Achoo

By: Saad Kothawala, Christian Lanzer, Ryan Sevilla

The Initial Problem

- We aren’t doctors (yet)…
  - We had to get the diseases and the symptoms from somewhere
  - So we created a scraper using a python script that goes through MayoClinic’s website and pulls down all the diseases from their website.
  - This way we could accurately store all of the necessary info that our project is using.

TextEdit File Edit Format View Window Help

54% Wed Apr 22 7:01 PM

Scrape.py

```
18
19  rtTime = time.time()
20
21  easesAdded = []
22
23  nt = 0
24  letter in string.ascii_lowercase:#Go through all the alphabet
25    page = urllib2.urlopen(DISEASE_LIST_LINK_BASE + letter).read() #get the html
26    tree = html.fromstring(page) #make the html easier to use
27    links = tree.xpath('//*[@id="index"]/ol/li/a') #Get all diseases in the listing
28
29    for l in links:
30        diseaseName = l.text
31        if diseaseName[0].lower() == letter and diseaseName not in diseasesAdded:#to avoid duplicates
32            toWrite = diseaseName + "\t";
33            count += 1
34            print count, ": ", diseaseName
35            diseaseLink = MAIN_LINK + l.attrib['href'].replace('definition', 'symptoms')
36
37            diseasePage = urllib2.urlopen(diseaseLink).read()
38            diseaseTree = html.fromstring(diseasePage)
39            symptoms = diseaseTree.xpath('//*[@id="main-content"]/ul[1]/li')
40
41            flag = 0
42            for symptom in symptoms:
43                if not "Tweet" in symptom.text_content() and not "Share" in symptom.text_content() and not "Definition" in symptom.text_content() and not "Cau
44                    toWrite += (symptom.text_content().replace("\r\n", "") + "\t").encode("utf-8")
45                    flag = 1
46                    diseasesAdded.append(diseaseName)
47
48
49
50            toWrite += "\n";
51            toWrite.encode('utf-8')
52            if flag is 1:#write only if symptopms present
53                obj = open(FILE_NAME, 'ab')
54                obj.write(toWrite)
55                obj.close
56
57
58
59
```

Line 1, Column 1 Tab Size: 4 Python

scraping.png

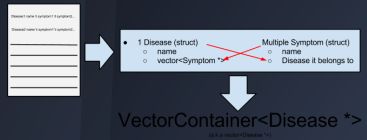
Scraper — bash — 80×24

```
1123 :  Waldenstrom macroglobulinemia
1124 :  Water on the knee
1125 :  Wegener's granulomatosis
1126 :  West Nile virus
1127 :  Wheat allergy
1128 :  Whiplash
1129 :  Whipple's disease
1130 :  Whooping cough
1131 :  Wilms' tumor
1132 :  Wilson's disease
1133 :  Wolff-Parkinson-White syndrome
1134 :  Wrinkles
1135 :  Wrist pain
1136 :  X-linked agammaglobulinemia
1137 :  Yeast infection (vaginal)
1138 :  Yellow fever
1139 :  Yips
1140 :  Zollinger-Ellison syndrome
--------------Done Running Mayo Clinic Scraper--------------

Elements Added:  1140
Run Time:  1372.73157096  seconds

Saads-MacBook-Pro:Scraper Saad$
```

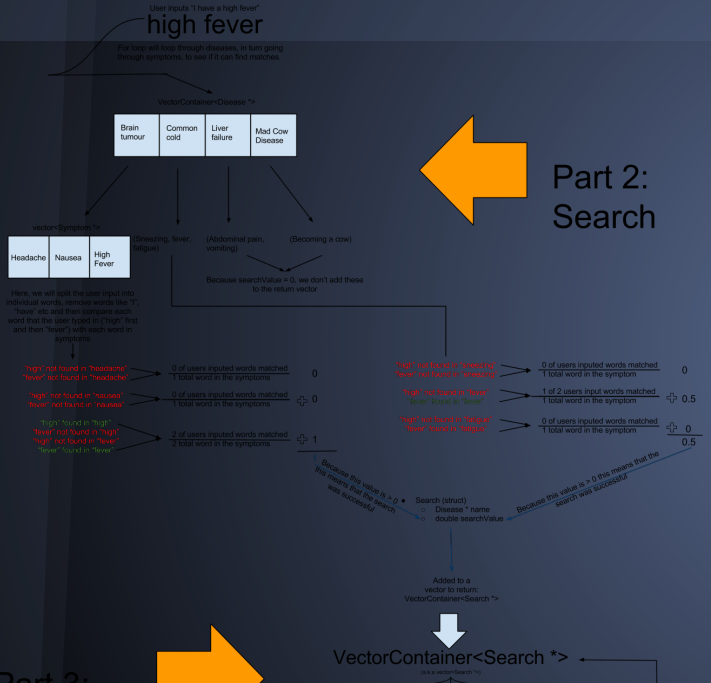
20 mins b/c 1 thread

diseases.txt

```
Achilles tendon rupture Pain, possibly severe, and swelling near your heel
        An inability to bend your foot downward or "push off" the injured
leg when you walk       An inability to stand on your toes on the injured
leg     A popping or snapping sound when the injury occurs
ACL injury      A loud "pop" sound      Severe pain and inability to
continue activity       Knee swelling that usually worsens for hours after
the injury occurs       A feeling of instability or "giving way" with
weight bearing
Acne    Whiteheads (closed plugged pores)       Blackheads (open plugged
pores — the oil turns brown when it is exposed to air)  Small red, tender
bumps (papules)         Pimples (pustules), which are papules with pus at
their tips      Large, solid, painful lumps beneath the surface of the skin
(nodules)       Painful, pus-filled lumps beneath the surface of the skin
(cystic lesions)
Acoustic neuroma        Hearing loss, usually gradual — although in some
cases sudden — and occurring on only one side or more pronounced on one
```



Part 1: Initialization



Part 2: Search

Part 3: Post-Search

https://docs.google.com/drawings/d/1GgqMBUnPPHhXHlT-2XGo_auM7FKmRt-RfRkKsWHXPnY/edit?usp=sharing

Demo