MINISTÈRE
DE L'ÉDUCATION
NATIONALE
Liberté
Égalité
Fraternité

EAE INF 3

SESSION 2025

AGREGATION
CONCOURS EXTERNE

Section : INFORMATIQUE

ÉPREUVE SPÉCIFIQUE SELON L'OPTION CHOISIE :
- **ÉTUDE DE CAS INFORMATIQUE**
- **FONDEMENTS DE L'INFORMATIQUE**

Durée : 6 heures

L'usage de tout ouvrage de référence, de tout dictionnaire et de tout matériel électronique (y compris la calculatrice) est rigoureusement interdit.

Il appartient au candidat de vérifier qu'il a reçu un sujet complet et correspondant à l'épreuve à laquelle il se présente.

Si vous repérez ce qui vous semble être une erreur d'énoncé, vous devez le signaler très lisiblement sur votre copie, en proposer la correction et poursuivre l'épreuve en conséquence. De même, si cela vous conduit à formuler une ou plusieurs hypothèses, vous devez la (ou les) mentionner explicitement.

NB : Conformément au principe d'anonymat, votre copie ne doit comporter aucun signe distinctif, tel que nom, signature, origine, etc. Si le travail qui vous est demandé consiste notamment en la rédaction d'un projet ou d'une note, vous devrez impérativement vous abstenir de la signer ou de l'identifier.
Le fait de rendre une copie blanche est éliminatoire.

Tournez la page S.V.P.

INFORMATION AUX CANDIDATS

Vous trouverez ci-après les codes nécessaires vous permettant de compléter les rubriques figurant en en-tête de votre copie.

Ces codes doivent être reportés sur chacune des copies que vous remettrez.

► **Etude de cas informatique** :

Concours	Section/option	Epreuve	Matière
EAE	6200A	103	9424

► **Fondement de l'informatique** :

Concours	Section/option	Epreuve	Matière
EAE	6200A	103	9425

Épreuve spécifique

IMPRIMERIE NATIONALE – 25 0115 – D'après documents fournis

Système de gestion de projets informatiques

Préliminaires

L'énoncé s'inspire d'un système de gestion de projets informatiques tel que *SourceForge* ou *GitHub*. Chaque partie s'intéresse à une problématique à résoudre, présente des objectifs concrets et amène une réflexion sur les moyens de les atteindre.

Attendus. Il est attendu des candidates et des candidats des réponses construites. Les copies sont évaluées sur la précision, le soin et la clarté de la rédaction. On veillera toujours à rendre saillante la logique générale du code. Les pré-conditions des fonctions demandées n'ont pas besoin d'être vérifiées dans le code mais peuvent être énoncées sous forme de commentaires.

Dépendances. Ce sujet contient six parties. Les différentes parties et un grand nombre de leurs questions sont largement indépendants. Il est possible d'aborder les différentes parties dans un ordre quelconque en groupant les questions d'une même partie et en indiquant clairement quelle question est répondue.

Langages. Les parties I et V sont à traiter en Python. La partie II est à traiter en SQL. La partie III est à traiter en HTML/CCS et JavaScript. La partie IV est à traiter en C.

Partie I. Mise en place de chaînes de traitement d'intégration et de déploiement continus (*chaînes CI/CD*)

Une chaîne d'intégration et de déploiement continus (*chaîne CI/CD*) permet de gérer le processus de production de nouvelles versions d'un produit logiciel. La chaîne assure la bonne mise en relation entre les activités des équipes de développement et celles des équipes opérationnelles. Un système d'intégration et de déploiement continus permet la définition et l'exécution *automatique* de chaînes CI/CD.

Une chaîne CI/CD est composée de plusieurs *travaux* où chaque travail définit une suite d'*actions*. Les travaux d'une chaîne sont tous distincts. Ils peuvent être indépendants et s'exécuter en parallèle, ou avoir des dépendances qui imposent qu'un travail attende la terminaison d'un autre. Au sein d'un travail, les actions s'exécutent en séquence l'une après l'autre, chaque action ne débutant qu'après la fin de l'action qui la précède. Un exemple de chaîne CI/CD pour une application informatique est une chaîne qui a plusieurs travaux de construction indépendants (création de différentes versions de l'application), ainsi qu'un travail de packaging qui dépend de ces derniers. Chaque travail de construction est typiquement composé d'une suite d'actions qui mélange compilation et tests.

L'exécution d'une chaîne CI/CD, qui se déclenche suite à un événement ou de manière périodique, consiste en l'exécution des travaux qui la composent.

Question 1. Quel est l'intérêt de la mise en place des chaînes CI/CD ? La réponse s'appuiera sur un minimum de deux arguments.

Nous représentons les entités constitutives d'une chaîne CI/CD par trois classes Python : la classe `CICDPipeline`, la classe `Work` et la classe `Action`, qui obéissent aux relations décrites ci-dessus. Chaque instance de l'une des trois classes dispose d'un attribut entier `id` distinct (un identifiant) dont la valeur est fournie par l'usager à la création. Chacune des trois classes est munie d'une méthode `run`. La méthode `run` définit les traitements que la chaîne, le travail ou l'action doit effectuer.

Question 2. Proposer un diagramme de classes contenant les trois classes `CICDPipeline`, `Work` et `Action`, ainsi que d'éventuelles classes supplémentaires permettant de factoriser certaines fonctionnalités le cas échéant.

Question 3. Quels sont les concepts de la programmation orientée-objet qui permettent à la méthode `run` de s'appliquer à des instances relevant de trois classes différentes ?

Question 4. Écrire une implémentation de la classe `Work` en Python.

Travaux indépendants.

Dans la question 5, nous considérons le cas simplifié des chaînes CI/CD composées de travaux *indépendants*. Une telle chaîne lance l'exécution de ses travaux en parallèle à l'aide de fils d'exécution (ou *threads*) : elle crée un fil d'exécution par travail. L'exécution de la chaîne se termine quand tous les travaux, et donc les fils d'exécution correspondants, se sont terminés.

Question 5. Écrire en Python une première version de la classe `CICDPipeline`. Des rappels sur la manipulation de fils d'exécution en Python sont donnés en Annexe A.

Mise en place des dépendances.

Dans les questions 6 à 13, nous considérons le cas général d'une chaîne où les travaux peuvent avoir des dépendances. L'exécution d'un travail qui a des dépendances ne peut démarrer avant la terminaison de l'exécution de tous les travaux dont il dépend.

Question 6. Modifier le code Python de la classe `Work` écrite à la question 4 en adjoignant un troisième argument au constructeur : une liste `dependencies` des travaux dont dépend le travail construit et qui est passée comme attribut à l'objet créé.

Le bon fonctionnement de l'exécution d'une chaîne repose sur la synchronisation des fils d'exécution utilisés pour exécuter les travaux. Dans ce but, une chaîne CI/CD associe un sémaphore distinct à chacun de ses travaux. Quand un travail t_1 doit attendre la terminaison d'un autre travail t_2, il se bloque sur le sémaphore associé à t_2. Il est débloqué par t_2 à la fin de l'exécution de t_2.

Question 7. Modifier le code Python de la classe `CICDPipeline` écrite à la question 5 afin d'y inclure une gestion des sémaphores : définir un attribut, procéder à son initialisation dans le constructeur et fournir un accesseur d'entête `get_semaphore(work_id)`. Des rappels sur la synchronisation de fils d'exécution en Python sont donnés en Annexe B.

Pour que tous les travaux qui attendent un certain travail puissent être débloqués, il est nécessaire de connaître leur nombre. Comme pour les sémaphores, l'information sur le nombre de travaux que chaque travail doit débloquer est à gérer au niveau de la chaîne CI/CD correspondante.

Question 8. Dans la continuité de la question 7, modifier la classe `CICDPipeline` pour pouvoir gérer l'association entre un travail de la chaîne et le nombre de travaux qui en dépendent. Rajouter également un accesseur d'entête `get_waiting_for(work_id)`.

Question 9. Pour se synchroniser avec les autres travaux, un travail a besoin d'avoir accès aux informations (sémaphores et compteurs) de sa chaîne CI/CD. Modifier la classe `Work` pour y intégrer une référence vers sa chaîne CI/CD.

Question 10. Décrire, en langue française, une solution pour la synchronisation de travaux avec dépendances. La mettre en œuvre en réécrivant la méthode `run` de la classe `Work`.

Interblocages.

Question 11. Les dépendances entre travaux étant introduites exclusivement par l'intermédiaire du constructeur de la classe `Work`, selon la spécification de la question 6, démontrer que l'exécution de la méthode `run` de la classe `CICDPipeline` ne connaît pas d'interblocages.

Lors de la définition de grandes chaînes CI/CD avec de nombreux travaux, il peut être commode de rajouter des dépendances au fur et à mesure de la définition de différents travaux.

Question 12. Enrichir le code Python de la classe `Work` écrite à la question 6 en adjoignant un transformateur d'entête `set_dependency(prec_work)` de sorte que la commande `work.set_dependency(prec_work)` ajoute le travail `prec_work` comme dépendance devant précéder le travail `work`.

L'utilisation de la méthode `set_dependency` risque d'introduire des interblocages.

Tournez la page S.V.P.

Question 13. Citer un problème d'algorithmique classique permettant de modéliser la détection d'interblocages. Nommer un algorithme classique le résolvant. Écrire en langage Python un script permettant de détecter si une chaîne fait l'objet d'un interblocage.

Question 14. Proposer une modification du code écrit à la question 12 pour informer l'usager de l'introduction d'un interblocage dès que possible.

Partie II. Modélisation et stockage persistant

Les projets informatiques. Les chaînes CI/CD sont définies dans le cadre de *projets informatiques*. Un projet informatique est un projet de développement et de gestion d'un produit logiciel. Un projet dispose d'un identifiant unique au sein du système, ainsi que d'un nom permettant une manipulation plus facile par les équipes de développement et d'intégration. Un projet peut disposer de plusieurs chaînes CI/CD.

Les utilisateurs. Les *utilisateurs* qui travaillent sur un projet informatique sont dotés d'un *rôle*. Le rôle délimite les opérations qu'un utilisateur peut effectuer et inclut, entre autres, les rôles de *propriétaire*, *développeur* et *invité*. Un utilisateur peut être impliqué dans plusieurs projets avec des rôles distincts.

Les opérations de lecture/écriture. Les utilisateurs peuvent effectuer deux types d'opérations sur les projets : des consultations de projet (*opération de lecture*) et des modifications (*opération d'écriture*). Chaque opération d'un utilisateur sur un projet, qu'elle soit de lecture ou d'écriture, est consignée de manière persistante.

Question 15. Élaborer une modélisation relationnelle des données du système en faisant apparaître les tables `User`, `Project`, `CICD_Pipeline`, `Role` et une table de jointure `User_Project`.

Question 16. Écrire une requête SQL qui permet de créer la table contenant les données des projets informatiques.

Question 17. Écrire une requête SQL qui permet d'identifier les cinq projets ayant le plus d'utilisateurs.

Question 18. Écrire une requête SQL qui permet de trier les utilisateurs d'un projet `MyProject` selon le nombre de leurs contributions (modifications).

Partie III. Application WEB

Dans cette partie nous nous intéressons à l'application web qui permet de travailler avec les projets informatiques.

L'application web doit permettre aux utilisateurs d'accéder aux informations sur leurs projets et leur activité. Chaque utilisateur doit disposer de sa propre adresse pour consulter les informations le concernant. Toutefois, les informations concernant un projet informatique en particulier doivent être accessibles à la même adresse pour tous les utilisateurs.

Question 19. Proposer une structuration d'URL pour l'application web qui permette de consulter les informations d'un projet informatique P pour un utilisateur U.

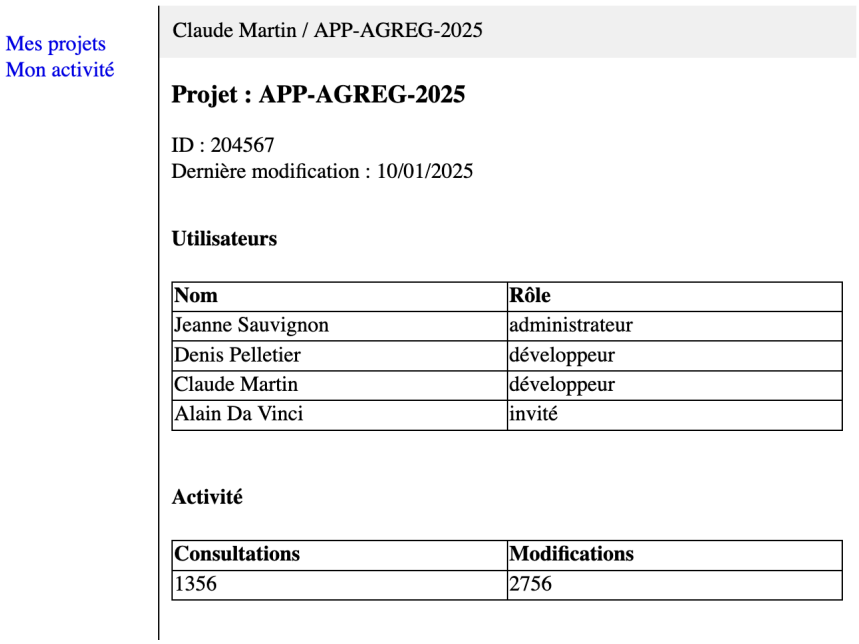

FIGURE 1 – Page web visualisée par l'utilisatrice *Claude Martin* participant au projet informatique `APP-AGREG-2025`. La page donne l'identifiant du projet et la dernière date de modification, avant de lister les utilisateurs impliqués et le nombre de consultations et de modifications. À gauche, un menu permet d'accéder aux informations sur les projets de *Claude Martin* et sur l'historique de ses contributions.

Question 20. Donner le code HTML/CSS statique qui correspond à l'esquisse d'IHM décrite dans la Figure 1.

Question 21. Enrichir le code HTML/CSS précédent pour proposer une solution supportant la mise-à-jour dynamique, i.e. sans rechargement de la page, du nom du projet, du nombre de consultations et du nombre de modifications.

Question 22. Écrire en JavaScript un script qui permet à l'application cliente de recevoir les informations de mise-à-jour du nombre de consultations et du nombre de modifications.

Tournez la page S.V.P.

Question 23. Quels sont les couches réseau et les protocoles impliqués entre la partie cliente et la partie serveur de l'application web lorsqu'un utilisateur veut consulter la page à l'adresse `https://agregsystem.fr` ?

Partie IV. Gestion de versions

Dans cette partie nous allons nous pencher sur un des aspects des plus importants d'un système de gestion de projets informatiques : la gestion des versions. La gestion des versions concerne l'ensemble des fichiers qui définissent un produit logiciel : fichiers source, scripts, fichiers de configuration, etc. Ces fichiers sont typiquement organisés de manière arborescente, avec plusieurs répertoires et sous-répertoires.

Pour gérer les différentes versions, le système utilise deux types d'objets : les *objets binaires (blobs)* et les *arbres*. Les *blobs* représentent les fichiers, alors que les *arbres* représentent les répertoires. Pour chaque objet, le système calcule une clé unique d'identification obtenue par hachage du contenu de l'objet.

Nous supposons pour la suite que le système fournit les définitions suivantes, écrites en langage C :

```
// types d'objets de gestion de versions
typedef enum {BLOB_TYPE, TREE_TYPE} object_t;

// structure décrivant un objet de gestion de versions
struct object_info {
  object_t type;
  unsigned long size;
  hash_t oid;
};

// fonction de hachage de contenu
hash_t hash(void* content, unsigned long size);

// structure de blob
struct blob {
    struct object_info header;
    void* content;
};
```

La structure `object_info` est utilisée aussi bien pour les *blobs*, que pour les *arbres*. Elle contient trois champs : le type de l'objet, donné à l'aide du type énuméré `object_t`, la taille en octets du contenu de l'objet, ainsi qu'un identifiant, `oid`, qui correspond au haché de l'objet. L'identifiant est obtenu à l'aide de la fonction `hash`, que nous supposons fournie.

La structure pour représenter un *blob* contient, en plus des informations de type, de taille et d'identifiant, le contenu du fichier correspondant, `content`. Le champ `content` est par conséquent une suite d'octets de taille variable.

Question 24. Écrire en C une fonction

```
struct blob* init_blob(void* content, unsigned long size);
```

qui initialise une structure `struct blob` à partir du contenu et de la taille d'un fichier passés en argument. La fonction renvoie la structure initialisée qui doit être allouée par la fonction. En cas d'échec, la fonction renvoie `NULL`.
Des fonctions C qui peuvent vous être utiles pour cette question sont données en Annexe C.

Arbres. Pour les *arbres*, le système de gestion de versions s'inspire de la logique des répertoires UNIX. Dans UNIX, le contenu d'un répertoire est une liste comprenant des informations sur des fichiers et des sous-répertoires. Dans le système de gestion de versions, le contenu d'un *arbre* est une liste comprenant des informations sur des blobs et des sous-arbres.

Pour la suite, nous supposons que le contenu d'un arbre est représenté à l'aide d'un tableau de taille prédéfinie. Chaque élément de ce tableau correspond à un objet distinct de l'arbre et contient deux parties : la structure `struct object_info`, qui décrit l'objet, et le nom du fichier ou du répertoire représenté par cet objet.

Question 25. En suivant la logique de la structure `struct blob` et les indications ci-dessus, déclarer en C une structure de données `struct tree` pour représenter un arbre. Donner le code C d'une fonction `struct tree* empty_tree()` qui crée un arbre vide. Bien spécifier comment un objet non défini est représenté dans le contenu de l'arbre.

Question 26. Donner le code C des deux fonctions suivantes qui permettent de changer le contenu d'un arbre

```
bool tree_add_object(struct tree* tree, char* name, struct object_info* header);
bool tree_remove_object(struct tree* tree, hash_t oid);
```

Question 27. Écrire en C une fonction

```
bool diff(void* o1, void* o2);
```

qui vérifie si les deux objets `o1` et `o2` passés en paramètre sont identiques.

Chemins. Dans UNIX, l'emplacement d'un fichier dans l'arborescence de fichiers est défini par son *chemin* depuis la racine. UNIX définit des conventions pour représenter un tel chemin sous forme d'une chaîne de caractères. Ainsi, la racine de l'arborescence de fichiers est notée **/** et le chemin `/code/test.c` correspond au fichier `test.c` qui se trouve dans le répertoire `code` qui est lui-même un sous-répertoire de la racine.

Le système de gestion de versions sauvegarde tous les objets de blob ou d'arbre dans une table associative entre l'objet et son identifiant. Dans la suite, nous supposerons les fonctions de manipulation de cette table fournies et déclarées comme suit :

```
bool put_hashtable(hash_t oid, void* o);
void* get_hashtable(hash_t oid);
```

La fonction `put_hashtable` lie un identifiant à son objet dont il est le haché (*blob* défini par `struct blob` ou *arbre* défini par `struct tree`) correspondant. La fonction `get_hashtable` permet d'avoir accès à l'objet identifié par un haché donné.

Tournez la page S.V.P.

Nous supposons également fournies les fonctions suivantes :

```
struct tree* get_root();
char** parse_path(char* path);
```

La fonction `get_root` récupère l'objet arbre correspondant au répertoire racine (/). La fonction `parse_path` renvoie la liste des chaînes de caractères qui composent le chemin `path`, terminée par NULL. Ainsi `parse_path("/code/test.c")` renvoie `{"code","test.c",NULL}`.

Question 28. Écrire en C une fonction `object_t get_type(char* path)` qui permet de connaître le type de l'objet (*blob* ou *arbre*) accessible à l'emplacement défini par le chemin `path` passé en paramètre. Nous supposons que le paramètre `path` est une chaîne de caractères non vide qui respecte les conventions de nommage des chemins UNIX.

Question 29. Déclarer en C une structure de données `struct commit` qui permet de contenir les informations relatives à la définition d'une nouvelle version d'un arbre : l'identifiant de l'arbre en question, la liste des commits précédant le commit en question, le nom de l'utilisateur qui a créé le commit et un commentaire.

Partie V. Performances à l'exécution

Dans cette partie, nous nous intéressons à la durée d'exécution d'une chaîne CI/CD. Nous supposons que toute action d'un travail CI/CD s'exécute en une unité de temps.

Question 30. Enrichir le code Python de la classe `Work`, introduite à la question 4, afin que la commande `work.duration()` renvoie le temps d'exécution du travail `work`.

Le système de gestion de projets informatique exécute les chaînes CI/CD sur des entités dédiées appelées *exécutants*. Les exécutants fournissent les ressources de calcul et de stockage nécessaires aux traitements CI/CD, et permettent une éventuelle exécution parallèle. Pour la suite, nous supposons que les dépendances d'une chaîne CI/CD sont stabilisées et ne peuvent plus changer.

Question 31. Enrichir le code Python de la classe `CICDPipeline`, introduite à la question 5, afin que la commande `pipeline.sequential_duration()` renvoie le temps d'exécution de la chaîne `pipeline` dans le cas où un seul fil d'exécution peut être exécuté à la fois.

Question 32. Enrichir le code Python de la classe `CICDPipeline`, introduite à la question 5, afin que la commande `pipeline.parallel_duration()` renvoie le temps d'exécution de la chaîne `pipeline` dans le cas où les fils d'exécution peuvent être de nombre illimité. On décrira et on justifiera avec soin l'algorithme employé.

Partie VI. Autour du CI/CD

Question 33. En quoi consiste la phase d'intégration d'un projet informatique ? Quels sont ses liens avec les tests et la gestion de versions ?

Question 34. Le déploiement correspond au processus de transfert du produit logiciel modifié vers l'environnement de production. Citer trois risques majeurs pour un déploiement.

Question 35. Que signifie le fait que l'intégration et le déploiement soient *continus* ? Quels en sont les bénéfices ?

* *
*

Tournez la page S.V.P.

ANNEXE A : `threading` — Parallélisation à base de Threads

Extrait de la documentation Python 3

Code source : `Lib/threading.py`

Objets `Threads`

La classe `Thread` représente une activité exécutée dans un fil d'exécution distinct. Il existe deux manières de spécifier l'activité : en passant un objet appelable au constructeur ou en redéfinissant la méthode `run()` dans une sous-classe. Aucune autre méthode (à l'exception du constructeur) ne doit remplacée dans une sous-classe. En d'autres termes, remplacez uniquement les méthodes `__init__()` et `run()` de cette classe.

Une fois qu'un objet fil d'exécution est créé, son activité doit être lancée en appelant la méthode `start()` du fil. Ceci invoque la méthode `run()` dans un fil d'exécution séparé.

D'autres fils d'exécution peuvent appeler la méthode `join()` d'un fil. Ceci bloque le fil appelant jusqu'à ce que le fil dont la méthode `join()` est appelée soit terminé.

Un fil d'exécution a un nom. Le nom peut être passé au constructeur, et lu ou modifié via l'attribut name.

`class threading.`**`Thread`**`(`*`group=None, target=None, name=None, args=(), kwargs=, *, daemon=None`*`)`

Les arguments sont :

`group` devrait être `None` ;

`target` est l'objet appelable qui doit être invoqué par la méthode `run()`. La valeur par défaut est `None`, ce qui signifie que rien n'est appelé.

`name` est le nom du fil d'exécution. Par défaut, un nom unique est construit de la forme `"Thread-N"` où N est un entier. Si `target` est défini, le nom est de la forme `"Thread-N (target.__name__)"`

`args` est une liste ou un tuple d'arguments pour l'invocation de `target`. La valeur par défaut est ().

`kwargs` est un dictionnaire d'arguments nommés pour l'invocation de l'objet appelable. La valeur par défaut est .

Si la sous-classe réimplémente le constructeur, elle doit s'assurer d'appeler le constructeur de la classe de base (`Thread.__init__()`) avant de faire autre chose au fil d'exécution.

`start()`

Lance l'activité du fil d'exécution.

Elle ne doit être appelée qu'une fois par objet de fil. Elle fait en sorte que la méthode `run()` de l'objet soit invoquée dans un fil d'exécution.

`run()`

Méthode représentant l'activité du fil d'exécution.

Exemple :

```
>>> from threading import Thread
>>> t = Thread(target=print, args=[1])
>>> t.run()
1
>>> t = Thread(target=print, args=(1,))
>>> t.run()
1
```

`join(timeout=None)`

Attend que le fil d'exécution se termine. Ceci bloque le fil appelant jusqu'à ce que le fil dont la méthode `join()` est appelée se termine – soit normalement, soit par une exception non gérée – ou jusqu'à ce que le délai optionnel timeout soit atteint.

Tournez la page S.V.P.

ANNEXE B : `threading` — Synchronisation de Threads en Python

Extrait de la documentation Python 3

Code source : `Lib/threading.py`

Verrous

Un verrou primitif n'appartient pas à un fil d'exécution lorsqu'il est verrouillé. En Python, c'est actuellement la méthode de synchronisation la plus bas-niveau qui soit disponible, implémentée directement par le module d'extension `_thread`.

Un verrou primitif est soit « verrouillé » soit « déverrouillé ». Il est créé dans un état déverrouillé. Il a deux méthodes, `acquire()` et `release()`. Lorsque l'état est déverrouillé, `acquire()` verrouille et se termine immédiatement. Lorsque l'état est verrouillé, `acquire()` bloque jusqu'à ce qu'un appel à `release()` provenant d'un autre fil d'exécution le déverrouille. À ce moment `acquire()` le verrouille à nouveau et rend la main. La méthode `release()` ne doit être appelée que si le verrou est verrouillé, elle le déverrouille alors et se termine immédiatement. Déverrouiller un verrou qui n'est pas verrouillé provoque une `RuntimeError`.

Lorsque plusieurs threads sont bloqués dans `acquire()` en attendant que l'état passe à déverrouillé, un seul thread continue lorsqu'un appel `release()` réinitialise l'état à déverrouillé ; lequel des threads en attente se débloque n'est pas défini et peut varier selon les implémentations.

class threading.**Lock**

La classe implémentant des verrous primitifs. Une fois qu'un fil d'exécution a acquis un verrou, tout appel consécutif pour acquérir le verrou est bloquant, jusqu'à libération du verrou. Un verrou peut être libéré par n'importe quel fil d'exécution.

acquire(blocking=True, timeout=-1)

Acquiert un verrou, bloquant ou non bloquant.

release()

Libérer un verrou. Peut être appelé par n'importe quel fil d'exécution, non seulement par celui qui a acquis le verrou.

Sémaphores

class threading.**Semaphore**(value=1)

Cette classe implémente les sémaphores.

Un sémaphore gère un compteur interne qui est décrémenté à chaque appel `acquire()` et incrémenté à chaque appel `release()`. Le compteur ne peut jamais descendre en dessous de zéro ; lorsque `acquire()` trouve qu'il est nul, il bloque, attendant qu'un autre thread appelle `release()`.

class threading.**Semaphore**(value=1)

acquire(blocking=True, timeout=None)

Acquiert un sémaphore.

release(n=1)

Libérer un sémaphore, incrémente le compteur interne de un.

Tournez la page S.V.P.

ANNEXE C : Fonctions utiles en C

NAME

memcpy – copy memory area

SYNOPSIS

```
#include <string.h>

void *
memcpy(void *restrict dst, const void *restrict src, size_t n);
```

DESCRIPTION

The **memcpy()** function copies n bytes from memory area src to memory area dst. If dst and src overlap, behavior is undefined. Applications in which dst and src might overlap should use **memmove(3)** instead.

RETURN VALUES

The **memcpy()** function returns the original value of dst.

NAME

memmove – copy byte string

SYNOPSIS

```
#include <string.h>

void *
memmove(void *dst, const void *src, size_t len);
```

Transducteurs finis

Les transducteurs finis sont des machines théoriques similaires aux automates finis permettant de transformer des mots donnés en entrée. Ils sont utilisés pour l'analyse morphologique et phonologique dans le domaine de la linguistique, mais peuvent également avoir des applications en analyse syntaxique. Certains cas particuliers de transducteurs peuvent également avoir des applications dans des machines simples, tels que des distributeurs automatiques.

Le sujet est découpé en trois parties. La première présente les transducteurs séquentiels et contient quelques exemples et implémentations. La deuxième généralise le concept en introduisant les transducteurs sous-séquentiels. Enfin, la troisième et dernière partie établit deux théorèmes qui caractérisent les fonctions séquentielles et sous-séquentielles et leurs démonstrations.

Préliminaires

Notations Si Σ est un alphabet, on note Σ^* l'ensemble des mots sur Σ. Pour $u \in \Sigma^*$, on note $|u|$ la taille de u. L'unique mot de taille 0 sera noté ε, indépendamment de l'alphabet. Pour n un entier naturel, on note :

- Σ^n l'ensemble des mots de taille **exactement** n, défini par induction par $\Sigma^0 = \{\varepsilon\}$ et $\Sigma^{n+1} = \Sigma\Sigma^n$;
- $\Sigma^{\leqslant n}$ l'ensemble des mots de taille **au plus** n, c'est-à-dire $\Sigma^{\leqslant n} = \{\varepsilon\} \cup \Sigma^1 \cup \Sigma^2 \cup \ldots \cup \Sigma^n$.

On définit un **automate fini déterministe complet** comme un quintuplet $A = (Q, \Sigma, \delta, q_0, F)$ où :

- Q est un ensemble fini dont les éléments sont appelés **états** ;
- Σ est un alphabet appelé **alphabet d'entrée** ;
- δ est une fonction de $Q \times \Sigma$ dans Q appelée **fonction de transition** ;
- $q_0 \in Q$ est appelé **état initial** ;
- $F \subseteq Q$ est l'ensemble des **état finaux**.

On étend inductivement l'ensemble de définition de δ à $Q \times \Sigma^*$ par :

- $\delta(q, \varepsilon) = q$;
- si $u \in \Sigma^*$ et $a \in \Sigma$, alors $\delta(q, ua) = \delta(\delta(q, u), a)$.

On dit qu'un mot $u \in \Sigma^*$ est **reconnu** par l'automate A si et seulement si $\delta(q_0, u) \in F$. On note $L(A)$ l'ensemble des mots reconnus par A. Un langage L est dit **rationnel** (ou régulier) s'il existe un automate A tel que $L = L(A)$.

Dépendances. Ce sujet contient plusieurs parties. Chaque partie utilise des définitions et des résultats des parties précédentes. Sauf mention explicite du contraire, les questions restent néanmoins indépendantes, au sens où toute question peut être traitée en admettant les résultats énoncés dans les questions précédentes.

Attendus. Les questions de programmation doivent être traitées en langage OCaml. On pourra utiliser toutes les fonctions des modules `Array` et `List`, ainsi que les fonctions de la bibliothèque standard (celles qui s'écrivent sans nom de module, comme `max`, `incr` ainsi que les opérateurs comme `@`). L'utilisation d'autres modules est interdite.

Il est attendu des candidates et des candidats des réponses construites. Ils seront aussi évalués sur la précision, le soin et la clarté de la rédaction.

Partie I. Transducteurs séquentiels

Informellement, un transducteur séquentiel est un automate fini déterministe complet qui écrit un mot en sortie lors de la lecture d'un mot. Dans la définition suivante, il n'y a pas de notion d'état final.

On définit un **transducteur séquentiel** (ou simplement transducteur quand il n'y a pas d'ambiguïté) comme un sextuplet $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ tel que :

- Q est un ensemble fini dont les éléments sont appelés **états** ;
- Σ est un alphabet appelé **alphabet d'entrée** ;
- Γ est un alphabet appelé **alphabet de sortie** ;
- δ est une fonction de $Q \times \Sigma$ dans Q appelée **fonction de transition** ;
- λ est une fonction de $Q \times \Sigma$ dans Γ^* appelée **fonction de sortie** ;
- $q_0 \in Q$ est appelé **état initial**.

Comme pour un automate, on étend l'ensemble de définition de δ à $Q \times \Sigma^*$. On fait de même pour λ :

- $\lambda(q, \varepsilon) = \varepsilon$;
- si $u \in \Sigma^*$ et $a \in \Sigma$, alors $\lambda(q, ua) = \lambda(q, u)\lambda(\delta(q, u), a)$.

On appelle **transition de** T et on note $q \xrightarrow{a|v} q'$ un quadruplet (q, a, q', v) tel que $(q, a) \in Q \times \Sigma$, $\delta(q, a) = q'$ et $\lambda(q, a) = v$. L'interprétation d'une telle transition est « si, depuis l'état q, on lit la lettre a en entrée, alors on va vers l'état q', et on écrit le mot v en sortie ».

Un **calcul dans** T est une suite de transitions de la forme $p_0 \xrightarrow{a_0|v_0} p_1 \xrightarrow{a_1|v_1} \dots \xrightarrow{a_{k-1}|v_{k-1}} p_k$ telle que $p_i \xrightarrow{a_i|v_i} p_{i+1}$ est une transition de T pour $i \in [\![0, k-1]\!]$. On dit qu'un tel calcul est d'origine p_0, d'extrémité p_k, d'entrée $a_0a_1\dots a_{k-1}$ et de sortie $v_0v_1\dots v_{k-1}$.

Enfin, on définit la **fonction associée** à T par $\begin{array}{rcl} \varphi_T : \Sigma^* & \longrightarrow & \Gamma^* \\ u & \longmapsto & \lambda(q_0, u) \end{array}$. Autrement dit, $\varphi_T(u) = v$ si un calcul d'entrée u est de sortie v. On dit qu'une fonction $f : \Sigma^* \to \Gamma^*$ est **séquentielle** s'il existe un transducteur séquentiel T tel que $f = \varphi_T$.

Schématiquement, on représente un transducteur séquentiel comme dans la figure 1, qui représente le transducteur $T_0 = (Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ où $Q = \{q_0, q_1, q_2\}$, $\Sigma = \{a, b\}$, $\Gamma = \{0, 1\}$ et δ et λ sont définies par les tableaux en figure 2. Sur ce transducteur, on a $\varphi_{T_0}(aabba) = 001010$ et $\varphi_{T_0}(baaab) = 01$.

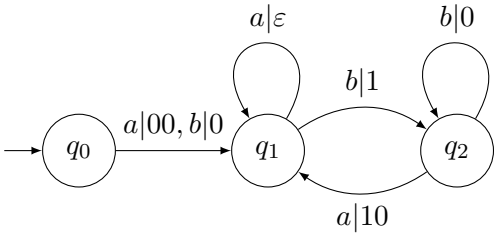

FIGURE 1 – Le transducteur séquentiel T_0

δ	a	b
q_0	q_1	q_1
q_1	q_1	q_2
q_2	q_1	q_2

λ	a	b
q_0	00	0
q_1	ε	1
q_2	10	0

FIGURE 2 – La table de transition et la table de sortie de T_0

I.1 Premiers exemples

Question 1 Pour T_0 le transducteur défini par la figure 1, déterminer $\varphi_{T_0}(bbbaaa)$. Déterminer un antécédent par φ_{T_0} de 01101.

Question 2 On considère le transducteur T_1 de la figure 3. Pour un mot $u \in \{a, b\}^*$ quelconque, décrire en français comment est construit $\varphi_{T_1}(u)$.

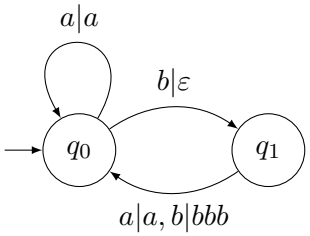

FIGURE 3 – Le transducteur séquentiel T_1

Question 3 Représenter graphiquement un transducteur séquentiel T sur $\Sigma = \Gamma = \{a, b\}$ dont la fonction séquentielle remplace toute séquence de a consécutifs en un seul a et ne modifie pas les séquences de b consécutifs. Justifier brièvement la correction de la construction.

Par exemple, on doit avoir $\varphi_T(aaaaabaaabbba) = ababbba$ et $\varphi_T(babba) = babba$.

Tournez la page S.V.P.

Pour $b \in \mathbb{N} \setminus \{0,1\}$ et $n \in \mathbb{N}^*$, une représentation en base b de n est une suite notée $[a_0 a_1 \ldots a_{m-1}]_b$ telle que $m \in \mathbb{N}^*$, $\forall i \in [\![0, m-1]\!]$, $a_i \in [\![0, b-1]\!]$ et :

$$n = \sum_{i=0}^{m-1} a_i b^i$$

Si de plus on impose $a_{m-1} \neq 0$, alors cette suite est unique.

Par exemple, la représentation en base 4 de 237 est $[1323]_4$ car $237 = 1+3\times4+2\times16+3\times64$. Notons que nous représentons ici les chiffres de poids faible à gauche.

Question 4 Représenter graphiquement un transducteur sur $\Sigma = \{0,1,2,3\}$ et $\Gamma = \{0,1\}$ qui transforme un mot u représentant une écriture de $n \in \mathbb{N}$ en base 4 en un mot v représentant une écriture de n en base 2. Justifier brièvement la correction de la construction.

Question 5 Représenter graphiquement un transducteur sur $\Sigma = \{0,1\}$ et $\Gamma = \{0,1,2,3\}$ qui transforme un mot u représentant une écriture de $n \in \mathbb{N}$ en base 2, tel que u est de taille paire (terminant éventuellement par un 0), en un mot v représentant une écriture en base 4 de n. Justifier brièvement la correction de la construction.
Par exemple, l'image de 011110 doit être 132, car $[011110]_2 = [132]_4$.

I.2 Implémentation

On représente un transducteur séquentiel en OCaml en utilisant les types suivants :

```
type lettre = int
type mot = lettre list
type etat = int
type transducteur = (etat * mot) array array
```

On choisit de représenter un alphabet Σ par $\Sigma = \{0, 1, \ldots, |\Sigma| - 1\}$ (de même pour Γ). Un mot de Σ^* sera représenté par une liste d'entiers. Par exemple, l'objet `[0; 1; 1; 2; 1; 0]` représente le mot 011210 sur l'alphabet $\{0,1,2\}$.

Lorsque le transducteur séquentiel $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ est représenté par la matrice `t`, dont la première dimension est n et la deuxième dimension est p, on a :

- $|Q| = n$;
- $|\Sigma| = p$;
- $q_0 = 0$;
- pour tous $q \in Q$ et $a \in \Sigma$, `t.(q).(a)` est un couple `(q', v)` où `q'` représente $q' = \delta(q, a)$ et `v` représente $v = \lambda(q, a)$.

Par exemple, le transducteur séquentiel T_0 de la figure 1 peut être représenté par le morceau de code suivant :

```
let t0 = [| [|(1, [0; 0]); (1, [0])|];
            [|(1, []); (2, [1])|];
            [|(1, [1; 0]); (2, [0])|] |]
```

On a assimilé ici la lettre a à `0` et la lettre b à `1`.

Question 6 Écrire une fonction `calcul (t : transducteur) (u : mot) : mot` telle que si `t` est un transducteur séquentiel représentant $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ et `u` est un mot $u \in \Sigma^*$, alors `calcul t u` renvoie le mot v représentant $\lambda(q_0, u)$.

À T fixé, cette fonction devra avoir une complexité linéaire en $|u|$ et on demande de le justifier.

Question 7 Écrire une fonction `liste_mots (p : int) (m : int) : mot list` qui prend en argument un entier p et un entier m et renvoie une liste contenant tous les mots de taille m sur l'alphabet $\{0, 1, \ldots, p-1\}$ dans un ordre arbitraire.

Question 8 En déduire une fonction `antecedent (t : transducteur) (m : int) (v : mot) : mot option` telle que si `t` est un transducteur séquentiel représentant $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0)$, `m` est un entier naturel et `v` est un mot $v \in \Gamma^*$, alors `transducteur t m v` renvoie `Some u` où u est un mot le plus court possible vérifiant $|u| \leqslant m$ et $\varphi_T(u) = v$ s'il en existe un, et `None` sinon.

On n'impose aucune restriction sur la complexité temporelle de cette fonction.

I.3 Morphismes de mots

Soit $\Phi : \Sigma^* \longrightarrow \Gamma^*$. On dit que Φ est un **morphisme de mots** si et seulement si pour tous mots u et v de Σ^*, $\Phi(uv) = \Phi(u)\Phi(v)$.

Question 9 Soit $\Phi : \Sigma^* \longrightarrow \Gamma^*$ un morphisme de mots. Déterminer en justifiant la valeur de $\Phi(\varepsilon)$.

Question 10 Soit Φ et Ψ deux morphismes de mots de Σ^* dans Γ^*. Montrer que si pour tout $a \in \Sigma$, $\Phi(a) = \Psi(a)$, alors $\Phi = \Psi$.

Question 11 Soit $\Phi : \Sigma^* \longrightarrow \Gamma^*$ un morphisme de mots. Montrer que Φ est une fonction séquentielle.

Tournez la page S.V.P.

Question 12 Soit T un transducteur séquentiel dont tous les états sont **accessibles** (c'est-à-dire que pour tout état $q \in Q$, il existe $u \in \Sigma^*$ tel que $\delta(q_0, u) = q$). Montrer qu'il y a équivalence entre :

1. φ_T est un morphisme de mots
2. pour tout $q \in Q$ et $a \in \Sigma$, $\lambda(q, a) = \lambda(q_0, a)$.

I.4 Machines de Mealy et de Moore

Un transducteur séquentiel $(Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ est appelé **machine de Mealy** si pour toute transition $q \xrightarrow{a|v} q'$, on a $|v| = 1$. C'est-à-dire que pour chaque lettre lue en entrée, la machine écrit une lettre en sortie. La fonction de sortie peut être vue comme définie de $Q \times \Sigma$ dans Γ.

Une machine de Mealy est appelée **machine de Moore** si de plus pour toute paire de transitions $q_1 \xrightarrow{a_1|b_1} q$ et $q_2 \xrightarrow{a_2|b_2} q$, on a $b_1 = b_2$. Autrement dit, la lettre écrite en sortie ne dépend que de l'état dans lequel on arrive lors de l'exécution d'une transition.

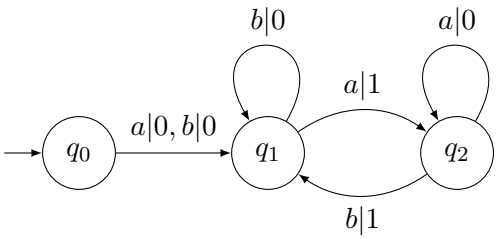

FIGURE 4 – Une machine de Mealy

On rappelle qu'une représentation en base b commence par les chiffres de poids faible.

Question 13 Représenter graphiquement une machine de Mealy sur $\Sigma = \Gamma = \{0, 1\}$ qui transforme un mot u représentant l'écriture de $x \in \mathbb{N}$ en base 2 en le mot v représentant l'écriture en base 2 de $2x$ modulo $2^{|u|}$. Justifier brièvement la correction de la construction.

Question 14 Représenter graphiquement une machine de Moore sur $\Sigma = \Gamma = \{0, 1\}$ qui transforme un mot u représentant l'écriture de $x \in \mathbb{N}$ en base 2 en le mot v représentant l'écriture en base 2 de $x + 1$ modulo $2^{|u|}$. Justifier brièvement la correction de la construction.

Question 15 Écrire une fonction `est_mealy (t : transducteur) : bool` qui détermine si un transducteur est une machine de Mealy ou non. Déterminer sa complexité temporelle dans le pire cas.

Question 16 Soit M une machine de Mealy. Montrer qu'il existe une machine de Moore M' telle que $\varphi_{M'} = \varphi_M$.

Une machine de Mealy est appelée **machine de Moore alternative** si pour toute paire de transitions $q \xrightarrow{a_1|b_1} q_1$ et $q \xrightarrow{a_2|b_2} q_2$, on a $b_1 = b_2$. Autrement dit, la lettre écrite en sortie ne dépend que de l'état duquel **on part** lors de l'exécution d'une transition. La définition est similaire à celle d'une machine de Moore, mais c'est cette fois-ci l'état de départ qui détermine la lettre en sortie.

Question 17 Montrer qu'il existe une machine de Mealy M à moins de deux états telle qu'aucune machine de Moore alternative n'est équivalente à M.

Partie II. Transducteurs sous-séquentiels

Dans cette partie, on étend la définition des transducteurs séquentiels en autorisant l'écriture d'un préfixe et d'un suffixe lors du calcul de l'image d'un mot.

II.1 Définition et exemples

On définit un transducteur sous-séquentiel comme un septuplet $(Q, \Sigma, \Gamma, \delta, \lambda, q_0, \rho)$ tel que $(Q, \Sigma, \Gamma, \delta, \lambda, q_0)$ est un transducteur séquentiel et ρ est une fonction de Q dans Γ^* appelée **fonction de suffixe**.

Si $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0, \rho)$ est un transducteur sous-séquentiel, on définit la fonction associée à T comme la fonction $\begin{array}{rcl} \varphi_T : \Sigma^* & \longrightarrow & \Gamma^* \\ u & \longmapsto & \lambda(q_0, u)\rho(\delta(q_0, u)) \end{array}$. Autrement dit l'image de u est constituée de la sortie du calcul d'entrée u suivie du suffixe associé à l'état final du calcul.

On représente graphiquement un transducteur sous-séquentiel de la même façon qu'un transducteur séquentiel, en rajoutant des flèches sortantes étiquetées par les suffixes. Par exemple, la figure 5 représente un transducteur sous-séquentiel T_2 vérifiant $\rho(q_0) = 00$, $\rho(q_1) = \varepsilon$ et $\rho(q_2) = 1$.

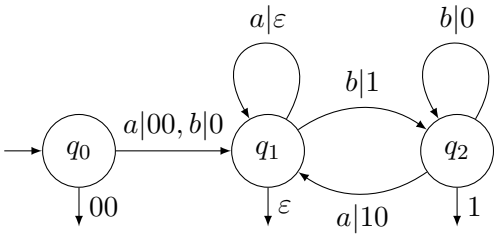

FIGURE 5 – Le transducteur sous-séquentiel T_2

On a par exemple $\varphi_{T_2}(\varepsilon) = 00$ et $\varphi_{T_2}(aabbaa) = 001010$.

Enfin, si f est une fonction de Σ^* dans Γ^*, on dit que f est une **fonction sous-séquentielle** s'il existe un transducteur sous-séquentiel T tel que $f = \varphi_T$.

Question 18 Expliquer comment modifier la machine de Moore de la question 14 en un transducteur sous-séquentiel qui fait une addition exacte, c'est-à-dire sans modulo.

Question 19 On considère le transducteur sous-séquentiel T_3 de la figure 6. Montrer que si u est un mot non vide qui représente une écriture de $x \in \mathbb{N}$ en base 2, alors $\varphi_{T_3}(u)$ représente une écriture de $3x$ en base 2.

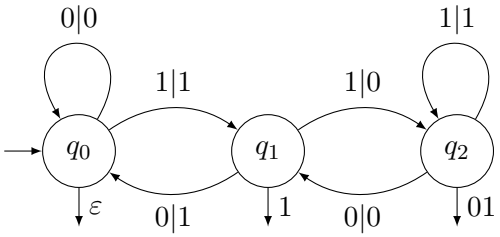

FIGURE 6 – Le transducteur sous-séquentiel T_3

Question 20 Représenter graphiquement sans justifier un transducteur sous-séquentiel à 5 états sur $\Sigma = \Gamma = \{0, 1\}$ qui transforme un mot u non vide représentant une écriture de $x \in \mathbb{N}$ en base 2 en un mot v représentant une écriture en base 2 de $5x$.

II.2 Langages rationnels

Dans cette sous-partie, on considère $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0, \rho)$ un transducteur sous-séquentiel.

Question 21 Montrer que si L est un langage rationnel sur Σ, alors $\varphi_T(L)$ est un langage rationnel sur Γ.

Indication : on pourra admettre qu'un automate fini dont les transitions sont étiquetées par des mots (plutôt que des lettres) reconnaît aussi un langage rationnel.

Question 22 Montrer que la réciproque de la question précédente est fausse.

Soit L un langage sur Γ. On appelle image réciproque de L par T, notée $\varphi_T^{-1}(L)$, l'ensemble des mots dont l'image par φ_T est dans L, c'est-à-dire $\varphi_T^{-1}(L) = \{u \in \Sigma^* \mid \varphi_T(u) \in L\}$.

Question 23 Montrer que si L est un langage rationnel sur Γ, alors $\varphi_T^{-1}(L)$ est un langage rationnel sur Σ. On explicitera un automate fini, construit à partir de T et d'un automate

reconnaissant L, et on montrera qu'il reconnait exactement $\varphi_T^{-1}(L)$.

Indication : on commencera par traiter le cas où T est séquentiel.

II.3 Fonctions sous-séquentielles

Question 24 Soit $f : \Sigma^* \longrightarrow \Gamma^*$ et $g : \Gamma^* \longrightarrow \Lambda^*$ deux fonctions sous-séquentielles. Montrer que $g \circ f$ est une fonction sous-séquentielle.

On pourra commencer par traiter le cas de fonctions séquentielles.

Soit $f : \Sigma^* \longrightarrow \Gamma^*$ une fonction. On dit que f **conserve les préfixes** si $f(\varepsilon) = \varepsilon$ et si pour tous mots u et v de Σ^*, si u est un préfixe de v, alors $f(u)$ est un préfixe de $f(v)$.

Question 25 Montrer par un exemple qu'il existe une fonction sous-séquentielle qui ne conserve pas les préfixes.

Soit $f : \Sigma^* \longrightarrow \Gamma^*$ une fonction. On cherche à montrer l'équivalence entre les deux propositions suivantes :

(a) f est une fonction séquentielle.

(b) f est une fonction sous-séquentielle et f conserve les préfixes.

Question 26 Montrer l'implication (a) $\Rightarrow$ (b) de l'équivalence précédente.

Question 27 Soit $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0, \rho)$ un transducteur sous-séquentiel tel que φ_T conserve les préfixes. On suppose que tous les états de Q sont accessibles. Montrer que pour tout $q \in Q$ et $a \in \Sigma$, il existe un mot qu'on notera $\lambda'(q, a) \in \Gamma^*$ tel que $\rho(q)\lambda'(q, a) = \lambda(q, a)\rho(\delta(q, a))$.

Question 28 Avec les notations de la question précédente, on étend l'ensemble de définition de λ' à $Q \times \Sigma^*$ comme pour les fonctions de sortie des transducteurs. Montrer que pour $q \in Q$ et $u \in \Sigma^*$, $\rho(q)\lambda'(q, u) = \lambda(q, u)\rho(\delta(q, u))$.

Question 29 En déduire l'implication (b) $\Rightarrow$ (a) de l'équivalence précédente.

Tournez la page S.V.P.

Partie III. Théorèmes de Ginsburg-Rose et Choffrut

III.1 Distance préfixe

Pour u et v deux mots de Σ^*, on note $u \wedge v$ le plus long préfixe commun à u et v. On définit la **distance préfixe** entre u et v par $d(u,v) = |u| + |v| - 2|u \wedge v|$. Par exemple, si $u = abbaba$ et $v = abbbbaaab$, alors $u \wedge v = abb$ et $d(u,v) = |u| + |v| - 2|u \wedge v| = 6 + 9 - 2 \times 3 = 9$.

Question 30 Montrer que la fonction de distance préfixe est une distance, c'est-à-dire que pour tous u, v, w mots de Σ^*, d vérifie les quatre propriétés suivantes :

- $d(u,v) = d(v,u)$;
- $d(u,v) \geqslant 0$;
- $d(u,v) = 0 \Leftrightarrow u = v$;
- $d(u,w) \leqslant d(u,v) + d(v,w)$.

Question 31 Soit $X \subseteq \Sigma^*$ un ensemble de mots non vide. Soit u le plus long préfixe commun à tous les mots de X. Montrer que pour $v \in X$, $d(u,v) \leqslant \max_{(x,y)\in X^2} d(x,y)$.

On dit qu'une fonction $f : \Sigma^* \longrightarrow \Gamma^*$ est **lipschitzienne** s'il existe un entier K strictement positif tel que :

$$\forall (u,v) \in \Sigma^* \times \Sigma^*, d(f(u), f(v)) \leqslant K d(u,v)$$

On cherche dans cette partie à montrer les théorèmes suivants :

- **Théorème de Ginsburg et Rose (1966)** : Soit $f : \Sigma^* \longrightarrow \Gamma^*$. La fonction f est une fonction séquentielle si et seulement si elle vérifie les trois propriétés suivantes :
  (a) f conserve les préfixes ;
  (b) pour tout langage rationnel L sur Γ^*, $f^{-1}(L)$ est rationnel ;
  (c) f est lipschitzienne.
- **Théorème de Choffrut (1978)** : Soit $f : \Sigma^* \longrightarrow \Gamma^*$. La fonction f est une fonction sous-séquentielle si et seulement si elle vérifie les deux propriétés suivantes :
  (a) pour tout langage rationnel L sur Γ^*, $f^{-1}(L)$ est rationnel ;
  (b) f est lipschitzienne.

Question 32 On suppose vrai le théorème de Choffrut. Montrer le théorème de Ginsburg et Rose.

Question 33 Soit $f : \Sigma^* \longrightarrow \Gamma^*$ une fonction sous-séquentielle. Montrer que f est lipschitzienne. En déduire le sens direct du théorème de Choffrut.

Indication : si $f = \varphi_T$ avec $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0, \rho)$, on posera $K = K_1 + 2K_2$, où K_1 et K_2 sont définies par $K_1 = \max\{|\lambda(q,a)| \mid q \in Q, a \in \Sigma\}$ et $K_2 = \max\{|\rho(q)| \mid q \in Q\}$.

La suite de cette partie a pour objectif de montrer le sens réciproque du théorème de Choffrut. On considère pour la suite une fonction $f : \Sigma^* \longrightarrow \Gamma^*$ qui vérifie :

(a) pour tout langage rationnel L sur Γ^*, $f^{-1}(L)$ est rationnel ;

(b) f est lipschitzienne.

On cherche à montrer que f est une fonction sous-séquentielle.

III.2 Découpage de $f(uv)$

Pour $u \in \Sigma^*$, on définit $\pi(u)$ comme le plus long préfixe commun aux mots $f(uv)$, pour $v \in \Sigma^{\leqslant 1}$. Ainsi, pour tout $v \in \Sigma^{\leqslant 1}$, il existe $\sigma_u(v)$ tel que :

$$f(uv) = \pi(u)\sigma_u(v)$$

Question 34 Montrer qu'il existe un entier naturel M tel que pour tout $u \in \Sigma^*$ et tous v_1, v_2 dans $\Sigma^{\leqslant 1}$, on a :

$$d(f(uv_1), f(uv_2)) \leqslant M$$

Question 35 En déduire que pour tout $u \in \Sigma^*$ et $v \in \Sigma^{\leqslant 1}$, $|\sigma_u(v)| \leqslant M$ puis que l'ensemble $S = \{\sigma_u \mid u \in \Sigma^*\}$ est fini.

Pour $s \in S$, on pose alors $X_s = \{u \in \Sigma^* \mid \sigma_u = s\}$.

III.3 Construction d'un transducteur sous-séquentiel

On admet temporairement que pour $s \in S$, l'ensemble X_s est un langage rationnel sur Σ.

Question 36 Soit L_1 et L_2 deux langages rationnels sur Σ. Montrer qu'il existe un automate fini déterministe complet $A = (Q, \Sigma, \delta, q_0, F)$ ainsi que deux sous-ensembles $Q_1 \subseteq Q$ et $Q_2 \subseteq Q$ tels que $L_1 = \{u \in \Sigma^* \mid \delta(q_0, u) \in Q_1\}$ et $L_2 = \{u \in \Sigma^* \mid \delta(q_0, u) \in Q_2\}$.

Un automate fini déterministe est dit **standard** si aucune transition ne pointe vers l'état initial.

Question 37 Montrer qu'il existe un automate fini déterministe complet et standard $A = (Q, \Sigma, \delta, q_0, F)$ tel qu'en posant, pour tout $s \in S$, $Q_s = \{\delta(q_0, u) \mid u \in X_s\}$, alors on a :

Tournez la page S.V.P.

– $\forall s, s' \in S, Q_s \cap Q_{s'} \neq \emptyset \Leftrightarrow s = s'$;
– $\bigcup_{s \in S} Q_s = Q$.

Pour $q \in Q$, on pose $U_q = \{u \in \Sigma^* \mid \delta(q_0, u) = q\}$. On pose de plus $\beta(q)$ le plus long suffixe des $\pi(u)$, pour $u \in U_q$. Il existe donc une fonction α telle que pour $u \in U_q$:

$$\pi(u) = \alpha(u)\beta(q)$$

On veut montrer, par disjonction de cas, que pour tout $u \in U_q$ et $a \in \Sigma$, $\alpha(u)$ est un préfixe de $\alpha(ua)$. Pour cela, on note $s \in S$ tel que $q \in Q_s$, $q' = \delta(q, a)$ et s' tel que $q' \in Q_{s'}$.

Question 38 Montrer que $\alpha(u)\beta(q)s(a) = \alpha(ua)\beta(q')s'(\varepsilon)$.

On suppose pour les deux questions suivantes que $|s(a)| \leqslant |s'(\varepsilon)|$.

Question 39 Montrer qu'il existe $v \in \Gamma^*$ tel que $\pi(u) = \pi(ua)v$, et que v ne dépend pas du mot u choisi dans U_q.

Question 40 Montrer que nécessairement $|\beta(q')v| \leqslant |\beta(q)|$, puis qu'il existe $w \in \Gamma^*$ tel que $\alpha(u)w = \alpha(ua)$, et que w ne dépend pas du mot u choisi dans U_q.

On admet que le cas $|s(a)| > |s'(\varepsilon)|$ se traite de manière similaire. En conclusion, il existe une fonction $\lambda : Q \times \Sigma \to \Gamma$ telle que pour tout $q \in Q$, $u \in U_q$ et $a \in \Sigma$, $\alpha(u)\lambda(q, a) = \alpha(ua)$.

On pose, pour $q \in Q$ tel que $q \in X_s$, $\rho(q) = \beta(q)s(\varepsilon)$.

Question 41 En considérant le transducteur sous-séquentiel $T = (Q, \Sigma, \Gamma, \delta, \lambda, q_0, \rho)$, montrer le théorème de Choffrut.

III.4 Rationalité des X_s

On cherche dans cette partie à montrer que pour tout $s \in S$, l'ensemble X_s est un langage rationnel. Pour cela, on définit successivement plusieurs familles intermédiaires de langages permettant de reconstruire X_s.

Pour $i \in \{0, 1, \ldots, 2M\}$ et $z \in \Gamma^M$, l'entier M étant celui définit à la question 34, on définit $B(i, z)$ par :

$$B(i, z) = \{x \in \Gamma^* \mid |x| \equiv i \pmod{2M+1} \text{ et } (x \in \Gamma^* z \text{ ou } z \in \Gamma^* x)\}$$

c'est-à-dire que $x \in B(i, z)$ si et seulement si la taille de x est de la forme $|x| = (2M+1) \times k + i$ avec k entier, et x est un suffixe de z ou que z est un suffixe de x.

Question 42 Montrer que $B(i,z)$ est un langage rationnel sur Γ.

Pour $i \in \{0,1,\ldots,2M\}$, $z \in \Gamma^M$, $s \in S$ et $v \in \Sigma^{\leqslant 1}$, on définit $C_s(i,z,v)$ par :

$$C_s(i,z,v) = \{u \in \Sigma^* \mid f(uv) \in B(i,z)s(v)\}$$

Question 43 Montrer que $C_s(i,z,v)$ est un langage rationnel sur Σ.

Enfin, pour $s \in S$, on définit Y_s par :

$$Y_s = \bigcup_{i=0}^{2M} \bigcup_{z\in\Gamma^M} \bigcap_{v\in\Sigma^{\leqslant 1}} C_s(i,z,v)$$

Le résultat de la question 43 et les propriétés de stabilité des langages rationnels garantissent que Y_s est un langage rationnel sur Σ.

Question 44 Montrer que pour $s \in S$, $X_s \subseteq Y_s$.

On veut maintenant démontrer que $Y_s \subseteq X_s$. On considère $y \in Y_s$, et $i \in \{0,1,\ldots,2M\}$, $z \in \Gamma^M$ tels que $y \in \bigcap_{v\in\Sigma^{\leqslant 1}} C_s(i,z,v)$.

Question 45 Soient $v_1, v_2 \in \Sigma^{\leqslant 1}$ et $b_1, b_2 \in B(i,z)$ tels que $f(yv_1) = b_1 s(v_1)$ et $f(yv_2) = b_2 s(v_2)$. Montrer que $|b_1| = |b_2|$. Pour la suite, on pose n_s cette taille commune.
Indication : on pourra utiliser, après l'avoir montrée, l'inégalité $||u| - |v|| \leqslant d(u,v)$.

Question 46 Montrer qu'il existe $v_1, v_2 \in \Sigma^{\leqslant 1}$ tels que $s(v_1) \wedge s(v_2) = \varepsilon$. En déduire que $|\pi(y)| \leqslant n_s$.

On admet que, de même, il existe $v_1, v_2 \in \Sigma^{\leqslant 1}$ tels que $\sigma_y(v_1) \wedge \sigma_y(v_2) = \varepsilon$.

Question 47 En considérant $v_1, v_2 \in \Sigma^{\leqslant 1}$ tels que $\sigma_y(v_1) \wedge \sigma_y(v_2) = \varepsilon$, montrer que $s(v_1)$ est un suffixe de $\sigma_y(v_1)$ et que $s(v_2)$ est un suffixe de $\sigma_y(v_2)$. En déduire que $|\pi(y)| = n_s$.

Question 48 En déduire que $Y_s \subseteq X_s$.

* *
*