

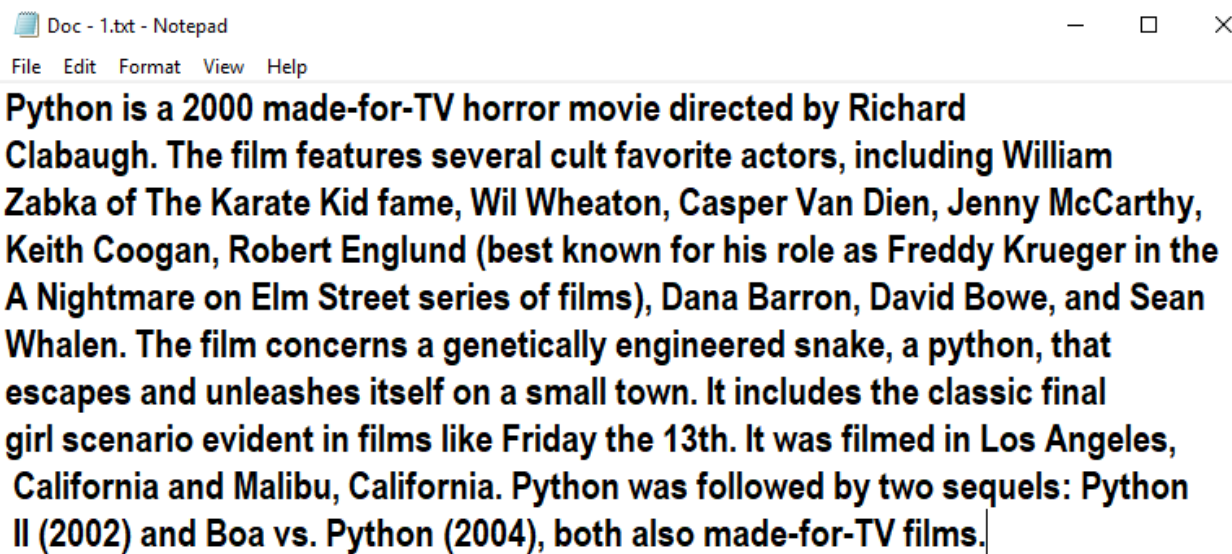
Web Mining Lab Assignment 4

Name: Om Ashish Mishra

Registration Number: 16BCE0789

Slot: F2

The Document 1:

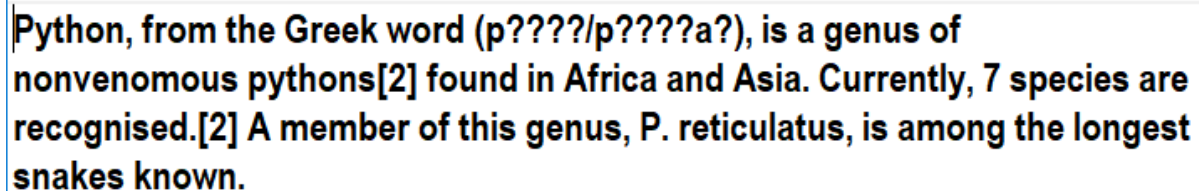


Doc - 1.txt - Notepad

File Edit Format View Help

Python is a 2000 made-for-TV horror movie directed by Richard Clabaugh. The film features several cult favorite actors, including William Zabka of The Karate Kid fame, Wil Wheaton, Casper Van Dien, Jenny McCarthy, Keith Coogan, Robert Englund (best known for his role as Freddy Krueger in the A Nightmare on Elm Street series of films), Dana Barron, David Bowie, and Sean Whalen. The film concerns a genetically engineered snake, a python, that escapes and unleashes itself on a small town. It includes the classic final girl scenario evident in films like Friday the 13th. It was filmed in Los Angeles, California and Malibu, California. Python was followed by two sequels: Python II (2002) and Boa vs. Python (2004), both also made-for-TV films.

The Document 2:



Python, from the Greek word (p????/p????a?), is a genus of nonvenomous pythons[2] found in Africa and Asia. Currently, 7 species are recognised.[2] A member of this genus, P. reticulatus, is among the longest snakes known.

The Document 3:

The Colt Python is a .357 Magnum caliber revolver formerly manufactured by Colt's Manufacturing Company of Hartford, Connecticut. It is sometimes referred to as a "Combat Magnum".[1] It was first introduced in 1955, the same year as Smith & Wesson's M29 .44 Magnum. The now discontinued Colt Python targeted the premium revolver market segment. Some firearm collectors and writers such as Jeff Cooper, Ian V. Hogg, Chuck Hawks, Leroy Thompson, Renee Smeets and Martin Dougherty have described the Python as the finest production revolver ever made.

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:

```

-----
Top word in Document - 1 is ---- film    with tf-idf =  0.024138979216251643
Top word in Document - 2 is ---- P.      with tf-idf =  0.03433163402087843
Top word in Document - 3 is ---- targeted with tf-idf =  0.02857142857142857
>>> |

```