# Práctica III– Lenguajes de programación

## Contexto

Frigyes Karinthy propuso en su cuento “Cadenas” de 1930 la teoría de los seis grados de separación. Esta teoría propone que una persona puede contactar a cualquier otra persona del mundo usando un máximo de cinco contactos. Esta teoría se ha intentado probar por varias compañías como IBM y Facebook, llegando a diversos resultados. Si bien es una teoría planteada hace muchos años, con el auge de internet en la primera década de los 2000, este problema resurgió.

## Enunciado

En esta práctica emularán un experimento que determine si una persona (p1) está conectada con otra (p2). Esta conexión se determinará a partir de una red de contactos que se formará de manera automática con base en las personas que aparecen en las páginas de Wikipedia de p1, p2 y sus contactos derivados.

### Ejemplo 1: verificar conexión entre Barack Obama y Osama bin Laden

- Si revisamos la página de Wikipedia de Barack Obama nos encontramos diversas menciones a Osama bin Laden de manera directa, así que podemos concluir que están conectados.

### Ejemplo 2: verificar conexión entre Ian Gillan (vocalista de Deep purple) y Diomedes Díaz (Cantante de vallenato)

- En este caso, si se verifica la página de Wikipedia de Ian Gillan no hay menciones directas a Diomedes Díaz, así que se deben analizar sus conocidos con base en la información registrada en esta página. Una conexión posible es la siguiente: Ian Gillan (p1)->Jerry Lee Lewis (Cantante y pianista)->Jimmy Page (Guitarrista de Led Zeppelin)->David Beckham (Futbolista)->Victoria Beckham (Cantante)->Jennifer López (Cantante)->Marc Anthony (Cantante)->Maluma (Cantante)->Andrés Cepeda (Cantante) -> Jorge Celedón (Cantante) -> Diomedes Díaz (p2). Esto implica que sí están conectados y que un camino posible involucra a las otras personas mencionadas.

### Ejemplo 3:

https://www.tiktok.com/@tioatiok/video/7092901391491976454?is_from_webapp=v1&item_id=7092901391491976454

### Ejemplo 4:

https://www.tiktok.com/@tpshake/video/6907772678397070597?is_from_webapp=v1&item_id=6907772678397070597 (Este ejemplo tiene la misma intención, sin embargo, se hace sobre conceptos)

### Enfoque

Hacer este proceso de manera manual es altamente demandante, pues en la página de Wikipedia de una persona famosa puede haber menciones y enlaces a muchas otras personas, así que esta tarea se debe automatizar:

- Para hacerlo, crearán un *web scraper* que se encargue de extraer todos los enlaces de la página de Wikipedia de p1 y visitar los enlaces que hagan referencia a una persona.
- Para determinar si el texto del enlace se refiere a una persona usarán *Named Entity Recognition*, lo que les permitirá visitar sólo enlaces relevantes.
- Para cada uno de estos enlaces, deberán hacer el mismo proceso hasta encontrar una mención a p2 o hasta llegar a el límite predefinido.
- Como esta búsqueda puede ser muy profunda, definirán un límite de 50 personas analizadas, así que, si después de analizar la página de Wikipedia de la persona 50 no se ha encontrado una mención a p2, se determinará que p1 y p2 no están conectados.
- Para almacenar la información debe usar un grafo dirigido con la representación que prefiera (lista de adyacencia o matriz de adyacencia). En este grafo, cada persona será un nodo y se conectará con cada una de las otras personas mencionadas en su página de Wikipedia.
- A medida que se avanza en la extracción de información, el grafo deberá crecer con base en los nuevos nodos (personas) y conexiones (aristas) que se van agregando.
- Su programa deberá recibir el nombre de dos personas o la *URL* de sus páginas de Wikipedia y deberá retornar el listado de personas más corto que conecten a p1 y a p2 si es que existe una conexión (camino) o el mensaje “no están conectados” si no existe conexión entre p1 y p2.
- Adicionalmente, deberá retornar el grafo con toda la información extraída.
- Librerías recomendadas para Python: Beautiful Soup (*Web scraping*) y SpaCy (*Named Entity Recognition*).

## Un par de ayudas con las librerías

*Named Entity Recognition*: https://www.youtube.com/watch?v=Gn_PjruUtrc

*Web scraping*: https://www.youtube.com/watch?v=YY5skv756pc

## Entrega y sustentación

- 8 de noviembre 6:00 a.m.
- Se pueden hacer solos o en parejas.
- La sustentación será práctica, es decir, involucrará adiciones a la solución propuesta.
- El código deberá estar montado en un Github o Replit por grupo.
- Pueden usar el lenguaje de programación y paradigma de su preferencia.