Extractcity.py

import read

import codecs

town = []

f = open('town.txt','w')

temp = ''

with codecs.open("in.txt") as file:#open

count = 0

for line in file:

str1 = ''

longitude = ''

latitude = ''

app = ''

for k in line:

if ('\t' == k):

count = count+1

if (count == 2):

#print k

if not('\t' in k):

str1+=k

if (count == 4):

#print k

if not('\t' in k):

longitude+=k

if (count == 5):

#print k

if not('\t' in k):

latitude+=k

if (count == 6):

break

#town.append((str1,longitude,latitude))

print str1

print longitude

print latitude

if (temp != str1):

app = str1+'\t'+longitude+'\t'+latitude+'\n'

f.write(app)

count = 0;

temp = str1

analysis.py

import csv

import os

import codecs

import nltk

import json

import pprint

from nltk import pos\_tag, word\_tokenize

from time import time

from collections import Counter

from nltk.tokenize import RegexpTokenizer

India\_Province = ['ANDHRA PRADESH','ASSAM','BIHAR','CHHATTISGARH','DELHI','GOA','GUJARAT','HARYANA','HIMACHAL PRADESH','JAMMU & KASHMIR','JHARKHAND','KARNAKATA','KERALA','MADHYA PRADESH','MAHARASHTRA','ORISSA','PONDICHERRY','PUNJAB','RAJASTHAN','TAMIL NADU','UTTAR PRADESH','UTTARANCHAL','WEST BENGAL']

internet\_related = ['INTERNET','MOBILE','IP','ADDRESS','DOMAIN','WEB','SITES','PAGES','BANDWIDTH','PROVIDER','MOBIL','CARRIER','3G','DATA','TERMINAL','DEVICES','E-COMMERCE','ONLINE','BROADBAND'];

freq = {}

town\_freq = {}

max1 = 0

temp = '2012-12-31'

with codecs.open('all.txt') as file:

for line in file:

if (line == '\n'):

break

data = json.loads(line)

tokenizer = RegexpTokenizer(r'\w+')

text = tokenizer.tokenize(data['title'].upper())

print data['date'][2:5]

#making frequency table

#for w in text:

# if freq.has\_key(w):

# freq[w] = freq[w]+1

# max1 = max(max1,freq[w])

# else:

# freq[w] = 1

for w in text:

#if (w = '')

#break

if (w in (internet\_related[:1]+internet\_related[2:])):

print data['date'],'\t',data['city'].upper(),'\t\t\t',data['title'].encode('UTF-8')

if (town\_freq.has\_key(data['city'].upper())):

town\_freq[data['city'].upper()] = town\_freq[data['city'].upper()]+1

else:

town\_freq[data['city'].upper()] = 1

if freq.has\_key(w):

freq[w] = freq[w]+1

max1 = max(max1,freq[w])

else:

freq[w] = 1

#print nltk.pos\_tag(text)

#print (data)

if (data['date'] == temp):

print town\_freq

#print town\_freq

#print freq

#print max1