1: For every 1 hour extract latest news article details from different
news site (includes TimeofPublish ,ups,downs)
step 2:Find the number of articles extracted
step 3: Using reddit hot ranking algorithm sort the news articles based on popularity rank
step 4:Display each article based on the rank calculated.

Approach 2 cont ...

reddit hot ranking algorithm[4]:

step 1:start

step 2:for each article extract the time of publish,upcount,downcount

step 3:set ep=1970,1,1 in datetime format

step 4.define a function epsec(date)

  step 4.1:td=date - ep

  step 4.2:return td.days * 86400 + td.seconds +
    (float(td.microseconds) / 1000000)

step 5:define a function score(upcount,downcount)

  step 5.1:return upcount-downcount

Approach 2 cont ...

step 6:define a function hot(upcount,downcount,date)

step 6.1:s = score(upcount, downcount)

step 6.2:order = log(max(abs(s), 1), 10)

step 6.3:sign = 1 if s is grater than 0 else -1 if s less than 0 else 0

step 6.4:seconds = epsec(date) - 1134028003

step 6.5:return round(sign * order + seconds / 45000, 7)

step 7:stop

Approach 3

- **Hit Search Algorithm**
  Input: article dataset
  Output : Article order according to popularity

Step 1: Read dataset containing articles
Step 2: Extract title,Number of articles n
Step 3: Do steps 4,5 for n articles
Step 4: Search title using any search engine
Step 5: Read and store corresponding number of hits
Step 6: sort articles according to number of hits
Step 6: Do steps 7 for n articles
Step 7: Print title and content of sorted articles

Approach 4

- View Sort Algorithm
  Input: Nil
  Output: Articles sorted based on popularity

Step 1:Search for all news sites with trending news using search engine
Step 2:Repeat Step 3,4 until no more chances left
Step 3:Views of every page is stored
Step 4:Compare views of all pages
Step 5:Content of page with high views is extracted

Approach 4 cont...

Step 6:All articles extracted are sorted according to the number of views

Step 7:Print title and content of sorted articles

Comparison

- On comparing Approach 1 and 4 ,Approach 1 takes only articles from twitter and articles may not be always available.
- comparing Approach 4 and 2,Approach 2 its more difficult to implement to news articles as downvote[3] and upvote[4] my not be present in all articles.

Comparison contd..

- Approach 3 has possible cases of HTTP error and popularity may not be accurate

Solution and Conclusion

- In this work, we propose a method to detect and to rank important news stories in web environment
- We use google search to fetch news from all website.
- Then we analyze the total views of all website using google analytics.
- After that we separate the title and content of the news articles and display it.

Relevance

- Our websites gives the latest and popular information to the public – political, social, sports, entertainment etc.
- The site play a vital role in educating and informing Mass with latest updates, current happenings around the globe.

Reference

[1]https://www.ibm.com/

[2]https://en.wikipedia.org/

[3]https://en.wiktionary.org/

[4]https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9