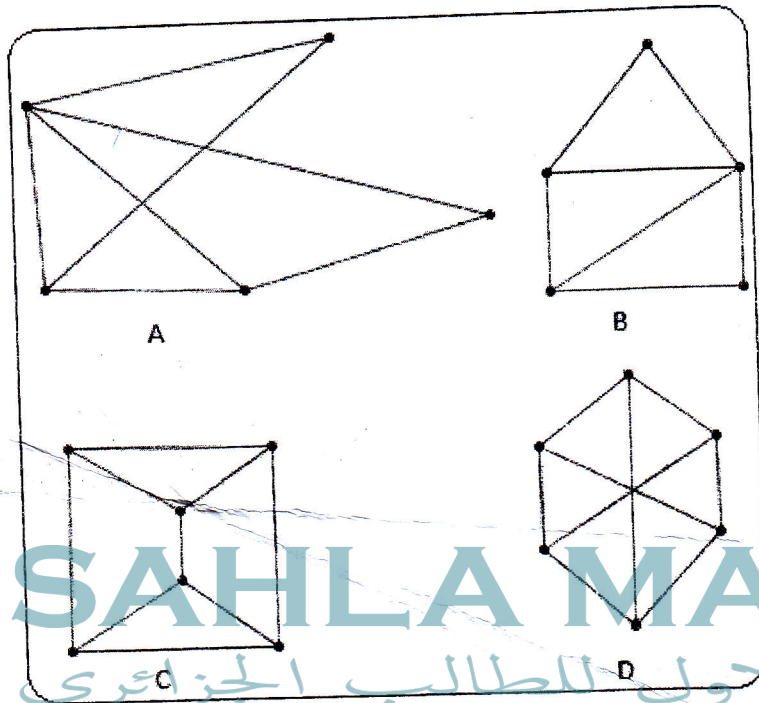


Concours d'accès au Doctorat LMD Informatique 2022-2023

Epreuve 1 : Algorithmique Avancée et Complexité
Date : 11/02/2023, Durée : 1h 30mn, 13.00 h -14.30 h

Exercice 1: <07 points>

1. Parmi les dessins suivants, lesquels représentent le même graphe, justifier votre réponse.



2. A un examen, les candidats peuvent choisir 2 ou 3 matières parmi les 6 matières en options proposées : M1, M2, M3, M4, M5, M6.

Certains élèves ont choisi les options : M1, M3, M4

D'autres : M2, M5

D'autres enfin : M3, M6.

Les élèves passent au plus une épreuve chaque jour. A l'aide de la théorie des graphes, répondez aux questions suivantes:

1- Combien peut-on programmer d'épreuves d'option au maximum dans une journée?

2- Quelle est la durée minimum de l'ensemble des épreuves optionnelles?

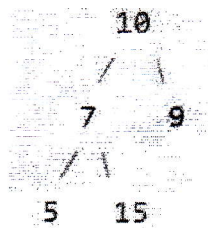
Concours d'accès au Doctorat LMD Informatique 2022-2023

Exercice 2: <07points>

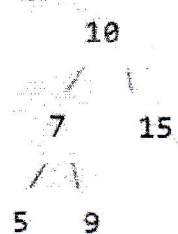
- 1) Soit la matrice M (n, m), écrire l'algorithme qui permet de permuter les deux colonnes de sommes maximale et minimale. Puis, évaluer sa complexité.
- 2) Pour chaque paire d'algorithmes A et B suivants, indiquer lequel est plus efficace en justifiant votre réponse :
 - a- Algorithme A en $O(n^2)$ et Algorithme B en $\Theta(n^2)$;
 - b- Algorithme A en $O(n^2)$ et Algorithme B en $\Theta(n \log n)$;
 - c- Algorithme A en $\Omega(n)$ et Algorithme B en $\Theta(n \log n)$

Exercice 3: <06 points>

1. On considère un arbre binaire ($A1$) illustré dans la figure ci-dessous.
En utilisant le parcours infixe (GRD), proposer un algorithme (*InsertPreOrder*) qui insère les clés de l'arbre dans un tableau T .
2. Sachant que dans cet arbre, exclusivement, deux valeurs sont permutées par erreur (9 et 15).
 - a- Proposer un algorithme (*FixBST*) qui trouve ces deux valeurs et restaure l'arbre de recherche binaire correct ($A2$).
 - b- Quelle est la complexité de cet algorithme.
3. Proposer un algorithme *FixBetterBST* de complexité $O(n)$ qui fait le même travail.



A1



A2

Bon courage



Concours d'accès au Doctorat LMD Informatique 2022-2023

Epreuve 2 : Bases de données et Data mining

Date : 11/02/2023, Durée : 2h00, 15.00 h -17.00 h

Partie 1 : Bases de données <08 points>

Exercice 1: <3 points>

Soit la relation $R(A, B, C, D, E)$ et \mathcal{DF} un ensemble de dépendances fonctionnelles défini sur les attributs de R , tel que :

$$\mathcal{DF} = \{ AB \rightarrow C, BC \rightarrow D, CD \rightarrow E, BCE \rightarrow F \}$$

Question : En utilisant les axiomes d'Armstrong, montrer que :

- 1) $AB \rightarrow E$
- 2) $AB \rightarrow F$

Exercice 2: <5 points>

Soit le schéma relationnel de la base de données de gestion d'une bibliothèque :

Livre (livre_id, titre, nombre_copies)

Personne(personne_id, nom, sexe)

Emprunt(#personne_id, #livre_id, date_emprunt, date_retour)

Sachant que:

- i) Pour chaque livre, la bibliothèque possède une ou plusieurs copies (**nombre_copies**).
- ii) Pour une même date d'emprunt (**date_emprunt**), une personne peut emprunter une seule copie du même livre, mais elle peut l'emprunter une autre date.
- iii) Avant de retourner une copie d'un livre, sa date de retour (**date_retour**) reste égale à **NULL**.
- iv) La valeur du sexe de la personne (**sexe**) a deux valeurs possibles 'homme' ou 'femme'.

Remarque : Les clés primaires sont soulignées et les clés étrangères sont précédées par le symbole (#)

Questions :

1- Qu'elle est la clé primaire de la table **Emprunt** ?

2- Exprimer à l'aide du langage SQL les requêtes suivantes :

R1: Retourner, pour chaque livre, le nombre de copies déjà **empruntées** et **non encore retournées**.

R2: Retourne les titres des livres qui ont été empruntés par des personnes de sexe 'homme'.

R3: Retourne les titres des livres qui ont été emprunté uniquement par des hommes.

R4: Retourne les titres des livres qui ont été emprunté à la fois par des personnes des deux sexes.



Concours d'accès au Doctorat LMD Informatique 2022-2023

Partie 2 : Data mining <12 points>

Exercice 3: <4.5 points>

Considérons l'approche suivante pour tester si un classifieur A est plus performant qu'un classifieur B. Soit N la taille d'un ensemble de données (dataset). P_A l'accuracy du classifieur A, P_B l'accuracy du classifieur B et $p = \frac{P_A + P_B}{2}$ l'accuracy moyenne des deux classifieurs A et B. Pour tester si un classifieur A est plus performant que B, on utilise la formule suivante :

$$Z = \frac{P_A - P_B}{\sqrt{\frac{2p(1-p)}{N}}}$$

On considère que le classifieur A est plus performant que le classifieur B si $Z > 1.96$. Le classifieur A est aussi performant que le classifieur B si $-1.96 \leq Z \leq 1.96$. Enfin, le classifieur A est moins performant que le classifieur B si $Z < -1.96$

Le tableau suivant compare trois différents classifieurs sur différents datasets selon leur accuracy. Les classifieurs sont : Arbre de Décision, Bayésienne Naïve et Support Vector Machine (SVM).

Dataset	Taille (N)	Arbre de Décision (%)	Bayésienne Naïve (%)	SVM (%)
Australia	690	85.51	76.81	84.78
Auto	205	81.95	58.05	70.73
Breast	699	95.14	95.99	96.42
Diabetes	768	72.40	75.91	76.82
Labor	57	78.95	94.74	92.98
Vehicule	846	71.04	45.04	74.94

On vous demande de résumer dans le tableau suivant la performance de chacun des trois classifieurs. Chaque cellule dans le tableau donne le nombre de datasets où le classifieur A est plus performant que B (+), moins performant que B (-), et aussi performant que B (=).

Performance	Arbre de décision			Bayésienne Naïve			SVM		
	+	-	=	+	-	=	+	-	=
Arbre de décision	0	0	6						
Bayésienne Naïve				0	0	6			
SVM							0	0	6



Concours d'accès au Doctorat LMD Informatique 2022-2023

Exercice 4: <7.5 points>

Un développeur de logiciels cherche à acheter un ordinateur portable à haute performance à des fins professionnelles. Il collecte des informations auprès de ses collègues développeurs qui ont déjà effectué un achat d'ordinateur portable et évalue leur niveau de satisfaction. Dans ce problème, le développeur tente de construire un arbre de décision pour prédire le niveau de satisfaction en fonction des caractéristiques de l'ordinateur portable futur.

Dans le tableau suivant, vous pouvez trouver les données de 500 transactions de développeurs achetant des ordinateurs portables. Les caractéristiques de chaque transaction sont CPU ("Intel" ou "AMD"), carte graphique ("Oui" ou "Non"), RAM ("Haute") et SSD ("Oui", "Non"). Chaque ligne montre les valeurs des caractéristiques et du niveau de satisfaction, ainsi que la fréquence de ces valeurs dans les transactions. Par exemple, il y a 90 développeurs satisfaits d'avoir acheté un ordinateur portable avec CPU Intel, carte graphique, RAM haute et SSD.

CPU	Carte Graphique	RAM	SSD	Niveau de Satisfaction	Fréquence
Intel	oui	haute	oui	satisfaire	90
AMD	oui	haute	oui	non satisfaire	95
Intel	non	haute	oui	satisfaire	80
AMD	non	haute	oui	satisfaire	80
Intel	oui	haute	non	non satisfaire	X
AMD	oui	haute	non	non satisfaire	102

1. Trouver la valeur de X.
2. En utilisant le gain en information, construire l'arbre de décision, incluant tous les détails du calcul.
3. Le développeur de logiciels a trouvé les ordinateurs portables suivants dans un magasin à proximité et souhaite utiliser l'arbre de décision construit dans la question 2 pour déterminer le niveau de satisfaction pour chaque ordinateur portable. Quel serait la sortie de l'arbre de décision?

N°	CPU	Carte Graphique	RAM	SSD	Niveau de Satisfaction
1	Intel	non	haute	Non	?
2	AMD	oui	haute	Oui	?

Bon courage