Práctica 4

1. Patrones de Hearst

En el archivo "wikipedia_es_abstracts.txt" está el *dump* de un millón y medio de resúmenes de la Wikipedia en Español. Diseñe un script que reciba una consulta simple y regrese un árbol con las relaciones de hiponimia en los resumenes relacionados a dicha consulta. Para ello, será necesario:

1. Encontrar patrones léxico-sintácticos que definan relaciones de hiperonimia-hiponimia en Español
   - Les recomiendo seguir el protocolo de Hearst usando el archivo provisto.
2. Por cada patrón, definir una expresión regular que lo describa.
   - Contemplen frases nominales (NP) simples.
3. Crear un script "relaciones.py" que reciba como consulta un número no acotado de palabras clave, separadas por espacios.

   `python relaciones.py subcadena1 subcadena2 ... subcadenaN`

   a) El script seleccionará sólo los resumenes que contengan **todas** las subcadenas (`subcadena1, subcadena2, ..., subcadenaN`).

   b) En los resumenes seleccionados, el script buscará relaciones de hiponimia mediante los patrones lexico-sintácticos que definieron en el paso 1.

   c) Con las relaciones encontradas, construirán una red semántica que desplegarán mediante el paquete Networkx (`pip install networkx`).

Un ejemplo de un programa con Networkx se presenta en el Listing 1.

```
# -*- coding:utf-8 -*-

import networkx as nx
import matplotlib.pyplot as plt
import random

# Crea grafica
G = nx.DiGraph()

# Vertices
G.add_nodes_from(range(1,7))

# Aristas
```

```
aristas = [(1,5), (2,1), (2,3), (2,5),
           (3,4), (4,5), (6,4)]
G.add_edges_from(aristas)

pos=nx.spring_layout(G)
nx.draw_networkx_labels(G, pos,
                        labels=dict([(i,i) for i in range(1,7)]))

# Dibuja la gráfica
nx.draw(G,pos)

# Muestra en pantalla lo dibujado
plt.show()
```

Listing 1: Gráfica dirigida con Networkx

1.1. Construcción de jerarquía

Ejecuten el script de la siguiente manera: `python relaciones.py matemática`

- Guarden la imagen resultante.

2. Repositorio

Guarden su código en GitHub. En el README del repositorio:

1. Desplieguen la imágen de la jerarquía para la cadena `matemática`.