

## Programación III - Universidad Nacional de General Sarmiento

### Trabajo Práctico 2: Clustering humano

---

El objetivo del trabajo práctico es implementar una aplicación para identificar automáticamente grupos de personas sobre la base de sus características. Tenemos una lista de personas, y para cada persona  $i$  tenemos su nombre y los siguientes datos:

- $d_i$  = Interés por los deportes
- $m_i$  = Interés por la música
- $e_i$  = Interés por las noticias del espectáculo
- $c_i$  = Interés por la ciencia

Cada uno de estos datos se expresa como un entero entre 1 y 5, siendo 1 el menor interés y 5 el máximo interés en el tema. Buscamos dividir a esta población en dos grupos con intereses similares. Para esto, utilizamos el siguiente algoritmo:

1. Construir un grafo completo  $G$  con un vértice por cada persona, una arista entre cada par de personas, y de modo tal que cada arista tiene peso igual al *índice de similaridad* entre las dos personas.
2. Construir un árbol generador mínimo  $T$  de  $G$ .
3. Eliminar la arista de mayor peso del árbol  $T$ .
4. Las dos componentes conexas del grafo resultante son los dos grupos que estamos buscando.

El *índice de similaridad* entre dos personas  $i$  y  $j$  se define como

$$\text{similaridad}(i, j) = |d_i - d_j| + |m_i - m_j| + |e_i - e_j| + |c_i - c_j|.$$

Este valor es cero cuando  $i$  y  $j$  tienen exactamente los mismos valores de interés en los cuatro temas, y es más alto para las personas con intereses distintos.

La aplicación implementada debe contar con una interfaz para cargar y visualizar los datos de las personas, y una interfaz para ver los dos grupos que se formaron luego del algoritmo anterior. Se puede contar con un botón para lanzar la ejecución del algoritmo.

Como objetivos opcionales no obligatorios, se pueden contemplar los siguientes elementos:

1. Mostrar estadísticas de los dos grupos generados, como por ejemplo similaridad promedio, promedio de interés en cada tema, etc.
2. Ejecutar el algoritmo para recalculer los grupos cada vez que se agrega una nueva persona a la lista.
3. Permitir que se generen más de dos grupos, preguntándole al usuario cuántos grupos quiere generar.

**Condiciones de entrega:** El trabajo práctico se debe entregar por mail a los docentes de la materia. Además del código, se debe incluir un documento en el que se describa la implementación y se detallen las decisiones tomadas durante el desarrollo. El código responsable del funcionamiento de la aplicación debe estar claramente separado del código responsable de la interfaz. Todas las clases de negocio deben tener tests unitarios. El trabajo práctico se puede hacer en grupos de hasta tres personas.

**Fecha de entrega:** Martes 10 de noviembre.