UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TOPICOS ESPECIALES

DOCENTE : Ing. Ivan Soria Solis

ESTUDIANTE : Mayuri Ccahuana Huamaní

ANDAHUAYLAS – APURÍMAC

PERÚ

OCTUBRE, 2015

PREGUNTA1:

```
#####======================PREGUNTA 1============
###=============================================
#itera sobre todas las entradas
for e in d.entries:
  if 'summary' in e: summary=e.summary
  else:summary=e.description
  #extrae una list palabras
  words=getwords(e.title+''+summary)
  wc={}
  for word in words:
    wc.setdefault(word,0)
    wc[word]+=1
    wc1[e.title]=wc
    titu[a]=e.title
    a=a+1
    return d.feed.title,wc,titu,wc1

###=============================================
```

PREGUNTA2:

Pruebe usando la distancia euclidiana para clustering de los blogs ¿Cómo cambia esto los resultados?

En matemáticas, la distancia euclidiana o euclídea es la distancia "ordinaria" (que se mediría con una regla) entre dos puntos de un espacio euclídeo, la cual se deduce a partir del teorema de Pitágoras.

Por ejemplo, en un espacio bidimensional, la distancia euclidiana entre dos puntos P_1 y P_2, de coordenadas cartesianas (x_1, y_1) y (x_2, y_2) respectivamente, es:

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Teniendo la siguiente formula nos vamos al código de clustering.

EJERCICIO2.py - E:\UNAJMA_2015_II\TOPICOS\T-PICOS_CCAHUANA\CAPITULO 3\EJERCICIO2.py (2.7.10)

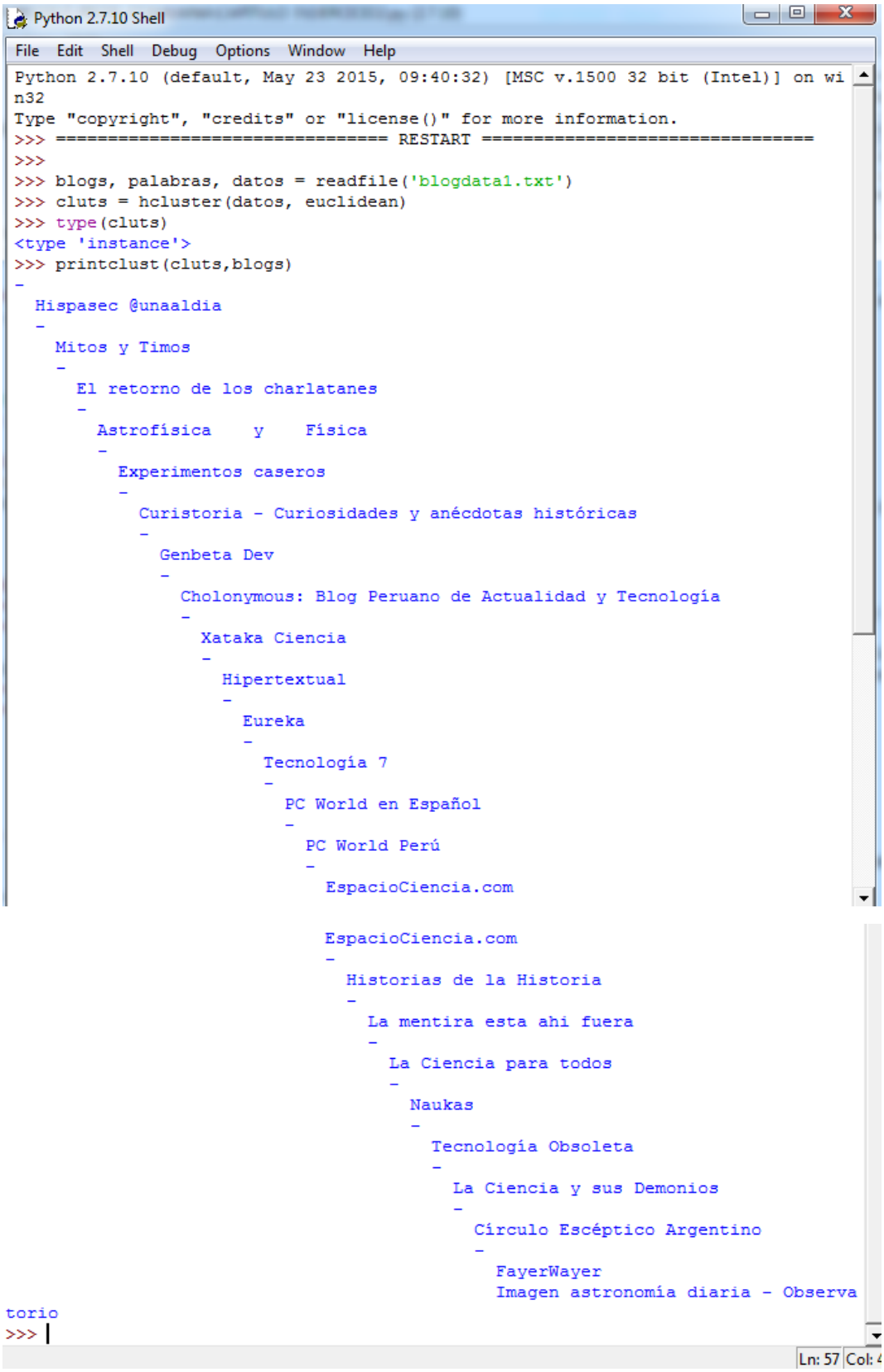
File Edit Format Run Options Window Help

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-

from PIL import Image,ImageDraw
import codecs
charset = "utf-8"

def readfile(filename):
  f=codecs.open(filename,encoding="utf-8", errors="ignore")
  lines=[line for line in f]

  # La primera linea son los titulos de columna
  colnames=lines[0].strip().split('\t')[1:]
  rownames=[]
  data=[]
  for line in lines[1:]:
    p=line.strip().split('\t')
    # La primera columna en cada fila es el nombre de la fila
    rownames.append(p[0])
    # Los datos para esta fila es lo que queda de la fila
    data.append([float(x) for x in p[1:]])
  return rownames,colnames,data


from math import sqrt
#####====================PREGUNTA 2 LA DISTANCIA EUCLIDIANA======
#####============================================================
#####============================================================
def euclidean(v1,v2):
  return sqrt(sum([(v1[i]-v2[i])**2 for i in range(len(v1))]))
#####=============================================================================================
#####=============================================================================================
def pearson(v1,v2):
  # Sumas simples
  sum1=sum(v1)
  sum2=sum(v2)

  n = float(len(v1)) # esto es necesario por que sino la division es entera

  # Sums of the squares
  sum1Sq=sum([pow(v,2) for v in v1])
```

AQUÍ SE TIENE LA FORMULA EUCLIDIANA

PARA VER LOS RESULTADOS VAMOS A EJECUTAR LA FUNCION (euclidean)

Después de ejecutar nos muestra los siguientes resultados:



```
Python 2.7.10 Shell
File  Edit  Shell  Debug  Options  Window  Help
Python 2.7.10 (default, May 23 2015, 09:40:32) [MSC v.1500 32 bit (Intel)] on wi
n32
Type "copyright", "credits" or "license()" for more information.
>>> ================================ RESTART ================================
>>>
>>> blogs, palabras, datos = readfile('blogdata1.txt')
>>> cluts = hcluster(datos, euclidean)
>>> type(cluts)
<type 'instance'>
>>> printclust(cluts,blogs)
-
  Hispasec @unaaldia
  -
    Mitos y Timos
    -
      El retorno de los charlatanes
      -
        Astrofísica    y    Física
        -
          Experimentos caseros
          -
            Curistoria - Curiosidades y anécdotas históricas
            -
              Genbeta Dev
              -
                Cholonymous: Blog Peruano de Actualidad y Tecnología
                -
                  Xataka Ciencia
                  -
                    Hipertextual
                    -
                      Eureka
                      -
                        Tecnología 7
                        -
                          PC World en Español
                          -
                            PC World Perú
                            -
                              EspacioCiencia.com

                              EspacioCiencia.com
                              -
                                Historias de la Historia
                                -
                                  La mentira esta ahi fuera
                                  -
                                    La Ciencia para todos
                                    -
                                      Naukas
                                      -
                                        Tecnología Obsoleta
                                        -
                                          La Ciencia y sus Demonios
                                          -
                                            Círculo Escéptico Argentino
                                            -
                                              FayerWayer
                                              Imagen astronomía diaria - Observa
torio
>>>
                                                                    Ln: 57  Col: 4
```

PREGUNTA2:

Investigue a cerca de la distancia de Manhattan. Cree una función para esta y vean cómo cambian los resultados?

MANHATTAN

Descripción general

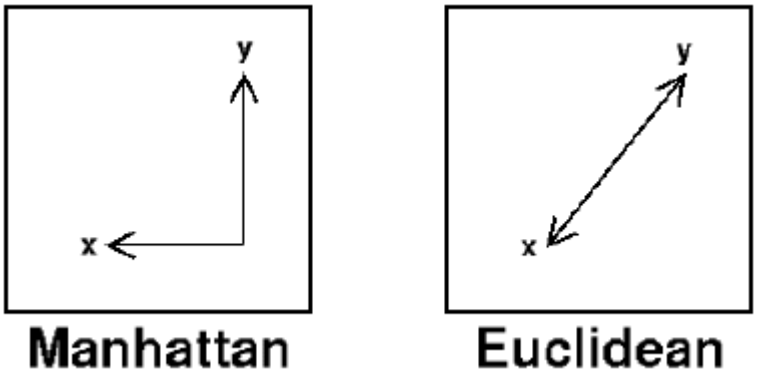
La función de la distancia Manhattan calcula la distancia que puede recorrer para llegar de un punto de datos a la otra si un camino en forma de rejilla es seguido. La distancia Manhattan entre dos elementos es la suma de las diferencias de sus correspondientes componentes.

La fórmula para esta distancia entre un punto ***X*** = ***(X*** 1, ***X*** *2, etc.) y un punto* ***Y*** = ***(Y*** 1, ***Y*** *2, etc.) es:*

$$d = \sum_{i=1}^{n} |x_i - y_i|$$

Donde n es el número de variables, y ***X*** *i* y ***Y*** *i* son los valores de la *i* ª variables, en los puntos ***X*** e ***Y, respectivamente.***

La siguiente figura ilustra la diferencia entre Manhattan distancia y distancia euclídea:



Teniendo la siguiente formula nos vamos al código de clustering.

EJERCICIO3.py - E:/UNAJMA_2015_II/TOPICOS/T-PICOS_CCAHUANA/CAPITULO 3/EJERCICIO3.py (2.7.10)

File Edit Format Run Options Window Help

```
  return rownames,colnames,data


from math import sqrt
#####====================PREGUNTA 2 LA DISTANCIA EUCLIDIANA==============
#####================================================================
#####================================================================
def euclidean(v1,v2):
  return sqrt(sum([(v1[i]-v2[i])**2 for i in range(len(v1))]))
#####================================================================
#####================================================================
#####====================PREGUNTA 2 LA DISTANCIA MANHATTAN===============
#####================================================================
#####================================================================================
def manjathan(v1,v2):
  return abs(sum([(v1[i]-v2[i]) for i in range(len(v1))]))
#####================================================================================
#####================================================================================
```

AQUÍ SE TIENE LA FORMULA DE LA DISTANCIA DE MANHATTAN

PARA VER LOS RESULTADOS VAMOS A EJECUTAR LA FUNCION (manjathan)

Después de ejecutar nos muestra los siguientes resultados:

Python 2.7.10 Shell

File Edit Shell Debug Options Window Help

```
Python 2.7.10 (default, May 23 2015, 09:40:32) [MSC v.1500 32 bit (Intel)] on wi
n32
Type "copyright", "credits" or "license()" for more information.
>>> ================================ RESTART ================================
>>>
>>> blogs, palabras, datos = readfile('blogdata1.txt')
>>> cluts = hcluster(datos, manjathan)
>>> type(cluts)
<type 'instance'>
>>> printclust(cluts,blogs)
-
  Mitos y Timos
  -
    -
      -
        -
          Cholonymous: Blog Peruano de Actualidad y Tecnología
          Xataka Ciencia
        -
          Genbeta Dev
          -
            Astrofísica    y    Física
            -
              Curistoria - Curiosidades y anécdotas históricas
              Experimentos caseros
      -
        Hipertextual
        -
          -
            PC World Perú
            Tecnología 7
          -
            -
              Imagen astronomía diaria - Observatorio
              -
                Naukas
                Historias de la Historia
```

```
            -
              -
                Eureka
                -
                  La mentira esta ahi fuera
                  PC World en Español
                -
                EspacioCiencia.com
                -
                  -
                    Tecnología Obsoleta
                    FayerWayer
                  -
                    La Ciencia y sus Demonios
                    -
                      Círculo Escéptico Argentino
                      La Ciencia para todos
    -
      El retorno de los charlatanes
      Hispasec @unaaldia
>>>
```