

Institut Supérieur d'Informatique de Médenine

Midterm Knowledge representation

MPLIC1

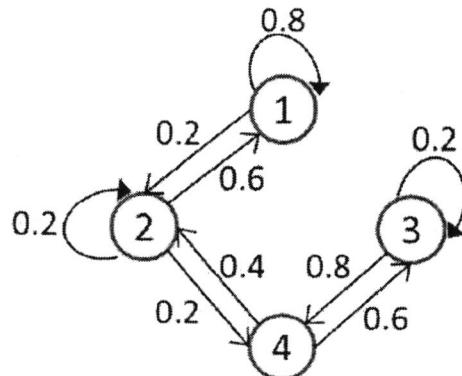
Duration: 1h

March 2023

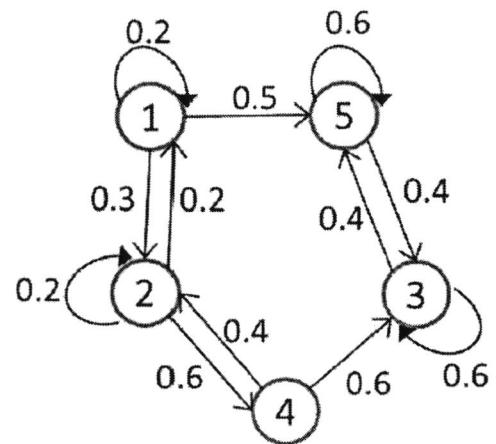
Teacher: F. Jarray

Exercise

Consider the following Markov chains



(a)



(b)

For each of them answer the following questions:

1. Determine the transition matrix
2. How many classes does it have. What is the type of each class?
3. Is the chain irreducible?
4. Determine the stationary distribution
5. Suppose that today, the chain is 40% at state 1 and 60% at state 2, what is its distribution for tomorrow?

Institut Supérieur d'Informatique de Médenine

Midterm Recherche d'information et indexation

GLSI2, ILC1

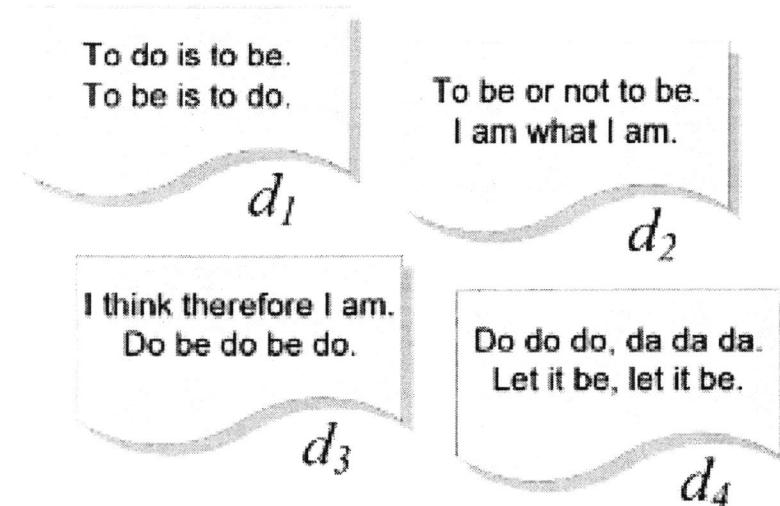
Duration : 1h

March 2023

F.Jarray

Exercise 1

Considering the following corpus of 4 documents



Assume that the following terms are stop words: "not", "or", ",", ":"

Question: Construct the inverted index after capitalization ignoring (transform to lowercase).

Exercise 2

Consider the following occurrences of terms in a corpus of 5 documents Doc1, ..., Doc5.

Term\doc	Doc1	Doc2	Doc3	Doc4	Doc5
Chatbot	3	2	0	5	31
GPT2	7	7	0	15	0
GPT3	0	1	8	1	1

Question1 : Compute the idf for each term

Question2: Compute the tf-idf weight for each term in each document.

Question3: Find the most relevant document to the query Q="Chatbot Chatbot GPT3"

Institut Supérieur d'Informatique de Médenine

Midterm Ingenierie linguistique

MPLIC1

Duration : 1h

March: 2023

F. Jarray

Exercise

Consider a corpus comprising the following sentences:

- S1: This is an NLP application
- S2: This is my NLP
- S3: I design my application
- S4: It is my application

Without removing any token, answer the following questions.

Question 1:

Determine the bigram probabilities table with smoothing

Question2:

Which of the most fluent following sentences, i.e., has a higher probability?

- S5: This is my application
- S6: This is my NLP

Question3:

Predict the most probable next word for the following word sequences?

- S7: This application is
- S8: This NLP is ...

Question 4

Compute the perplexity of the model on the test set=(S9,S10)

S9: This is my application my application.

S10: my application is NLP.

INSTITUT SUPÉRIEUR DE L'INFORMATIQUE DE MÉDENINE
AU 2022/2023 SEMESTRE II

DS

<i>Classe : MP1 ILC</i>	<i>Matière : Programmation web avancée</i>
<i>Documents Non Autorisés</i>	<i>Durée : 1 heure</i>

Nb pages : 1

On se propose d'écrire un code XML pour écrire un curriculum vitae (CV). Un CV comporte:

- Les coordonnées : nom, prénom, adresse, date de naissance, téléphone et e-mail.
 - o Une adresse est composée d'un code postal et d'une ville.
- Les formations : année de formation, diplôme, lieu de formation et éventuellement la mention.
- L'activité qui apparaît au plus une fois.
- Les langues sont optionnelles. Si elles existent, alors une langue comporte les attributs :
 - o nom_langue qui est obligatoire,
 - o langue_écrite qui peut avoir les valeurs *oui* ou *non*.
 - o et langue_parlée qui a la valeur par défaut *oui*.

1- Ecrire le code XML pour l'exemple de CV suivant :

2- Ecrire la DTD correspondante.

nom	prénom	Adresse		date naissance	téléphone	email
		code postal	ville			
Ben Salah	Salah	1111	Tunis	03/04/1996	98333333	salah@yahoo.fr

Formations :

année	diplôme	Lieu	mention
2022	Licence	Medenine	Assez bien
2023	Mastere	Medenine	Très bien

Activité : Chef de service

Langues :

nom_langue	Langue_parlée	Langue_écrite
arabe	oui	Oui
italien	oui	Non
français	oui	Oui
anglais	oui	Oui

Bon Travail

Institut Supérieur d'Informatique de Médenine

Midterm Ingenierie linguistique

MPILC1

Duration : 1h

March: 2023

F. Jarray

Exercise

Consider a corpus comprising the following sentences:

- S1: This is an NLP application
- S2: This is my NLP
- S3: I design my application
- S4: It is my application

Without removing any token, answer the following questions.

Question 1:

Determine the bigram probabilities table with smoothing

Question2:

Which of the most fluent following sentences, i.e., has a higher probability?

- S5: This is my application
- S6: This is my NLP

Question3:

Predict the most probable next word for the following word sequences?

- S7: This application is
- S8: This NLP is ...

Question 4

Compute the perplexity of the model on the test set=(S9,S10)

S9: This is my application my application.

S10: my application is NLP.

	<h1 style="margin: 0;">Devoir surveillé</h1>	Année universitaire : 2022 /2023 Matière: Vision d'un ordinateur Classe : MP1-ILC Durée : 1h Enseignantes : Dr. N. Jaouedi
-----------------------------------------------------------------------------------	----------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------

Exercice 1 :

La vision par ordinateur permet aux ordinateurs de dériver des informations significatives à partir des entrées visuelles.

Développer cette phrase et expliquer son importance dans quelques domaines. (5 pts)

Exercice 2 :

La segmentation consiste à partitionner l'image en des zones.

Développer cette phrase, expliquer les types des approches et décrire le principe d'une méthode que vous choisis. (5 pts)

Exercice 3 :

Soient la représentation numérique d'une image I1 et le filtre F1 les suivants. (5pts)

100	100	100	100	100
100	100	100	100	100
100	100	150	100	100
100	100	100	100	100
100	100	100	100	100

I1

0	-1	0
-1	5	-1
0	-1	0

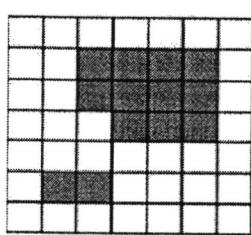
F1

Appliquer le filtre F1 sur le pixel central I1(3,3) de la valeur 150 et conclure le type de F1.

Appliquer un autre filtre F2 à vos choix pour rendre l'image I1 homogène.

Exercice 4 :

Soient la représentation matricielle R1 d'une image et un élément structurant E1 les suivants. (5pts)



R1

0	1	0
1	1	1
0	1	0

E1

Appliquer l'opération de l'ouverture sur cette représentation et expliquer les résultats trouvés

Université de Gabes
Institut Supérieur de l'Informatique de Médenine
Framework Symfony

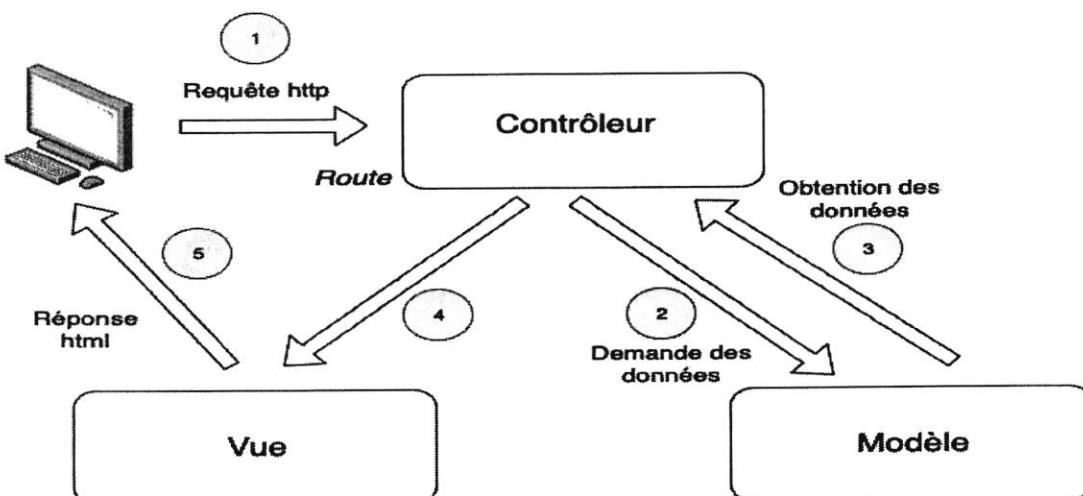
Enseignant : Dr. Yassine Adouani

Niveau : MP1- ILC

Documents non autorisés

Exercice 01 (07 pts)

1- Soit l'architecture : **Modèle, Vue et Contrôleur (MVC)** suivante:



a- Présenter les composants du modèle MVC :

- Contrôleur
- Vue
- Modèle

b- En se basant sur le schéma ci-dessus du modèle MVC, dans quel cas on aura besoin seulement de : **1- 4-5** ?

2- Présenter le Framework symfony.

3- Le Framework Symfony est basé sur quel langage ? (donner un seul langage)

4- Quelle est la version actuelle de Symfony ?

5- Quelle est la dernière version de PHP ?

6- Quels sont les **avantages** d'utiliser symfony ?

7- Quel est le rôle de « **composer** » dans un Framework symfony ?

8- Donner la commande permettant de créer un projet d'application web

9- Donner la commande permettant de démarrer un serveur symfony

10- C'est quoi le **Git-hub** ?

Exercice 02 (13 pts)

Soit la partie Front-End d'un site web intitulé « Amphi virtuelle » est constituée par les composants suivants :

- **Vue** : Page inscription.html.twig
- **Contrôleur** : Ficher AmphiControlleur.php
- Fichier style.css
- Fichier verif.js

Une partie de code source de la page **inscription.html.twig** (Vue) est comme suit :

```
<tr>
    <th>&nbsp;<label for="M">Matricule</label></th>
    <td><input name="M" type="text" required="required" id="M" size="10"> </td>
    <td>&nbsp;</td>
</tr><tr>
    <th scope="row"><label for="NP">Nom</label></th>
    <td><input name="NP" type="text" required="required" id="NP"></td>
    <td>&nbsp;</td>
</tr><tr>
    <th scope="row"><label for="A">Age</label></th>
    <td><input name="A" type="number" required="required" id="A"></td>
    <td>&nbsp;</td>
</tr><tr>
    <th scope="row"><label for="H">Genre</label></th>
    <td><input type="radio" name="G" id="H" value="H">
        <label for="H">Homme&nbsp;</label></td>
    <td><input type="radio" name="G" id="F" value="F">
        <label for="F">Femme</label></td>
</tr><tr>
    <th scope="row"><label for="E">Email</label></th>
    <td><input name="E" type="email" required="required" id="E"></td>
```

- 1- Ajouter le code nécessaire permettant de créer un lien entre le fichier **inscription.html.twig** et les deux fichiers « **style.css** » et « **verif.js** »
- 2- Développer la partie contrôleur **AmphiControlleur.php**
- 3- Donner le contenu de fichier **routes.yaml**
- 4- Créer la fonction **verif()** en javascript permettant de vérifier les champs suivants :
 - Le champ « **Matricule** » doit être numérique de longueur 8
 - Le champ « **Nom** » doit être alphabétique.
 - Le champ « **Age** » doit être de valeur entre 20 et 6?.
 - Le champs « **Email** » doit contenir obligatoirement les caractères « @ » et « . ».

Bon Travail



Devoir Surveillé



Semestre : 1 2

Session : Principale

Unité d'enseignement : Analyse des données massives (Big Data)

Enseignante : Dr. Sawsan Selmi

Documents autorisés : OUI NON

Nombre de pages : 2 pages Classe : MP1- ILC

Internet autorisée : OUI NON

Durée : 60 minutes

Exercice 1 : Questions de compréhension [10 points]

- 1- Quels sont les 5 Vs du big data. Expliquez les brièvement en donnant de exemples de domaines d'utilisation.
- 2- Quels sont les modules principaux existants dans la version 1.0 de Hadoop. (càd les composants de l'écosystème Hadoop). Rappeler leurs principes de fonctionnement.
- 3- Quels sont les différences entre Hadoop V 1.0 et v 2.0 ?
- 4- Comment le NameNode gère-t-il les défaillances de DataNode ?
- 5- Citer les principales différences entre la base de données relationnelle et HDFS ?

Exercice 2 : Calculs de moyenne pondérée en MapReduce [10 points]

On considère des gros volumes de données issues de mesures regroupés dans des Records, stockés dans des fichiers comme le «FichierMesures1.txt » ci-dessous :

On peut ainsi analyser de multiples mesures concernant l'ensemble des moteurs. Afin d'obtenir un système fiable, plusieurs capteurs sont installés sur chaque moteur, qui mesurent la même grandeur, mais qui n'ont pas tous la même fiabilité.

On souhaite coder un programme permettant de calculer la valeur moyenne de toute une série de mesures concernant un même moteur. Chaque mesure participera à la moyenne en étant pondérée par le coefficient de fiabilité de son capteur.

FichierMesures1.txt

```
{ moteurID : RefDuMoteur,  
capteurID : RefDuCapteur,  
rafale : [v0,..., v9],  
unité : UnitéDeLaMesure,  
fiabilitéCapteur : PoidsDeLaMesure }
```

- 1- Décrire le principe de stockage de ses fichiers de données à l'aide de HDFS en se basant sur un schéma explicatif.

2- Expliquer le traitement à faire, en se basant sur un système simple constituée seulement de Mappers et de Reducers.

3- Faut il créer un ou plusieurs Mappers et Reducers?

4- Définissez les paires clé valeurs de votre solution.

5- Il arrive qu'un capteur soit perturbé et produise une valeur aberrante.

Pour cela, nous considérerons qu'une valeur de mesure inférieure à 100 ou supérieure à +100 est une valeur aberrante.

Proposez un schéma expliquant les valeurs issues de votre solution en entrée et sortie des Mappers et Reducers.

6- Indiquez les eventuels changements de votre système appliqué à cet exercice si on utilise Hadoop V2.0.

😊 Bon courage 😊

Institut Supérieur de l'informatique de Médenine

Semestre : 2

Devoir Surveillé N°1

Matière : Techniques de communication

Niveau : 1^{re} année mastère ILC

Durée : 1 heure

Enseignante : Mme Emna Smida

Partie I

- 1) Quelle est la différence entre la communication interne et externe de l'entreprise ? et où réside leur importance ?
- 2) Le poster scientifique : un moyen efficace pour mettre en valeur un projet.

NB. La réponse doit être entièrement rédigée sous forme d'un paragraphe.

Partie II

Élaborer un résumé pour cet extrait

Traitement automatique du langage naturel (TALN)

Le traitement automatique du langage, récemment à la croisée de la linguistique, de l'informatique et de l'intelligence artificielle, voit ses applications, ses programmes et beaucoup de techniques informatiques, au service du langage humain en vue d'appréhender le sens des données en langage naturel. Une compréhension de haut niveau pour ce raisonnement humain a été longtemps recherchée et considérée comme le but extrême des premiers travaux.

On regroupe sous le vocable de traitement automatique du langage naturel (TALN) l'ensemble des recherches et développements visant à modéliser et reproduire, à l'aide de machines, la capacité humaine à produire et à comprendre des énoncés linguistiques dans des buts de communication. Il sera donc question ici de langage humain, d'où l'adjectif naturel, et non pas de langage formel, tel que C ou encore ADA. Ce naturel fait d'ailleurs tout le sel de

l'affaire : les langages formels sont précisément conçus et optimisés dans l'optique de manipulations algorithmiques. Il en va tout autrement pour le langage naturel, dont le traitement automatique pose des difficultés majeures.

L'étude du langage naturel et des mécanismes nécessaires à la mise en œuvre à son traitement automatique par des machines est un domaine d'études foisonnant, et riche en applications potentielles ou émergentes. De nombreux progrès restent à accomplir pour mieux comprendre cette faculté et pour bâtir des systèmes capables de soutenir la comparaison avec l'humain, mais l'état des connaissances en permet aujourd'hui de proposer de nombreuses solutions efficaces à des problèmes et des demandes réels. Une des limitations de pratiquement tous les systèmes de traitement un peu sophistiqués est qu'ils font appel à une somme importante de connaissance d'expert : lexiques, règles de grammaires, réseaux sémantiques... Ceci explique en partie pourquoi il n'existe pas de système de traitement qui soit à la fois complet (i.e. intégrant tous les niveaux de traitement) et indépendant du domaine (i.e. capable de traiter avec une même efficacité n'importe quel type de texte). Il existe une autre raison, moins visible, qui limite l'avancée des progrès en TALN, et qui est que, pour un bon nombre de phénomènes, l'état de la connaissance linguistique est insuffisamment formalisée pour pouvoir être utilisée par les concepteurs de systèmes de TALN. En réaction, de nombreux développements récents font appel à l'utilisation massive de corpus ou de lexiques (mono ou multilingues) pour extraire directement ces connaissances à l'aide de techniques statistiques. Ainsi, des analyseurs morphologiques, des étiqueteurs, des grammaires, des systèmes de traduction basés sur l'exploitation de corpus et bien d'autres outils encore ont été développés dans les années passées, avec des résultats plus qu'encourageants. Les avantages de ces outils sont (en théorie) multiples :

- Ils sont indépendants (au moins partiellement) de la langue, c'est-à-dire qu'ils peuvent être utilisés pour extraire des connaissances relatives à des langues différentes (suffisamment voisines toutefois) ;
- Ils sont facilement portables d'une application à l'autre, et l'adaptation à un nouveau domaine est facilitée.

François Yvon

BON COURAGE !

***Examen de CC
Entreprenariat***

Questions de cours :

- 1- Pourquoi une entreprise est considérée comme une unité de production ?
- 2- La finalité financière est-elle la seule finalité d'une entreprise ?
- 3- Quelle est la différence entre micro et macro-environnement ?
- 4- Selon vous, quels peuvent être les facteurs contextuels ou personnels favorisant la création d'entreprise ?

Exercice :

DOCUMENT : « L'avenir ne se prévoit pas, il se prépare »

Entreprendre, c'est faire un pari sur l'avenir. On peut reprendre la vision de Schumpeter qui part du principe que l'entrepreneur incarne le pari de l'innovation. Entreprendre c'est :

- Produire de nouveaux produits ;
- Développer de nouvelles méthodes de production ;
- Introduire de nouvelles formes d'organisation ;
- Conquérir de nouveaux marchés.

- 1- Qu'est-ce qu'une démarche entrepreneuriale ?
- 2- Expliquez le titre du document : « **L'avenir ne se prévoit pas, il se prépare** ».
- 3- Quelles sont les conditions nécessaires à la réussite d'un projet d'entreprise ?
- 4- Quels sont les enjeux de la logique entrepreneuriale ?

Bon Travail



Enseignante : Boukouba Rihab

DS du module « Datawarehouse»

MP1ILC

Nom :

Prénom :

Groupe :

Exercice 1 : Répondre par vrai ou faux. Toute réponse doit être justifiée. **(8 points)**

1. Les entrepôts de données sont destinés à stocker les données brutes de l'entreprise.
Vrai ou Faux?

2. L'ETL signifie Extraction, Transformation et Lecture de données. Vrai ou Faux?

3. Un datamart stocke toutes les données de l'entreprise pour une utilisation ultérieure.
Vrai ou Faux?

4. L'OLTP est un système qui traite les transactions en temps réel. Vrai ou Faux?

5. La transformation des données est la première étape du processus ETL. Vrai ou Faux?

6. Un entrepôt de données facilite l'analyse à grande échelle des données de l'entreprise. Vrai ou Faux?

7. L'objectif principal d'un datamart est de fournir des informations spécifiques à un domaine métier particulier. Vrai ou Faux?

8. Le chargement des données est la dernière étape du processus ETL. Vrai ou Faux?

Problème : (12 points)

1. Modéliser l'exemple de production et de fabrication en un **seul table de fait** et de **dimensions**.

Supposons que nous avons une entreprise de fabrication qui produit des pièces pour l'industrie automobile. Nous pouvons avoir **une table de fait** qui contient les données sur la production quotidienne de pièces. Cette table de fait peut inclure des informations telles que le nombre de pièces produites chaque jour, le temps de production, les matières premières utilisées et les coûts de production.

Les tables de dimensions peuvent inclure des informations sur les matières premières utilisées, les machines utilisées pour la production, les employés qui travaillent sur la production, les coûts des matières premières, etc.

Par exemple, la table de dimension "matières premières" peut contenir des informations sur la quantité de matières premières utilisées pour la production de chaque pièce. La table de dimension "machines" peut contenir des informations sur la capacité de production de chaque machine et le temps de fonctionnement de chaque machine. La table de dimension "employés" peut contenir des informations sur les employés travaillant sur la production, leur rôle et leur performance.

On vise d'avoir une **vue complète** de la production quotidienne de pièces, y compris les matières premières utilisées, les machines utilisées et les employés impliqués. Cette approche peut aider l'entreprise à optimiser la production en identifiant les goulots d'étranglement et en identifiant les domaines où des améliorations peuvent être apportées pour améliorer l'efficacité de la production.

2. Quel est le but de cette modélisation ?



Enseignante :Mme Saadaoui Zakia

Durée : 1H

La clarté et la propreté de la copie sont indispensables.
Le barème est indicatif.

NOM ET PRENOM :.....

Exercice N°1

Cocher la(les) bonne(s) réponse(s)

1. Quelle est l'unité d'échange entre les fichiers, la mémoire et les processus ?
A. DB_BLOCK_BUFFERS B. DB_KEEP_CACHE_SIZE
C. DB_RECYCLE_CACHE_SIZE D. DB_BLOCK_SIZE

2. Quelles sont les tailles possibles pour le bloc de données ?
A. 1K B. 2K C. 4K D. 6K E. 8K
F. 10K G. 14K H. 16K I. 32K

3. Quel est le mode de démarrage de l'instance pour pouvoir créer une base de données ?
A. STARTUP B. STARTUP NOMOUNT
C. STARTUP MOUNT D. STARTUP OPEN

4. Quelle est la vue du dictionnaire de données qui vous permet d'afficher la liste de tous les utilisateurs de la base de données et leurs caractéristiques ?
A. DBA_USERS B. USER_USER
C. ALL_USER D. V\$SESSION

5. Pour administrer la base de données à travers la console, a-t-on besoin du Server Manager ?
A.OUI B.NON

6. Votre base de données travaille en mode « ARCHIVELOG ».
Quel est le processus qui va lire les fichiers journaux et écrire ces informations dans les fichiers journaux archivés ?
A. LGWR B. CKPT
C. DBWn D. ARCn

7. Lesquelles de ces vues vous permettent d'afficher le nom et l'emplacement du fichier de contrôle ?

- A. V\$PARAMETER
- B. V\$DATABASE
- C. V\$CONTROLFILE_RECORD_SECTION
- D. V\$CONTROLFILE

8. Vous voulez définir le multiplexage dans votre base de données. Laquelle des définitions suivantes définit pour Oracle l'emplacement des fichiers de contrôle ?

- A. Valeur spécifiée dans « CONTROL_FILES »
- B. Valeur spécifiée dans V\$DATABASE
- C. Valeur spécifiée dans « BACKGROUND_DUMP_DEST »
- D. Aucun choix, Oracle connaît automatiquement l'emplacement de ces fichiers

9. Qu'est-ce qui détermine la taille initiale d'un tablespace ?

- A. L'argument « INITIAL » de la commande « CREATE TABLESPACE... »
- B. L'argument « MINEXTENTS » de la commande « CREATE TABLESPACE... »
- C. La somme des arguments « INITIAL » et « NEXT » de la commande « CREATE TABLESPACE... »
- D. La somme des tailles des fichiers des données de la commande « CREATE TABLESPACE... »

10. Qu'est-ce qui nous permet de récupérer les données qui n'ont pas été mises à jour dans les fichiers de données suite à l'arrêt brutal du serveur ?

- Les fichiers journaux
- Les segments UNDO
- Le tablespace « SYSTEM »

11. Vous voulez réduire la fréquence des points de contrôle, les checkpoints. Laquelle de ces options vous devez choisir, qui ne modifie pas le fichier de paramètres « SPFILE » ?

- FAST_START_MTTR_TARGET
- LOG_CHECKPOINT_TIMEOUT
- Arrêter le processus « ARCn »
- Augmenter la taille des fichiers journaux

Exercice N°2

1. Soit le script suivant

```
Create table Produits( numero number ,  
                      Nom varchar2(20)  
                    )  
Tablespace data ;  
Storage (initial 500 K NEXT 100K PCTINCREASE 50 MINEXTENTS 3) ;
```

```

Create table Clients( numero number ,
                      Nom varchar2(20)
)
Tuplespace data;
Create table Categories( numero number ,
                      Nom varchar2(20)
)
Storage (initial 200 K)
Tuplespace data;

```

- a. Calculer la taille initiale du segment réservé pour chaque table
- b. Calculer le nombre d'extensions dans ces deux cas :
 - La tablespace data en gestion uniforme des extensions de taille 128 ko
 - La tablespace data en gestion automatique des extensions
2. Créer un tablespace permanent nommé «TBS_Clients» géré localement et disponible en lecture et écriture. Ce tablespace doit avoir les caractéristiques suivantes :
 - Type de tablespace: SMALLFILE
 - Gestion des extensions : locale et avec des tailles gérées par Oracle
 - Nombre de fichiers : 2
 - Nom des fichiers : «CL1.DBF» et «CL2.DBF»
 - Taille initiale des fichiers 5Mo chaqu'un
 - Le fichier «CL1.DBF» doit être de taille fixe
 - Le fichier «CL2.DBF» doit avoir une expansion automatiquement. Chaque expansion devra ajouter 2Mo au fichier. La taille du fichier ne devra pas dépasser 10Mo.
 - Le tablespace doit être en ligne
3. Ajouter au tablespace actif un fichier «CL3.DBF» physiquement localisé sur la partition E:/ de taille 500 MO.
4. On souhaite rendre les données stockées dans TBS_Clients indisponibles , que vous devez faire ?
5. Renommer le tablespace «TBS_Clients». Le nouveau nom est « TBS_Gestion_Clients».