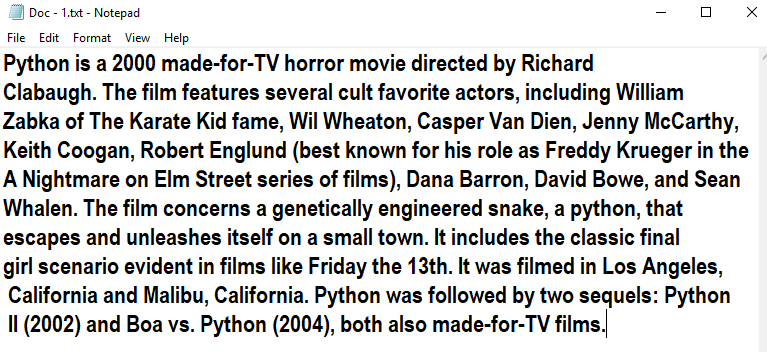
Web Mining Lab Assignment 4

Name: Om Ashish Mishra

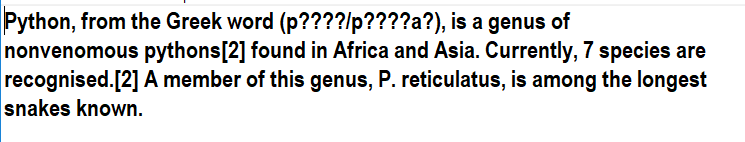
Registration Number: 16BCE0789

Slot: F2

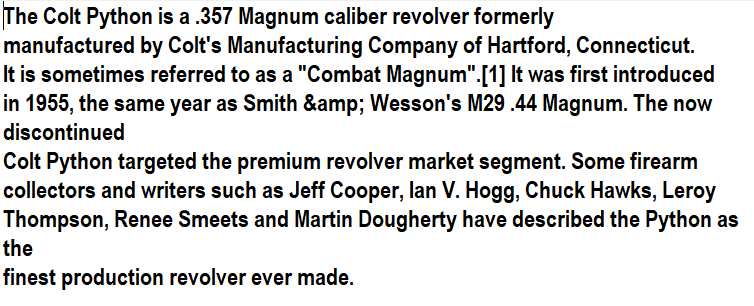
The Document 1:



The Document 2:



The Document 3:



# The Code:

import math

f1=open("Doc - 1.txt",'r')

f2=open("Doc - 2.txt",'r')

f3=open("Doc - 3.txt",'r')

l1=f1.read().split()

l2=f2.read().split()

l3=f3.read().split()

tf1=[]

tf2=[]

tf3=[]

s1=[i for i in l1]

s2=[i for i in l2]

s3=[i for i in l3]

s1=set(s1)

s1=list(s1)

s2=set(s2)

s2=list(s2)

s3=set(s3)

s3=list(s3)

for i in s1:

c=l1.count(i)

c=c/float(len(s1))

tf1.append(c)

for i in s2:

c=l2.count(i)

c=c/float(len(s2))

tf2.append(c)

for i in s1:

c=l2.count(i)

c=c/float(len(s3))

tf3.append(c)

i=0

k=0

for m in s1:

g1=1

if(m in s2):

g1+=1

if(m in s3):

g1+=1

g=math.log(3/g1)

tf1[i]=tf1[i]\*g

i+=1

j=0

for m in s2:

g2=1

if(m in s1):

g2+=1

if(m in s3):

g2+=1

g=math.log(3/g2)

tf2[j]=tf2[j]\*g

j+=1

for m in s3:

g3=1

if(m in s1):

g3+=1

if(m in s2):

g3+=1

g=math.log(3/g3)

tf1[k]=tf1[k]\*g

k+=1

myMax1 = tf1[0]

for num in tf1:

if myMax1 < num:

myMax1 = num

val1=tf1.index(myMax1)

myMax3 = tf3[0]

for num in tf3:

if myMax3 < num:

myMax3 = num

val3=tf3.index(myMax3)

myMax2 = tf2[0]

for num in tf2:

if myMax2 < num:

myMax2 = num

val2=tf2.index(myMax2)

print("Top word in Document - 1 is ---- ",s1[val1]," with tf-idf = ",myMax1)

print("Top word in Document - 2 is ---- ",s2[val2]," with tf-idf = ",myMax2)

print("Top word in Document - 3 is ---- ",s3[val3]," with tf-idf = ",myMax3)

# The Output:

