POLYTECHNIQUE MONTRÉAL

Corrigé examen final

INF2010

Sigle du cours

Q1	
Q2	
Q3	
Q4	
Q5	
Q6	
Total	

Identification de l'étudiant(e)		
Nom :	**Prénom :**	
Signature :	**Matricule :**	**Groupe :**

Sigle et titre du cours	*Groupe*	*Trimestre*
INF2010 – Structures de données et algorithmes	**Tous**	**20173**
Professeur	***Local***	***Téléphone***
Ettore Merlo, responsible Tarek Ould Bachir, chargé de cours		**5193**

Jour	*Date*	*Durée*	*Heures*
Mardi	**12 décembre 2017**	**2 h 30**	**9 h 30 – 12 h 00**

Documentation	*Calculatrice*	
☒ Aucune ☐ Toute ☒ Voir directives particulières	☐ Aucune ☐ Toutes ☒ Non programmable	**Les cellulaires, agendas électroniques ou téléavertisseurs sont interdits.**

Directives particulières

- Un cahier supplémentaire vous sera remis. Servez-vous de ce cahier comme brouillon. Toutes vos réponses doivent être faites sur le questionnaire. Le cahier supplémentaire n'est pas à remettre à la fin de l'examen.

Important

Ce corrigé contient **6** questions sur un total de **14** pages **(excluant cette page)**

La pondération de cet examen est de **40** **%**

Vous devez répondre sur : ☒ le questionnaire ☐ le cahier ☐ les deux

Vous devez remettre le questionnaire : ☒ oui ☐ non

L'étudiant doit honorer l'engagement pris lors de la signature du code de conduite.

Question 1 : Monceaux (18 points)

Pour cette question, référez-vous au code Java donné à l'Annexe 1.

1.1) **(1 pt)** Donnez la complexité asymptotique en pire cas d'un `findMin()`.

O(1)

1.2) **(1 pt)** Donnez la complexité asymptotique en pire cas d'un `deleteMin()`.

O(lg(*n*))

1.3) **(1 pt)** Donnez la complexité asymptotique en meilleur cas d'un `buildHeap()`.

O(*n*)

1.4) **(1 pt)** Donnez la complexité asymptotique en pire cas d'un `buildHeap()`.

O(*n*)

1.5) **(1 pt)** Donnez la complexité asymptotique en cas moyen d'un `insert(AnyType x)`.

O(1)

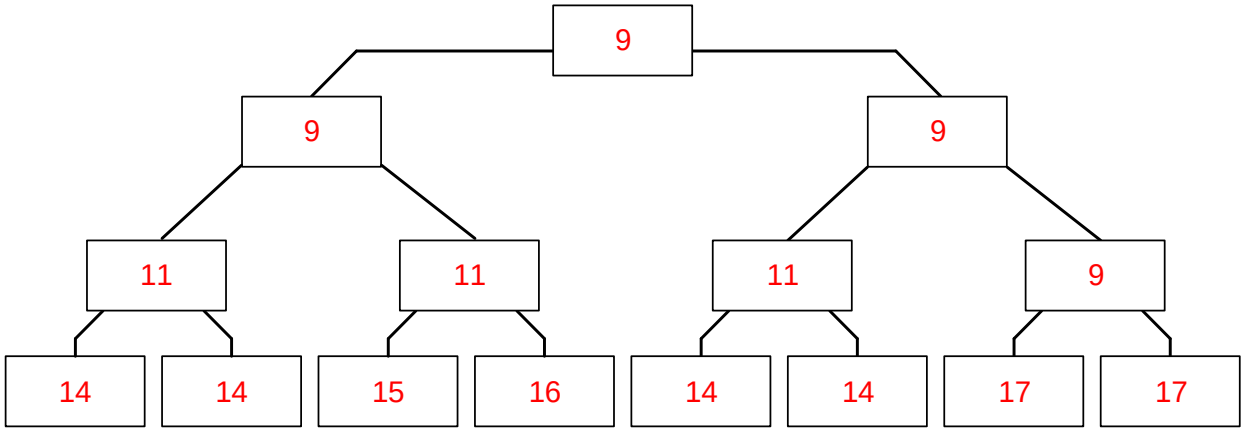
1.6) **(1 pt)** Donnez la complexité asymptotique en meilleur cas d'un `insert(AnyType x)`.

O(1)

1.7) **(1.5 pts)** Dessinez le monceau contenu en mémoire dans le tableau ci-après.

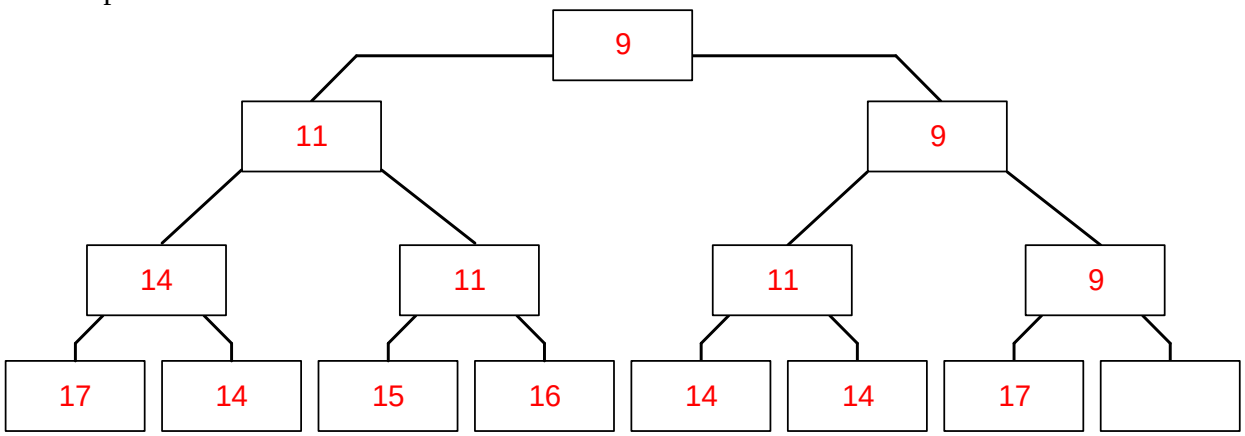
Indice	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Contenu	-	9	9	9	11	11	11	9	14	14	15	16	14	14	17	17

Votre réponse :



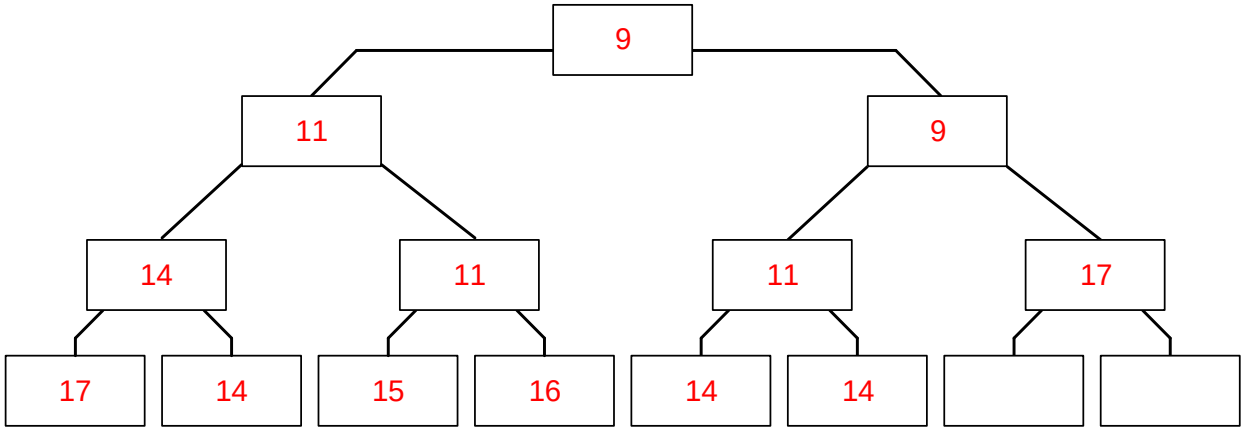
1.8) **(1.5 pts)** En partant du monceau de 1.7), effectuez un `deleteMin()`.

Votre réponse :



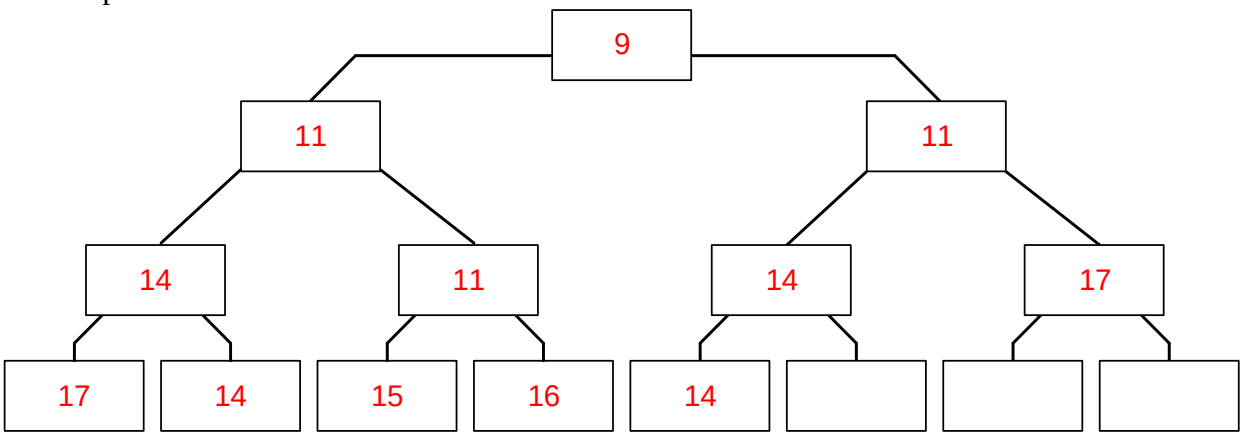
1.9) **(1.5 pts)** En partant du monceau de 1.8), effectuez un `deleteMin()`.

Votre réponse :



1.10) (1.5 pts) En partant du monceau de 1.9), effectuez un `deleteMin()`.

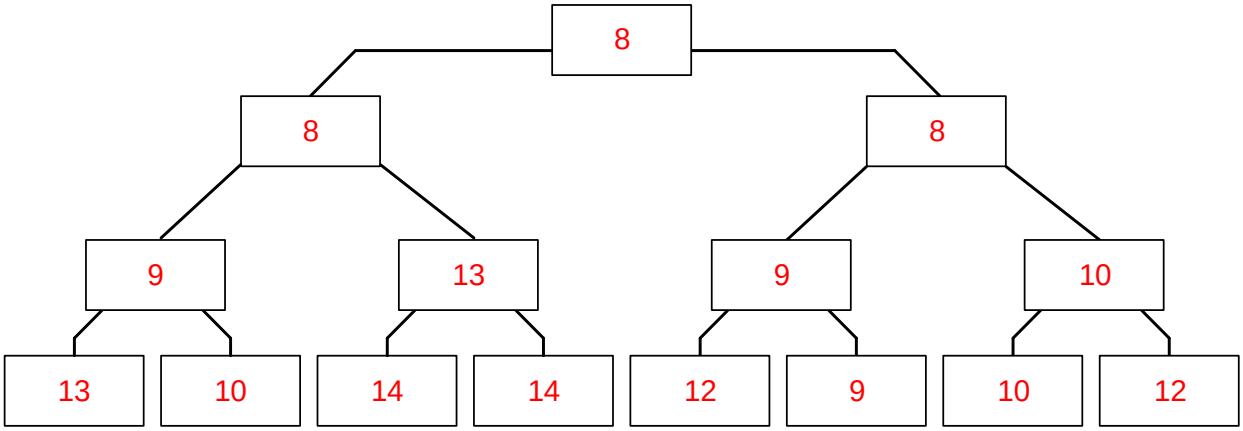
Votre réponse :



1.11) (2 pts) Dessinez le monceau résultant d'appel à `BinaryHeap(AnyType[ ] items)` où `items` est le tableau suivant :

Indice	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Contenu	13	14	12	13	8	9	10	9	10	8	14	8	9	10	12	

Votre réponse :



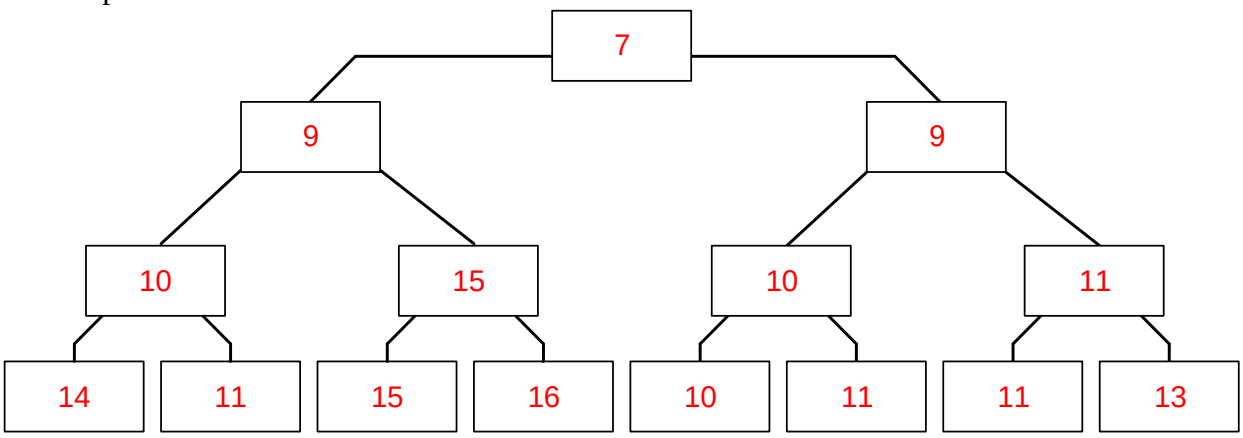
1.12) (4 pts) Dessinez le monceau résultant d'appel à `BinaryHeap(AnyType[ ] items)` où `items` est le tableau suivant :

Indice	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Contenu	16	15	11	14	9	10	11	10	11	7	15	9	10	11	13	

Et où la méthode `percolateDown` a été modifiée comme suit :

```
private void percolateDown( int hole ) {
    int child;
    AnyType tmp = array[ hole ];

    for( ; hole*2 <= currentSize; hole = child ) {

        child = hole*2;

        if( child != currentSize &&
            array[child+1].compareTo( array[child] ) =< 0 )
            child++;

        if( array[ child ].compareTo( tmp ) < 0 )
            array[ hole ] = array[ child ];
        else
            break;
    }
    array[ hole ] = tmp;
}
```

Votre réponse :



Question 2 : Recherche de patron (14 points)

L'algorithme Rabin Karp est un algorithme permettant de retrouver une chaîne de caractères dans un texte. La chaîne de caractères recherchée est alors remplacée par un entier via un calcul de hachage. Une implémentation de Rabin Karp vous est proposée à l'Annexe 2. On y considère l'alphabet $\Sigma = \{A, C, G, T\}$ où $d = |\Sigma| = 4$. On admettra l'encodage suivant :

Symbole	A	C	G	T
Code	0	1	2	3

Le hachage est calculé en base d modulo $q = 2^8+1 = 257$. Par exemple, la séquence « GATTACA », encodée **2, 0, 3, 3, 0, 1, 0** produira une valeur de hash :

$$\text{mod}((((((\mathbf{2}\cdot d + \mathbf{0})\cdot d + \mathbf{3})\cdot d + \mathbf{3})\cdot d + \mathbf{0})\cdot d + \mathbf{1})\cdot d + \mathbf{0}, q) = 161.$$

Sachant que pour deux entiers a et b pris dans l'intervalle $[0, 2^8)$,
$\text{mod}(2^8 \cdot a+b, 257) = 257+ b - a$ si $a \geq b$, et $b - a$ si $a < b$.

Sachant que pour deux entiers a et b pris dans l'intervalle $[0, 2^{16})$,
$\text{mod}(2^{16} \cdot a+b, 257) = \text{mod}(a + b, 257)$

2.1) (4 pts) Donnez le hash associé au patron « ATGATTACA ». Justifiez votre réponse par un calcul.

ATGATTACA peut être décomposé en trois parties A TGAT TACA
Le hash associé à A donne 0
Le hash associé à TGAC donne $\text{mod}(((3\times4+2)\times4+0)\times4+3, 257) = 227$
Le hash associé à TACA donne $\text{mod}(((3\times4+0)\times4+1)\times4+0, 257) = 196$
$p = \text{mod}(0\times2^{16} + 227\times2^8 + 196, 257) = 257 + 196 - 227 = 226$

Afin d'accélérer les calculs lors de la recherche du patron dans le texte, Rabin Karp exploite une formule récursive : $t_{s+1} = \text{mod}((t_s - hT[s])d + T[s+m+1], q)$, où $h = \text{mod}(d^{m-1}, q)$.

2.2) (2 pts) Donnez la valeur de la variable h pour le patron « ATGATTACA » :

Votre réponse :

$$h = \text{mod}(4^{9-1}, 257)$$
$$= \text{mod}(4^8, 257)$$
$$= \text{mod}(2^{16}, 257)$$
$$= 1$$

2.3) (2 pts) Simplifiez la formule $t_{s+1} = \text{mod}((t_s - hT[s])d + T[s+m+1], q)$ pour le cas où le patron recherché est « ATGATTACA » en utilisant le résultat de 2.2):

Votre réponse :

$$t_{s+1} = \text{mod}(4(t_s - T[s]) + T[s+m+1], 257)$$

2.4) (6 pts) Combien de faux positifs seront détectés si le patron « GAAAAAAAC » est recherché dans le texte « TAAAAAAAATGAAAAAAAC ». Fiez-vous au code java donné à l'Annexe 2. Remarquez que dans l'implémentation donnée de la fonction `updateHash`, le calcul du modulo renvoie toujours une valeur positive. Aidez-vous de la table de la page suivante.

```
private int updateHash(int oldHash, char oldChar, char newChar) {
   if( !ENCODING.containsKey( oldChar ) ) throw new IllegalArgumentException();
   if( !ENCODING.containsKey( newChar ) ) throw new IllegalArgumentException();

   int newHash = oldHash;
   newHash -= h*ENCODING.get( oldChar );
   if( newHash < 0 ) newHash += q; // valeur positive
   newHash *= d;
   newHash += ENCODING.get( newChar );
   newHash %= q;
   return newHash;
}
```

Décalage	Sous-séquence	t_s	Faux Positif?
0	TAAAAAAAA	3	√
1	AAAAAAAAT	3	√
2	AAAAAAATG	14	
3	AAAAAATGA	56	
4	AAAAATGAA	224	
5	AAAATGAAA	125	
6	AAATGAAAA	243	
7	AATGAAAAA	201	
8	ATGAAAAAA	33	
9	TGAAAAAAA	132	
12	GAAAAAAAC	3	

2.4.a) **(3 pts)** Nombre de faux positifs détectés :

2

2.4.b) **(3 pts)** Les sous-séquences associées aux faux positifs détectés.

TAAAAAAAA et AAAAAAAAT

Question 3 : Programmation dynamique (16 points)

On désire trouver le parenthésage idéal pour multiplier les matrices A_1 à A_5 permettant de minimiser le nombre de multiplications (scalaires) à effectuer. Les matrices sont dimensionnées comme suit :

$A_1 : 3 \times 1$; $A_2 : 1 \times 4$; $A_3 : 4 \times 1$; $A_4 : 1 \times 1$; $A_5 : 1 \times 3$

Considérez les tables **m** et **s** obtenues par l'exécution de l'algorithme dynamique vu en cours.

m	1	2	3	4	5
1	0	12	7	8	17
2		0	4	5	8
3			0	4	15
4				0	3
5					0

s	1	2	3	4	5
1		1	1	1	1 ou 4
2			2	3	4
3				3	3
4					4
5					

Complétez cette table pour répondre aux questions suivantes :

Rappel : $m[i, j] = \min\{m[i, k] + m[k+1, j] + p_{i-1}p_k p_j\}$ pour $k = i$ à $j-1$, sachant que la matrice A_i a une dimension $p_{i-1} \times p_i$.

3.1) (4 pts) Donnez le parenthésage optimal pour multiplier A_1 à A_4. Donnez son coût.

Parenthésage optimal: $A_1 ((A_2 A_3) A_4)$

Coût: 8

3.2) (4 pts) Donnez le parenthésage optimal pour multiplier A_2 à A_5. Donnez son coût.

Parenthésage optimal: $((A_2 A_3) A_4)A_5$

Coût: 8

3.3) (1 pts) Donnez le coût minimal de la multiplication matricielle suivante. Justifiez par un calcul.

$A_1(A_2A_3A_4A_5)$,

Coût : $3\times1\times3 + 8 = 17$

3.4) (1 pts) Donnez le coût minimal de la multiplication matricielle suivante. Justifiez par un calcul.

$(A_1A_2)(A_3A_4A_5)$,

Coût : $3\times4\times3+12+15 = 63$

3.5) (1 pts) Donnez le coût minimal de la multiplication matricielle suivante. Justifiez par un calcul.

$(A_1A_2A_3)(A_4A_5)$,

Coût : $3\times1\times3+7+3 = 19$

3.6) (1 pts) Donnez le coût minimal de la multiplication matricielle suivante. Justifiez par un calcul.

$(A_1A_2A_3A_4)A_5$,

Coût : $3\times1\times3+8 = 17$

3.7) (4 pts) Donnez le parenthésage optimal pour multiplier A_1 à A_5. Donnez son coût.

Parenthésage optimal: $A_1\ (((A_2\ A_3)\ A_4)\ A_5)$ ou $(A_1\ ((A_2\ A_3)\ A_4))\ A_5$

Coût: 17

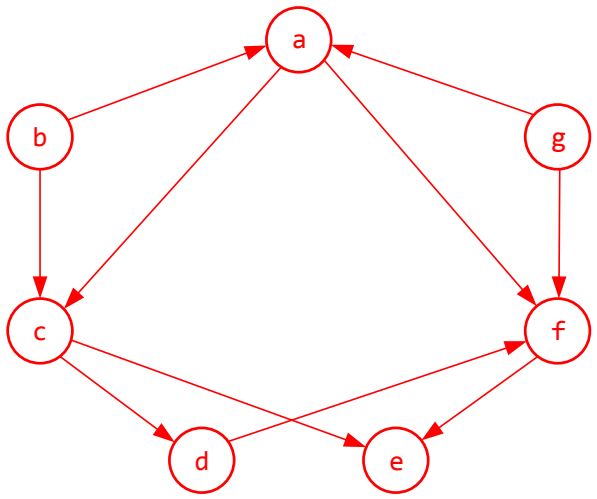
Question 4 : Algorithmes sur les graphes (24 points)

On veut connaître l'ordre topologique du graphe dirigé suivant :

$$V = \{a, b, c, d, e, f, g\}$$

$$E = \{(a, c), (a, f), (b, a), (b, c), (c, d), (c, e), (d, f), (f, e), (g, a), (g, f)\}$$

4.1) (2 pts) Reproduisez graphiquement le graphe G = (V, E) :



4.2) (7 pts) Donnez l'ordre topologique du graphe G en appliquant l'algorithme utilisant une file vu en classe.

Nœud	1	2	3	4	5	6	7
a	2	1	0	-	-	-	-
b	0	-	-	-	-	-	-
c	2	1	1	0	-	-	-
d	1	1	1	1	0	-	-
e	2	2	2	2	1	1	0
f	3	3	2	1	1	0	-
g	0	-	-	-	-	-	-
Entrée	b, g	-	a	c	d	f	e
Sortie	b	g	a	c	d	f	e

Ordre trouvé (débutez la numérotation à 1) :

Nœud	a	b	c	d	e	f	g
Ordre :	3	1	4	5	7	6	2

4.3) (7 pts) Donnez l'ordre topologique du graphe G en appliquant l'algorithme utilisant le parcours DFS post-ordre inverse. Partez du nœud `a` et visitez les nœuds alphabétiquement.

4.3.a) (2 pts) Donnez l'affichage DFS post-ordre obtenu :

e, f, d, c, a, b, g

4.3.b) (2 pts) Donnez l'affichage DFS post-ordre inverse obtenu :

g, b, a, c, d, f, e

4.3.c) (3 pts) Donnez l'ordre topologique trouvé (débutez la numérotation à 1) :

Nœud	a	b	c	d	e	f	g
Ordre :	3	2	4	5	7	6	1

4.4) (2 pts) Combien de composantes fortement connexes G contient-il? Donnez-les.

Il en existe 7, soit chacun des sommets individuellement.

4.5 (3 pts) Combien de composantes fortement connexes G contiendrait-il si on y ajoutait l'arc `(e, g)`? Donnez-les

Il n'y en aurait plus que 2, b d'un côté et l'ensemble des sommets restants ensemble.

4.6) (3 pts) Proposez l'ajout de deux arcs dirigés à G pour que le graphe devienne connexe.

Réponse :

Arcs à ajouter :

Arc 1 : (e , g)

Arc 2 : (e , b)

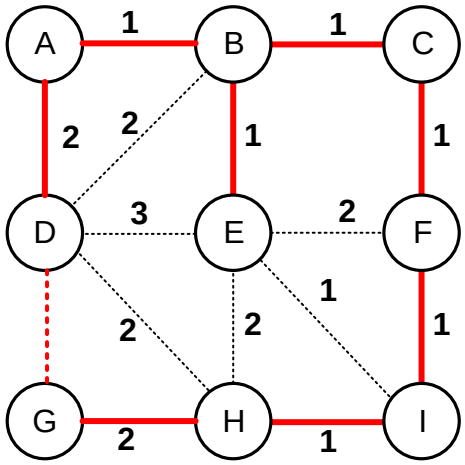
D'autres solutions existent.

Question 5 : Arbre sous-tendant minimum (12 points)

Donnez les arbres sous-tendant minimum obtenus par les algorithmes de Prim (le nœud de départ étant `A`) et Kruskal en reliant les arêtes retenues dans les graphes données ci-après.

Respectez l'ordre alphabétique pour visiter les nœuds voisins ou les arêtes. Utilisez les tables fournies pour ce faire (le remplissage des tables ne compte pas dans l'attribution des points pour la question).

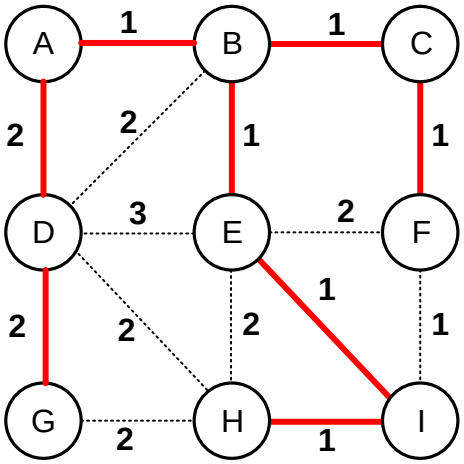
Par Prim (6 pts) :



Prim:

Nœud	Distance	Parent	Connu?
A	0	-	√
B	1	A	√
C	1	B	√
D	2	A	√
E	1	B	√
F	1	C	√
G	2	H	√
H	1	I	√
I	1	F	√

Par Kruskal (6 pts) :



Kruskal :

Arête	Poids	Retenue?
(A, B)	1	√
(A, D)	2	√
(B, C)	1	√
(B, D)	2	
(B, E)	1	√
(C, F)	1	√
(D, E)	3	
(D, G)	2	√
(D, H)	2	
(E, F)	2	
(E, H)	2	
(E, I)	1	√
(F, I)	1	
(G, H)	2	
(H, I)	1	√

Question 6 : Généralités (16 points)

Répondez aux assertions suivantes par « vrai » ou par « faux ». Justifier votre réponse brièvement. Les réponses non justifiées ne seront pas considérées.

6.1) **(2 points)** Le tri monceau est un tri interne.

Vrai, puisqu'il n'utilise pas de mémoire supplémentaire.

6.2) **(2 points)** Soit G = (V, E) un graphe dirigé. G^T (le graphe transposé de G) possède au moins deux composantes fortement connexes si et seulement si G est connexe.

Faux. Si G est connexe, alors GT aussi, et il ne possède qu'une seule CFC.

6.3) **(2 points)** Les algorithmes issus des méthodes de programmation dynamique requièrent un espace mémoire proportionnel à $O(n^2)$ (les besoins en mémoire croissent quadratiquement avec la taille du problème).

Faux. Cela varie d'un problème à l'autre. Exemple Fibonnaci peut-être résolu avec un espace mémoire (n), et même O(1).

6.4) **(2 points)** Un algorithme glouton est un algorithme dont les besoins en mémoire sont proportionnels à $O(n^2)$ (croissent quadratiquement avec la taille du problème).

Faux. O(n) pour Kruskal par exemple.

6.5) **(2 points)** La solution vue en classe au problème de parenthésage des matrices est un algorithme glouton.

Faux, algorithme issu des méthodes de programmation dynamique.

6.6) **(2 points)** Partant d'un sommet quelconque d'un graphe dirigé, le parcours DFS permet de statuer sur la connexité du graphe (dire si le graphe est fortement connexe ou pas).

Vrai. Si le graphe est fortement connexe, DFS visitera tous les nœuds en partant d'un sommet quelconque. Si ce n'est pas le cas lorsque le graphe n'est pas connexe.

6.7) **(2 points)** Si la longueur m d'un patron P[1 :m] est le quart de la longueur n du texte T[1 :n] dans lequel il est recherché (m=n/4), alors pour des valeurs de n suffisamment grandes, l'utilisation d'un automate est plus rapide que l'algorithme naïf, même en incluant le temps de construction de l'automate.

Faux. La construction du monceau se fait en $O(d^m)$. Si m est proportionnel à n, l'algorithme a une complexité exponentielle.

6.8) **(2 points)** La structure de données d'ensembles disjoints vue en classe permet d'identifier l'ensemble auquel appartient un élément en O(lg(n)) dans le pire cas.

Faux. O(n) en pire cas.