

## Tarea Recuperativa

### 1. Introducción

En esta tarea vas a computar el PageRank de un grafo del mundo real. La idea es que extraigas un grafo del *endpoint* de Wikidata, y luego lo cargues a un sistema en el que puedas calcular el PageRank de cada nodo del grafo.

### 2. Red de citación de artículos relacionados con el virus Zika

En el endpoint de Wikidata<sup>1</sup> vamos a rescatar la red de citación de artículos relacionados con el virus Zika. Esto es, un grafo donde los nodos son artículos científicos, en que un artículo *a* cita a un artículo *b* si el artículo *a* cita al *b*. De esta forma, podemos ver cuáles son los artículos más importantes de esta red utilizando el algoritmo de PageRank. Para obtener la red de citación, puedes correr la siguiente consulta.

```
1 SELECT ?node ?cite WHERE {  
2   ?node wdt:P31 wd:Q13442814 . ?node wdt:P921 wd:Q202864 .  
3   ?node wdt:P2860 ?cite .  
4   ?cite wdt:P31 wd:Q13442814 . ?cite wdt:P921 wd:Q202864 .  
5 }
```

El resultado de esta consulta lo puedes descargar (por ejemplo, como `.csv`). La idea de esta actividad es que entregues el (1) nombre de los 10 artículos más importantes, según el PageRank en esta red de citación, (2) junto con los autores de cada uno de los artículos. **Ojo!** Tiene que ser el nombre de los artículos y autores, no las URI, porque tienes que pensar que esto se lo estás presentando a alguien, y en el mundo real no es posible hacer mucho con la URI.

Queda a criterio tuyo cómo resuelves esta actividad, lo que tienes que tener presente es que sí o sí tienes que calcular el PageRank de toda la red. Luego para obtener los autores asociado a cada artículo, puedes usar el endpoint de Wikidata, o bien, buscando tu mismo el nombre de los artículos y sus respectivos autores.

---

<sup>1</sup><https://query.wikidata.org/>.

Lo que tienes que entregar es **un informe en formato .pdf** en el que incluyas la respuesta a la pregunta; es decir, el nombre de los 10 artículos con mayor PageRank, ordenados de mayor a menor, junto con el nombre de los autores de cada artículo. Además, debes explicar en a lo más una página cómo resolviste la tarea, y además, anexar el código o procedimiento con el que computaste el algoritmo de PageRank sobre la red.

**Detalles Académicos.** La tarea se entrega el día 11 de julio hasta las 20:00 horas. Pueden realizar la tarea en grupos de hasta dos personas. La entrega es un archivo en formato **.pdf** a través de un buzón en Webcursos.