

## **Enunciat del projecte de PROP**

### **Quadrimestre de primavera, curs 16/17**

#### **Extractor de prototipus de comportament o perfils**

Hem fet una enquesta i volem fer un programa per estudiar els diferents prototipus de comportament o perfils de les persones que l'han contestada. Més concretament, voldriem fer-nos una idea de quants comportaments/perfils hi ha, i com seria l'individu representatiu de cada comportament o perfil existent. Considerem que dues persones tenen un comportament similar si les respostes a l'enquesta també són similars, de manera que podem "agrupar" les persones per la semblança en les seves respostes.

Les dades de l'enquesta les tindrem usualment en un fitxer de text. El programa ha de poder tractar enquestes on les preguntes siguin del tipus "escolliu una d'aquestes opcions o com a màxim  $n$  opcions, o bé donar una resposta com un string en format lliure, o bé donar una quantitat numèrica". Tot i així, ens podem trobar respostes no contestades. Com a mínim, s'ha de poder tractar amb dades com les que ens trobem a les enquestes del CIS:

[http://www.cis.es/cis/opencm/ES/11\\_barometros/preguntasfijas.jsp](http://www.cis.es/cis/opencm/ES/11_barometros/preguntasfijas.jsp)

Volem una aplicació que permeti, com a mínim, les següents funcionalitats:

- Creació interactiva d'enquestes i importació d'enquestes
- Resposta interactiva i importació de respostes
- Gestió completa de enquestes i respostes (esborrar, modificar, etc.)
- Anàlisi d'enquestes (obtenció dels diferents comportaments o perfils, obtenció dels representants dels diferents grups, llistat de tots els individus de cada comportament o perfil, etc.)

#### **Dates dels lliuraments:**

Primer: divendres, 21 d'abril

Segon: divendres, 2 de juny

#### **Informació addicional:**

Es tracta d'un problema de *clustering*. Hi ha dues grans famílies d'algorismes de clustering: algorismes de clustering jeràrquic (que poden ser aglomeratius o divisius) i algorismes de clustering particional (algorisme *k-means*, algorisme *nearest-neighbour*, etc.). S'ha de resoldre el problema plantejat implementant un algorisme de clustering jeràrquic o un algorisme de clustering particional com *k-means* (ambdós en les seves versions més simples). Les següents referències contenen informació al respecte:

[http://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/](http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/)

[https://en.wikipedia.org/wiki/Cluster\\_analysis](https://en.wikipedia.org/wiki/Cluster_analysis)

[https://en.wikipedia.org/wiki/Hierarchical\\_clustering](https://en.wikipedia.org/wiki/Hierarchical_clustering)

[https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering)