

# Disseny D'Estructures de Dades

## PAC 1: Primera prova d'avaluació continuada

### Competències

Les competències que treballareu a la PAC són les següents:

#### Específiques

- Capacitat per analitzar un problema en el nivell d'abstracció adequat a cada situació i aplicar les habilitats i coneixements adquirits per a resoldre'l.
- Capacitat per dissenyar i construir aplicacions informàtiques mitjançant tècniques de desenvolupament, integració i reutilització.
- Capacitat per proposar i avaluar diferents alternatives tecnològiques per resoldre un problema concret.

#### Transversals

- Ús i aplicació de les TIC en l'àmbit acadèmic i professional.
- Capacitat per a innovar i generar noves idees.

### Objectius

Els objectius que es persegueixen en el desenvolupament de la PAC són els següents:

- Entendre el concepte de TAD i saber-ne fer l'especificació.
- Conèixer la biblioteca de TAD de l'assignatura i saber utilitzar-la per dissenyar i implementar noves estructures de dades.

- Saber calcular l'eficiència espacial i temporal d'una estructura de dades i dels algorismes associats per tal de comparar diferents alternatives i poder triar-ne la millor en termes d'eficiència (temporal o espacial)
- Ser capaç d'identificar l'estructura de dades utilitzada en un programa i entendre'n el funcionament.
- Entendre el funcionament dels contenidors seqüencials presentats a l'assignatura. Saber quan i com utilitzar aquests contenidors.

## Descripció de la PAC a realitzar

La PAC consta de 4 exercicis, alguns més teòrics i altres més pràctics, en els que posareu en pràctica els coneixements adquirits en l'estudi dels tres primers mòduls de l'assignatura.

## Recursos

Els recursos necessaris per a desenvolupar la PAC són els següents:

### **Bàsics** (material didàctic de l'assignatura)

- Mòdul 1: Tipus abstractes de dades.
- Mòdul 2: Complexitat algorísmica.
- Mòdul 3: Contenidors seqüencials.

## Complementaris

Llibreria de TAD de l'assignatura.

## Criteris de valoració

La puntuació global de la PAC és la suma de les puntuacions individuals obtingudes a cada un dels exercicis que la formen. La puntuació individual de cada exercici s'especifica a cadascun d'ells.

## Format i data de lliurament

### Data de publicació

5 d'octubre de 2020.

### Data de lliurament

22 d'octubre de 2020.

### Format de lliurament

El lliurament cal fer-lo a través de l'espai de "Lliurament i registre d'AC" amb un únic fitxer preferentment en format PDF o, si no és possible, Word o OpenOffice. Aquest fitxer contindrà la solució. Si us plau, no hi copieu l'enunciat, feu-hi constar el vostre nom a cada pàgina (per exemple, amb un peu de pàgina), i numereu les pàgines.

## Enunciat

Un grup d'alumnes de DED està preparant una eina que permeti ajudar a gestionar la farmacovigilància associada a les vacunes de la COVID-19 que s'estan desenvolupant actualment. En aquest moment existeixen diferents projectes que intenten produir vacunes i perquè siguin possible utilitzar-les en humans han de passar per un estricte procés de validació. Aquest procés de validació té diferents nivells. L'eina que proposen els alumnes de DED està centrada en el nivell 3 de farmacovigilància, en la qual es monitoritzen diferents grups d'usuaris realitzant preguntes de manera recurrent per a saber si tenen mal de cap, febre, indisposició, etc.

**Trial4C19** és una plataforma de farmacovigilància que té com a principal objectiu minimitzar la gestió de farmacovigilància i donar suport als grups que estan desenvolupant projectes associats a vacunes de COVID-19. Per fer-ho, cal dissenyar una estructura de dades per a **Trial4C19**. En aquesta PAC farem una prova pilot d'una part del sistema (un conjunt reduït de funcionalitats) i treballarem amb volums d'informació petits perquè els assajos clínics es puguin realitzar en petit format abans de fer un llançament més complex.

En aquesta PAC us familiaritzareu amb el problema a resoldre i practicareu una mica el disseny i la composició d'estructures de dades per donar forma i solució al problema plantejat. Concretament, per a la realització de l'exercici considereu que:

- El nombre d'usuaris (U) que realitzaran els assajos clínics és conegut i relativament petit, d'uns centenars.
- El nombre d'assajos clínics (T) és conegut i relativament petit i afitat (d'unes desenes).
- El nombre de grups de preguntes (G) serà petit, no variarà molt i és relativament petit, d'unes desenes.
- El nombre de preguntes dintre d'un grup de preguntes (QG) serà petit però en constant augment. Una pregunta només pot estar en un únic grup de preguntes.
- El nombre de grups de preguntes d'un assaig (QT) serà petit, però variable.

- El nombre d'usuaris que participen en un assaig concret (UT) és conegut i relativament petit, d'uns centenars.
- El nombre de respostes que realitza un usuari en un assaig concret (AUT) serà petit, però en constant augment.

En general, en tot l'exercici, s'ha d'optimitzar l'eficiència de les operacions de consulta encara que això pugui penalitzar les operacions d'escriptura.

## Exercici 1 (2'5 punts)

Especifiqueu un TAD **Trial4C19** que permeti les següents operacions:

- Afegir un nou usuari al sistema. De cada usuari sabrem el seu identificador, el seu nom i els seus cognoms. Si ja existeix un usuari amb aquest identificador, actualitzem les seves dades.
- Afegir un nou assaig clínic. De cada assaig clínic sabrem el seu identificador i una descripció. Si ja existeix l'assaig, retornarà un error.
- Afegir un grup de preguntes. De cada grup de preguntes sabrem el seu identificador i la seva prioritat. Si ja existeix un grup de preguntes, actualitzarem les seves dades.
- Afegir una nova pregunta en el sistema especificant el codi identificador de pregunta, l'identificador del grup de preguntes al que pertany la pregunta, el text de la pregunta i tipus de respostes (text lliure o escala Likert). Si ja existeix el codi identificador, actualitzarem les seves dades. Si un grup de preguntes no existeix, s'haurà d'informar de l'error.
- Consultar les preguntes d'un grup de preguntes a partir de l'identificador del grup. Si no existeix l'identificador del grup de preguntes, s'haurà d'informar de l'error.
- Assignar un grup de preguntes a un assaig clínic. Si el grup de preguntes o l'assaig clínic no existeixen, es retornarà l'error.
- Assignar un usuari a un assaig clínic. Considereu que l'usuari i l'assaig clínic sempre existeixen. Si l'usuari ja està assignat a un assaig clínic, es retornarà un error.
- Consultar la pregunta actual que un usuari ha de respondre. Els grups de preguntes tenen una prioritat i s'han de respondre primer a les preguntes dels grups de preguntes amb més prioritat.

- Respondre a una determinada pregunta per part d'un usuari en un moment donat. Si l'usuari o la pregunta no existeixen, es retornarà un error.
- Consultar les respostes d'un usuari ordenades per la data en què es va realitzar. Si l'usuari no existeix s'haurà de mostrar l'error. Si no hi ha respostes realitzades per l'usuari s'haurà d'indicar.
- Consultar l'usuari més actiu d'un assaig clínic. Es proporciona l'usuari dins d'un assaig clínic que més respostes ha realitzat. Considereu que l'assaig clínic sempre existeix.
- Consultar l'assaig clínic més actiu. Es proporciona l'assaig clínic amb més respostes realitzades.

### Apartat a) [1 punt]

Doneu la signatura del TAD **Trial4C19**. És a dir, indiqueu el nom que donaríeu a les operacions encarregades de cada funcionalitat requerida. Indiqueu també quins caldria que fossin els paràmetres d'entrada i el tipus de retorn en cas que es necessités.

#### Solució

- addUser(userID, name, surname)
- addTrial(trialID, description)
- addQuestionGroup(questionGroupID, priority)
- addQuestion(questionID, wording, type, groupID)
- getQuestions(groupID): Iterador
- assignQuestionGroup2Trial (groupID, trialID)
- assignUser2Trial(trialID, userID)
- getCurrentQuestion(userID)
- addAnswer(userID, questionID, date, answer)
- Iterador getAnswers(userID)
- mostActiveUser(idTrial)
- mostActiveTrial()

### Apartat b) [1,5 punts]

Feu l'especificació contractual de les operacions del TAD **Trial4C19**. En la redacció de l'especificació podeu fer servir, si us cal, qualsevol de les operacions del TAD. Preneu com a model l'especificació de l'apartat 1.2.3 del Mòdul 1 dels materials docents. Es valorarà especialment la concisió (absència d'elements redundants o innecessaris), precisió (definició correcta del resultat de les operacions), completesa (consideració de tots els casos possibles en què es pot executar cada operació) i manca d'ambigüitats (coneixement exacte de com es comporta cada operació en tots els casos possibles) de la solució. És important respondre aquest apartat usant una descripció condicional i no procedimental. L'experiència ens demostra que no sempre resulta fàcil distingir entre les dues descripcions, és per això que fem especial èmfasi insistint que poseu molta atenció a les vostres definicions.

A títol d'exemple indicarem que la descripció condicional (la correcta a utilitzar en el contracte) d'omplir un got buit amb aigua seria:

@pre el got es troba buit.

@post el got és ple d'aigua.

En canvi una descripció procedimental (incorrecta per utilitzar en el contracte) tindria una forma semblant a:

@pre el got hauria de trobar-se buit, si no es trobés buit s'hauria de buidar.

@post s'acosta el got a l'aixeta i s'hi tira aigua fins que estigui ple.

Cal també tenir en compte que un contracte hauria de disposar d'invariant sempre que aquest fos necessari per descriure el TAD.

## Solución

@pre cert.

@post si el codi d'usuari és nou, els usuaris seran els mateixos més un nou usuari amb les dades indicades. Sinó les dades de l'usuari s'hauran actualitzat amb els nous.

- addUser(userId, name, surname)

@pre cert.

@post Si el codi de l'assaig clínic no existeix els assajos clínics seran els mateixos més un de nou amb les dades indicades. Si no, s'haurà d'informar de l'error.

- addTrial(trialID, description)

@pre cert

@post si el codi de grup de preguntes és nou, els grups de preguntes seran els mateixos més un de nou amb les dades indicades. Si no, les dades del grup de preguntes s'hauran d'actualitzar amb les noves.

- addQuestionGroup(questionGroupID, priority)

@pre cert

@post Si el codi de pregunta és nou, les preguntes seran les mateixes més una nova. Si no, les dades de la pregunta s'actualitzaran amb les noves. Si el grup de preguntes no existeix, s'haurà d'informar de l'error.

- addQuestion(questionID, wording, type, choices, groupID)

@pre cert.

@post retorna un iterador per recórrer les preguntes d'un grup de preguntes.

- Iterador questions(groupID)

@pre

@post s'actualitza l'estructura de preguntes d'un assaig clínic. Si el grup de preguntes o assaig clínic no existeix, s'haurà d'informar de l'error

- assignQuestionGroup2Trial(questionGroup, trialID)

@pre l'usuari i l'assaig clínic existeixen

@post s'actualitza l'estructura d'usuaris d'un assaig clínic. Si l'usuari ja està en un assaig clínic, s'haurà d'informar de l'error

- assignUser2Trial(questionGroup, trialID)



@pre cert.

@post retorna la següent pregunta a respondre per un usuari.

- Question getCurrentQuestion(userID)

@pre cert.

@post el nombre de respostes d'un usuari serà el mateix més la nova resposta. Si l'usuari o la pregunta no existeix, s'haurà d'informar de l'error.

- addAnswer(userID, questionID, dati, answer)
- addAnswer(userID, dati, answer). Amb l'ajuda de l'operació getCurrentQuestion no seria necessari especificar la pregunta.

@pre cert.

@post retorna un iterador per recórrer les respostes d'un usuari ordenades per data. Si l'usuari no existeix o no hi ha respostes s'haurà de mostrar l'error.

- Iterador getAnswers(userID)

@pre cert.

@post retorna un iterador per recórrer les respostes d'un usuari ordenades per data. Si l'usuari no existeix o no hi ha respostes s'haurà de mostrar l'error.

- Iterador getAnswers(userID)

@pre l'assaig clínic existeix.

@post retorna l'usuari que més respostes ha realitzat en un assaig clínic

- User mostActiveUser(trial)

@pre cert.

@post retorna l'assaig actiu que més respostes ha generat

- Trial mostActiveTrial()

Nota: En tot l'exercici hem especificat com a precondicions les comprovacions que es donaven per assumides segons l'enunciat i com a comprovacions a realitzar dins de les operacions aquelles que havien de retornar un error.

## Exercici 2 (3'5 punts)

A l'exercici 1 heu definit l'especificació d'un nou TAD, el TAD **Trials**. Ara us demanem que feu el disseny de les estructures de dades que formaran aquest TAD. Dissenyau, doncs, el sistema per tal que sigui el màxim d'eficient possible, tant a nivell d'eficiència espacial com temporal, tenint en compte els volums d'informació i les restriccions especificades a l'enunciat.

**Per a la realització de la PAC1 us haureu de limitar als TAD del mòdul 3 dels materials. Queden descartats, per tant, els arbres i les taules de dispersió.**

Tingueu en compte només les operacions que es demanen a l'enunciat a l'hora de fer aquest disseny.

### Apartat a) [0,5 punts]

Dubtem entre utilitzar un vector, una llista encadenada, o una llista encadenada ordenada per a emmagatzemar les **preguntes**. Justifiqueu quina creieu que és la millor opció.

#### Solució

Per guardar les preguntes és millor utilitzar una estructura dinàmica ja que el nombre de preguntes és variable (i està en constant augment) i no ho coneixem a priori. Una estructura estàtica ens faria malgastar espai.

Dins de les estructures dinàmiques podem triar entre una llista encadenada o una llista encadenada ordenada. En aquest cas ens decantem per la **llista encadenada**, ja que l'operació d'alta d'una nova pregunta no haurà de reordenar i no tenim cap operació que hagi de retornar les preguntes de forma ordenada. En qualsevol cas, amb l'existència del grup de preguntes, no existirà un contenidor general de preguntes, ja que cada pregunta estarà classificada dins d'un grup.

## Apartat b) [0,5 punts]

Dubtem entre utilitzar una estructura estàtica o una estructura dinàmica per a emmagatzemar les **assajos clínics**. Suposem que volem una eficiència constant respecte el nombre d'assajos clínics en l'operació de consulta. Justifiqueu quina creieu que és la millor opció i digueu quina estructura concreta faríeu servir.

### Solució

Per guardar els assajos clínics és millor utilitzar una estructura estàtica ja que el nombre d'assajos clínics és conegut, petit i fitat. L'ús d'una estructura dinàmica com una llista encadenada o una llista encadenada ordenada ens faria malgastar espai pels apuntadors i l'eficiència temporal de les operacions de consulta seria lineal.

Dins de les estructures estàtiques podrem triar entre un vector ordenat que ens permetrà consultar els assajos amb eficiència logarítmica mitjançant cerca dicotòmica i un simple vector on cada assaig clínic ocupa la posició que defineix el seu identificador. Així, la cerca d'un assaig clínic la podrem fer accedint directament a la posició del vector (no tindrem ni que fer cerca dicotòmica) i tindrà cost constant  $O(1)$ .

L'elecció serà doncs un **vector** on cada assaig clínic ocupa la posició que defineix el seu identificador.

## Apartat c) [0,5 punts]

Dubtem entre utilitzar un vector, un vector ordenat, o una llista encadenada ordenada per a emmagatzemar els **usuaris** tenint en compte que ens interessa millorar els costos lineals per a les operacions de consulta. Justifiqueu quina creieu que és la millor opció.

### Solució

L'enunciat ens diu que el nombre d'usuaris  $O$  és relativament petit i conegut. Amb volums de dades al voltant del centenar l'elecció de l'estructura de dades no és crítica. Així doncs, qualsevol estructura de dades seria acceptable ja que no representaria una pèrdua d'eficiència important. No obstant això, l'estructura més adequada seria un vector ordenat pel seu identificador, ja que ocupa l'espai just i necessari i permet realitzar cerques dicotòmiques  $O(\log U)$  per localitzar a l'usuari a partir del seu identificador. En un vector sense ordenar hauríem de fer recorreguts lineals fins a trobar l'usuari  $O(U)$ . Les llistes encadenades, malgrat ser ordenades, no permeten realitzar cerques dicotòmiques i, per tant, també hauríem de fer recorreguts lineals  $O(U)$  a part de malbaratar espai amb els apuntadors.

### Apartat d) [1 punt]

Justifiqueu la resta d'estructures de dades per fer el disseny del TAD. La justificació de cada una de les estructures de dades ha de ser de l'estil:

“Per guardar XXX triem una llista encadenada ordenada ja que el número d'elements no és gaire gran, ens fa falta accés directe i ens calen recorreguts ordenats.”

#### Solució

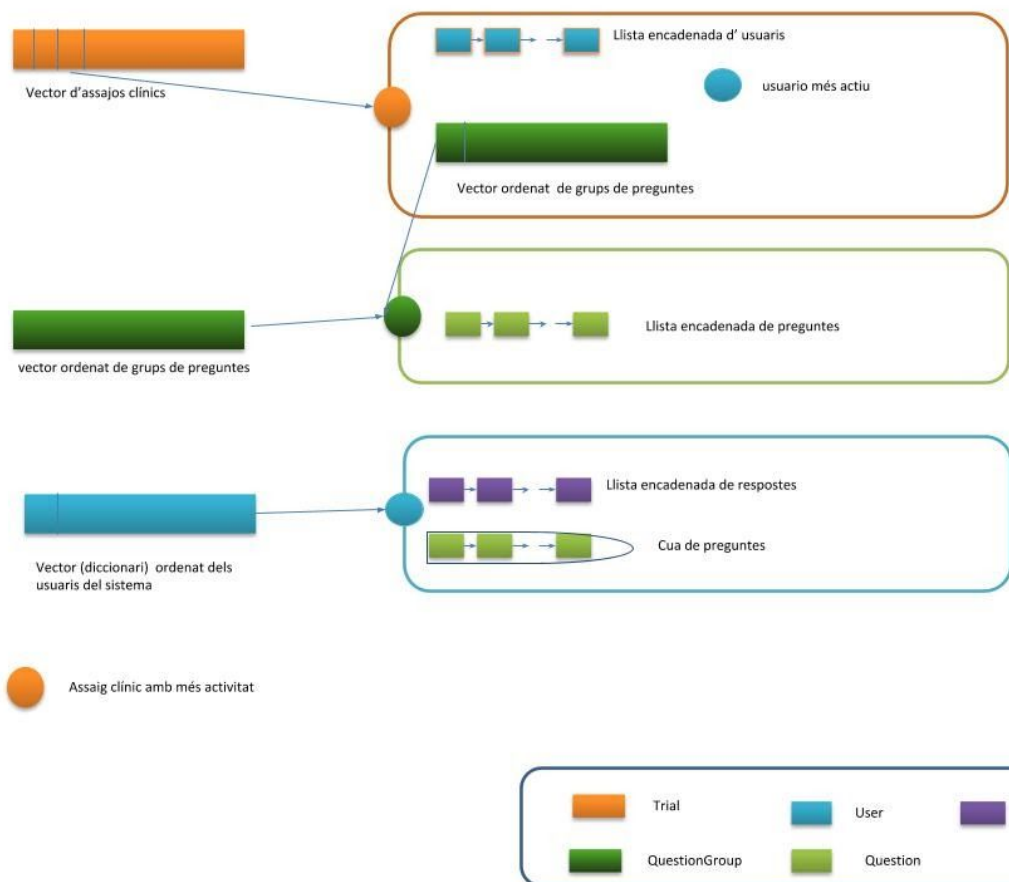
- Per guardar els grups de preguntes usarem un vector ordenat (per prioritat), ja que es tracta un nombre petit, no variarà molt i està fitat.
- Per guardar els usuaris d'un assaig clínic usarem una llista encadenada, ja que el nombre d'usuaris és petit, variable i no necessitem ordenació.
- Per guardar les preguntes (dins d'un grup) d'un assaig clínic usarem una llista encadenada, ja que el nombre de preguntes és petit, però en constant augment, no es poden fitar, i necessitem ordenació.
- Per guardar les respostes d'un usuari dins d'un assaig clínic usarem una llista encadenada ja que aniran en constant augment i inicialment estarà buida. El propi sistema manté l'ordre ja que va realitzant les preguntes en ordre.
- Per guardar les preguntes a realitzar per part de cada usuari s'utilitzarà una cua de preguntes que s'iniciarà al moment de l'assignació a un assaig clínic. Es pren la hipòtesi de partida que els usuaris tinguin precargadas

les preguntes per evitar que futures modificacions de preguntes/grups de preguntes a posteriori puguin afectar a una planificació inicial d'un assaig clínic. Amb altres escenaris de partida es poden proposar altres opcions.

- Per guardar les respostes realitzades per un usuari s'utilitzarà una llista encadenada.
- Per guardar l'usuari més actiu dins d'un assaig clínic s'haurà de guardar un apuntador
- Per guardar l'assaig més actiu s'haurà de guardar un apuntador.
- Per guardar l'assaig clínic d'un usuari s'haurà de guardar un apuntador.

## Apartat e) [1 punt]

Feu un dibuix de l'estructura de dades global pel TAD **Trial4C19** on es vegin clarament les estructures de dades que trieu per representar cada una de les parts i les relacions entre elles. Cal que feu el dibuix de l'estructura complerta, amb totes les estructures que us permetin implementar les operacions definides a l'especificació.



## Exercici 3 [3 punts]

A l'exercici 1 heu definit l'especificació del TAD **Trial4C19** amb les seves operacions, i a l'exercici 2 heu triat les estructures de dades per cada part del TAD. En aquest exercici us demanem que us fixeu en els algorismes que us serviran per implementar algunes de les operacions especificades i en l'estudi d'eficiència de les mateixes. Tingueu en compte que la implementació de les operacions va estretament lligada a l'elecció de les estructures de dades que hàgiu fet.

### Apartat a) [1,5 punts]

Descriviu i feu l'estudi d'eficiència de l'operació que hàgiu definit per **respondre a una pregunta**. Per fer-ho, heu de descriure breument el seu comportament indicant els passos que la componen (amb frases com ara: "inserir al vector / esborrar de la llista encadenada / consulta de l'element a la llista encadenada ordenada..."), dient l'eficiència asimptòtica de cada pas i donant l'eficiència total de l'operació.

#### Solució

- Buscar l'usuari en el vector ordenat d'usuaris  $\Rightarrow O(\log U)$ .
- Accedir a l'assaig clínic en el qual l'usuari participa  $\Rightarrow O(1)$
- Accedir a la pregunta, mitjançant la cua de preguntes associada a l'usuari  $\Rightarrow O(1)$
- Afegir una nova resposta en la llista de respostes de l'usuari  $\Rightarrow O(1)$
- Actualitzem l'apuntador de l'usuari més actiu dins d'un assaig, si correspon  $\Rightarrow O(1)$ .
- Actualitzem l'apuntador de l'assaig més actiu, si correspon  $= O(1)$

### Apartat b) [1'5 punts]

Suposeu ara que volem maximitzar l'eficiència espacial i no tindrem una estructura de dades per gestionar els grups de preguntes. Analitzeu l'operació **respondre a una pregunta** i expliqueu quines implicacions hi han respecte al contracte especificat, les estructures de dades necessàries i si hi ha afectacions a l'eficiència de les operacions.

### Solució

**Nota:** Si s'elimina l'estructura de dades que gestiona els grups de preguntes, tenim les següents implicacions:

- S'ha de replantejar l'operació assignar un grup de preguntes a un assaig clínic, ja que no serà possible assignar tot un grup de preguntes a un assaig clínic. Per aquest motiu l'assignació de preguntes a un assaig serà més tediosa per a les persones que configuren els assajos clínics.
- L'operació afegir una pregunta en el sistema, ja no indicarà l'identificador del grup, i s'haurà d'indicar la prioritat de la pregunta. Això implica que la inserció de la pregunta tindrà en compte la prioritat especificada.
- L'operació respondre a una pregunta si es manté la cua de preguntes precargadas, no hi ha afectacions.
- L'operació consultar les preguntes d'un grup ja no té sentit. Potser es pot plantejar el reenfoque d'aquesta operació proporcionant les preguntes d'una determinada prioritat.

## Exercici 4 [1 punt]

Indiqueu quins dels TAD de la biblioteca de TAD de l'assignatura et semblen més adients per utilitzar-los en la implementació de cadascuna de les estructures de dades definides pel TAD **Trial4C19**. Si no hi ha una implementació ja feta a la biblioteca de TAD de l'assignatura, digueu com la implementaríeu.

### Solució

- Per guardar els **assajos clínics** utilitzarem un **array de Java** ja que només necessitem accés directe per posició. (Trial[] )
- Per guardar els **usuaris**, implementarem una nova classe que implementi un vector ordenat amb operacions per afegir ordenadament, i fer consultes mitjançant cerca dicotòmica. Per integrar aquesta classe a la biblioteca de classes ha d'implementar les interfícies de ContenedorAcotado i de Diccioniari.
- Per guardar **les respostes d'un usuari** dins d'un assaig clínic utilitzarem una ListaEncadenada.



- Per guardar els **grups de preguntes**, implementarem una nova classe que implementi un vector ordenat amb operacions per afegir ordenadament. Com no haurem de fer consultes no cal implementar cerca dicotòmica. Per integrar aquesta classe a la biblioteca de classes ha d'implementar la interfície ContenedorAcotado.
- Per guardar **els grups de preguntes dins d'un assaig clínic** utilitzarem també un vector ordenat
- Per guardar les **usuaris dins d'un assaig clínic** utilitzarem una ListaEncadenada.
- Per guardar **les preguntes dins d'un grup de preguntes** utilitzarem una ListaEncadenada.