

Portail IE / INF2

Contrôle continu

20 février 2020
30 mn

Ce sujet comporte 2 pages. Répondre de façon concise mais précise aux questions suivantes en utilisant les cadres. Aucun document n'est autorisé, hormis un aide-mémoire personnel, manuscrit, de format A4, recto seul. Remplir le cadre « Identification » dès le début de l'épreuve.

1. Quel est le sujet du TP de recherche documentaire ?

Green IT (*What else?*) ou l'étude des TIC (le « numérique » ou *IT* en anglais) sous l'angle du développement durable.

Certains auront pu croire que je me moquais d'eux en posant cette question, mais pas du tout. Cette question est une prime à l'intention de ceux qui sont un minimum dans le coup. Des questions plus subtiles vont suivre évidemment.

Résultat, plus de 15 % de mauvaises réponses sur les questions 1 et 2 ! En particulier, des réponses qui donnaient le sujet de l'an dernier ! Je suppose qu'elles viennent de redoublants qui appliquent cette année les méthodes qui les ont fait redoubler l'an dernier. On ne change pas une équipe qui perd ; belle lucidité !

Attention : dans « Green IT », « IT » c'est pour *information technology* (équivalent de « le numérique »), et ce n'est donc pas Green Hit ! Ne pas confondre non plus avec Green500, qui existe mais n'est pas la même chose.

2. Quel thème allez-vous développer ? Ne pas développer ici.

Ici, il n'y a pas de vraiment mauvaise réponse, pour peu qu'on ait su répondre à la question 1. Par contre, une réponse qui est juste une paraphrase de Green IT n'apporte rien. Elle rapporte quand même une partie des points pour la distinguer des réponses franchement absurdes.

3. On a vu qu'il vaut mieux comprendre une entité informatique comme manipulant des symboles que des nombres. Donner 2 exemples.

Au tout début on a vu l'exemple du compteur de parité qui ne faisait pas vraiment un comptage ; il oscillait juste entre les états « pair » et « impair », alors qu'on aurait pu se dire qu'il fallait compter puis tester la parité.

Puis, on a vu l'exemple de la représentation de l'ancienneté des dernières utilisations des pages du cache qui ne faisait qu'utiliser les symboles « *used* » et « *not used* » représentés par 1 bit, alors qu'on aurait pu se dire qu'il fallait représenter des dates à la *vs* près.

Attention : ici, la question est « donner 2 exemples », pas « donner une définition » ni « donner une explication ».

4. Rappeler quelques éléments de la partie vue commande de la description d'un système de fichiers.

Il s'agit de commandes d'exploration et de visualisation comme `cd`, `ls`, `pwd`, et `cat`. Elles s'exécutent toujours par rapport à des répertoires conventionnels : répertoire racine, répertoire d'accueil, répertoire courant, et répertoire parent. Elles utilisent des chemins d'accès absolus (= qui partent de la racine) ou relatifs (= qui partent du répertoire courant).

Vu la place laissée pour répondre, il y avait une incitation à développer un peu : `'.'` dénote le répertoire courant, `'..'` le répertoire parent, `'/'` le répertoire racine, et `'~'` le répertoire d'accueil, `ls` liste le contenu d'un répertoire ou les caractéristiques d'un élément du répertoire courant, qui est le répertoire d'accueil par défaut, `cd` permet de changer de répertoire courant ou de revenir au répertoire d'accueil, `pwd` permet d'afficher le répertoire courant, et `cat` permet d'afficher le contenu d'un fichier texte. Ici, j'ai développé beaucoup ; un peu suffisait.

Attention : des réponses comme « la vue commande diffère de la vue graphique » disent des choses vraies, mais qui ne répondent pas à la question. C'est différent de dire des choses vraies que de dire des choses pertinentes. De la même façon, rappeler qu'un système peut être décrit en termes de son modèle, sa vue et son contrôle, vous fait dire des choses vraies mais pas pertinentes.

5. Rappelez quelques éléments de sa partie contrôle.

Il s'agit de commandes de création, de suppression et de modification comme `mkdir`, `touch`, `rm`, `rmdir`, et `mv` : `mkdir` crée un répertoire dans le répertoire courant, `touch` crée un fichier dans le répertoire courant, `rm` supprime un fichier, `rmdir` supprime un répertoire, et `mv` renomme un fichier ou un répertoire.

6. Rappeler les éléments de sa partie modèle. Les commenter rapidement en vue de la question suivante.

Des fichiers : de taille quelconque et de nom arbitraire donné par l'utilisateur. Ils contiennent des données que le SGF ne sait pas interpréter, mais que celui-ci se doit de préserver intactes. Les données d'un fichier sont utilisées par des applications qui les utilisent lorsqu'elles (les données) sont stockées en mémoire.

Des répertoires : de taille quelconque et de nom arbitraire donné par l'utilisateur. Ils contiennent des données que le SGF se doit de préserver intactes, et qu'il sait interpréter ; ces données constituent le dictionnaire des habitants (fichiers ou répertoires) de chaque répertoire. La relation « habiter dans » forme essentiellement un arbre (au sens où chaque élément d'un système de fichier n'habiterait qu'un répertoire), mais pour être plus conforme à la réalité il vaut mieux convenir que cette relation forme un graphe car il est utile que certains éléments habitent plusieurs répertoires.

Attention : la hiérarchisation ne consiste pas en une division des fichiers en « sous-fichiers ». C'est juste l'appartenance de répertoires à des répertoires. On peut

imaginer des applications qui éclatent des fichiers en sous-fichiers répartis dans des répertoires différents, mais ce n'est pas le problème du système de fichiers.

Ne confondez pas « en forme de graphe » et « graphique ». En forme de graphe signifie essentiellement qu'il y a des objets et des relations a priori quelconques entre les objets. Graphique (dans interface, tablette, ..., et vue) signifie visuel. Bien sûr, on peut envisager la représentation graphique d'un graphe, avec des points et des traits, mais ce n'est pas une raison pour confondre.

Ne confondez pas variables et variées ! Variable est ce qui change dans le temps, varié est ce qui change d'un individu à l'autre. Dans notre amphithéâtre, la couleur de nos yeux est variée, pas variable. Et la vitesse du vent en un point est variable, pas variée. Et pour les tailles de fichiers, ce qui importe d'abord est qu'elles soient variées et imprévisibles. On peut avoir des images petites ou grosses, des textes petits ou gros, des programmes petits ou gros ; il n'y a pas de loi, c'est arbitraire.

Ne parlez pas de disque dur. Un système de fichiers peut s'appuyer sur autre chose qu'un disque dur ; il n'y a vraisemblablement pas de disque dur dans votre téléphone, mais il y a presque certainement un système de fichiers. Les questions 4 à 6 portent sur la définition d'un système de fichiers, pas sur son implémentation sur un disque dur ; ça c'est le sujet des questions 7 à 10.

Et ne confondez pas arbitraire et aléatoire ! Il n'y a rien d'aléatoire dans la taille d'un fichier, et encore moins dans son nom !

7. Rappeler quelles caractéristiques d'un disque dur semblent contradictoires avec celles d'un système de fichiers.

Le disque dur est formé de pistes qui délimitent des secteurs de taille fixe, de nom conventionnel (n° piste, n° de secteur dans la piste), dans une structure essentiellement plate. De plus, le temps d'accès est environ 10^6 fois celui d'un accès mémoire.

Dans cette réponse et la précédente, j'ai souligné les termes qui se contredisent.

Attention à un détail. Ce qui est gênant dans la taille des fichiers n'est pas qu'elle varie dans le temps, c-à-d. qu'elle soit variable ; ça c'est juste une circonstance aggravante. Ce qui est gênant est qu'elle varie d'un fichier à l'autre, c-à-d. qu'elle soit variée. Même si on interdisait de diminuer ou augmenter la taille d'un fichier, on aurait un problème à résoudre.

Présenter comme une limitation le fait que le disque dur a une taille fixe est absurde. Votre vélo a aussi une taille fixe. C'est une propriété banale de n'importe quel objet physique. Et inversement, c'est une propriété qui ne s'applique pas sans précaution aux objets informatiques. Ex. parler de la taille d'un système de fichiers n'a pas grand sens. Tout au plus peut-on parler de sa capacité maximale ou de la taille de ce qu'il contient.

Faites aussi attention au bon usage des termes physiques : une durée est courte ou longue, mais pas rapide ni lente. Pareillement, une vitesse est petite ou grande, mais

ni courte ni longue, ni rapide ni lente. Bien différencier les propriétés qualitatives d'un objet, ex. être rapide ou non, être lourd ou non, être coûteux ou non, des propriétés quantitatives des objets, ex. la vitesse, la masse, le coût, qui ont des valeurs plus ou moins grandes.

Et faites vraiment attention au mot « instantané ». Aucun objet physique, ex. un disque dur ou un transistor, ne peut changer d'état instantanément ! Tout ce qu'on peut dire est que certains objets changent d'état plus vite que d'autres. « Instantané » fait partie du vocabulaire de la pensée magique qui veut faire croire que les opérations informatiques sont instantanées. De la même façon, « infini » en informatique fait partie le plus souvent du vocabulaire de la pensée magique ; rien n'est infiniment petit, ni infiniment rapide, ni infiniment grand ! Votre job d'étudiant en sciences est d'abandonner la pensée magique.

Et faites absolument attention à ne pas mélanger « contradictoire » et « contraposée » ! En logique, $\text{non } B \Rightarrow \text{non } A$ est la contraposée de $A \Rightarrow B$ et ces deux formules sont absolument équivalentes. En logique, A et B sont contradictoires si $A \wedge B$ est faux. Et là, les formules A et B ne peuvent pas être équivalentes ! La logique est un grand outil pour éviter la pensée magique, mais seulement si on s'en sert correctement.

8. Un principe très important a été présenté pour résoudre une de ces contradictions, lequel ? Développer un peu.

Il s'agit du principe de localité. Il se décompose en deux sous-principes :

- Principe de localité temporelle : selon lequel une opération réalisée dans un passé récent a de forte chance d'être réalisée à nouveau dans un futur proche.
- Principe de localité spatial : selon lequel une opération réalisée dans un passé récent sur une certaine donnée a de forte chance d'être réalisée dans un futur proche sur une donnée voisine.

Ce principe permet donc de prévoir le futur proche en observant le passé récent. Il permet à un système de s'adapter à des événements dont la survenue est probable.

Attention : le principe de localité ne doit pas être compris comme une loi universelle. C'est seulement une condition à valider dans chaque contexte. Elle est généralement pour les systèmes de fichiers, mais aussi pour les accès mémoire par les processeurs, ou encore pour les accès internet.

Une fois attesté, le principe de localité justifie deux types d'actions :

- Stocker les résultats d'opérations du passé récent au cas où on voudrait les réaliser à nouveau dans un futur proche. On appelle cela un cache, et ça n'est intéressant que si stocker coûte moins cher que réaliser à nouveau.
- Réaliser plus que l'opération demandée en l'appliquant aussi aux données voisines de la donnée visée. On appelle cela un calcul spéculatif, et ça n'est intéressant que si exécuter plusieurs opérations à la fois coûte moins cher que les exécuter une par une.

9. Quel dispositif du système de fichiers sur disque exploite ce principe ? Expliquer son rôle et l'avantage qu'on en attend.

Le cache utilise ces deux principes :

- Localité spatiale : des secteurs entiers sont lus et stockés dans le cache, plutôt que des données individuelles car il est probable que les données voisines vont servir prochainement elles aussi.
- Localité temporelle : les secteurs lus les plus récemment restent dans le cache car on s'attend à ce qu'ils doivent être relus prochainement.

Attention : on dit « cache » et pas « cash », et encore moins « cach ». De la même façon, on conjugue le verbe « stocker » ; « je stocke », « tu stockes », « il stocke », et pas « je stock », « tu stock », « il stock ». Plus généralement, quand on incorpore des termes étrangers dans le français, on leur applique la grammaire française : « nous pushons », « vous commitez », « ils scannent ». Et en général, on essaie de trouver un terme français qui va bien. Utiliser un langage de *geeks* ne suffira pas à faire de vous des informaticiens.

Il est vraiment important de distinguer le principe qui est appliqué et la façon de l'appliquer. En particulier, le principe de localité n'a pas pour « objectif » de stocker en mémoire des résultats de calculs passés. Il n'a pas non plus été « créé ». Le principe de localité est un principe d'observation purement statistique qui, quand il est vérifié, suggère de se souvenir des opérations récentes, voire même de les généraliser un peu. En particulier, il est important de l'énoncer sans recourir aux concepts de fichiers, secteurs et disque dur, car il s'applique aussi à Internet, au web, aux accès à la mémoire vive, ou à l'organisation des datacentres. Une des premières publications sur le sujet date de 1972, par Peter Denning, le conférencier de la vidéo présentée en début de la section cours de la page Moodle. Ce principe a donc près de 50 ans ! En voici une illustration d'époque : <http://denninginstitute.com/modules/vm/orange/realprog.html> .

À propos de ces questions, beaucoup ont répondu par la réponse à une question de l'an dernier sur l'architecture de Von Neumann ! Cela témoigne d'un travail à rebrousse-poil. Si je prends la peine de rédiger ces éléments de réponse c'est pour vous aider à comprendre vos résultats, mais en aucun cas cela est destiné à se substituer au cours. Le travail dans le sens du poil, c'est de venir en CM-TD-TP (tous obligatoires, rappelez-le vous), d'y faire le maximum pour que ce ne soit pas une perte de temps, de relire les documents déposés sur Moodle, et in fine de regarder les CC du passé pour vérifier que vous savez y répondre. Et si vous ne savez pas, retour aux documents.

10. Rappeler brièvement les rôles des bits *used* et *dirty*.



Afin d'économiser les accès disque on n'écrit pas directement dans le disque mais uniquement dans le cache. En conséquence, certains secteurs ont une version à jour dans le cache et une version ancienne sur le disque. On dit qu'ils sont *dirty*, et on veillera à resynchroniser les deux versions en écrivant le secteur sale sur le disque.

Afin de garder dans le cache les copies de secteurs qui ont le plus de pertinence à y résider on marque les copies de secteurs dans le cache du symbole *used* selon qu'elles ont été utilisées dans un passé récent. Pour ce faire, les bits *used* sont effacés régulièrement, et chaque opération de lecture ou écriture dans le cache

marque *used* le secteur qu'elle concerne. De cette façon, seuls les secteurs les plus récemment utilisés sont marqués *used*.

Attention, les informations stockées dans le cache sont des copies de secteurs du disque dur (on les appelle souvent des pages), et pas des fichiers. Se rappeler qu'il est complètement improbable qu'un fichier corresponde exactement à un secteur.

Attention enfin : le TP où on peut voir l'effet de ces bits n'est pas le « TP Python » ! Si on a décidé de vous le faire réaliser en Python, c'est juste pour vous offrir un premier contact avec ce langage qu'il est utile de connaître parmi d'autres. Il est exceptionnel qu'un enseignement d'informatique se consacre à un langage de programmation en particulier. Les langages de programmation sont juste des moyens pour réaliser des TP en relation avec le vrai sujet des enseignements. Vous ne devez pas vous définir comme des « codeu.rs.ses », mais plutôt comme des « résolveu.rs.ses de problème » sachant coder.

Bilan : moyenne « moyenne », entre un gros paquet de 4/20 et moins (sachant que les 4 premiers points étaient donnés par les questions 1 et 2) qui plombent l'ensemble et autant de 12/20 et plus, jusqu'à plus de 20/20. Facilement améliorable en adoptant un peu de méthode pour ceux qui sont entre les deux.

- 1- Lisez bien les questions avant d'y répondre, et même toutes les questions avant de répondre à une seule. Parfois, plusieurs questions se suivent sur le même sujet ; essayez de comprendre comment elles s'articulent, et éviter de répondre en vrac en une seule fois en me chargeant de faire le tri.
- 2- Ne réinventez pas les définitions ! Apprenez-les... En science, on utilise constamment des mots du vocabulaire commun, mais dans un sens spécialisé. C'est évidemment le sens spécialisé qui compte ici. Ce module parle constamment de choses dont vous avez une expérience commune, mais son but est d'aller au-delà de l'expérience commune, donc répondre par l'expérience commune est hors-sujet.
- 3- Respecter l'orthographe et la grammaire n'est pas optionnel. Une faute de temps en temps n'est pas tragique, mais quand ça vous rend illisible ça l'est... pour vous. En plus, on observe une corrélation entre les réponses absurdes ou hors-sujet et une orthographe/grammaire déficiente.
- 4- Utiliser tout le temps disponible et toute la place disponible pour être le plus précis possible. Répondre à l'économie ne vous favorise pas. Ça ne veut pas dire remplir l'espace de hors-sujet. Il faut repérer dans la question ce qui est attendu et s'en tenir là. Ex. à une question « Qu'est-ce que ... » on ne donne pas une réponse « Pourquoi ... » ou « Comment ... », et vice-versa. Vous devez répondre uniquement à la question posée, mais le plus précisément possible. Ex. lire les *study skills* de l'université de Cornell (lien dans la page Moodle de INF2 : <http://lsc.cornell.edu/study-skills/read-about/>).

Sur le point 3 : donnez-vous des objectifs :

- a- Orthographier les mots comme dans les questions. C'est la moindre des choses. Mal orthographier un mot qui est correctement écrit une ligne au-dessus donne une très mauvaise image de vous.

- b- Orthographier les mots comme dans le cours. Par exemple, code Morse ou morse, mais pas Mors ni Morses. Si vous n'avez pas eu à écrire ce mot, ne vous inquiétez pas, c'était dans des réponses hors-sujet.
- c- Classifier et utiliser à bon escient les terminaisons qui s'entendent comme é/è ('é', 'és', 'er', 'et', 'ez', 'ais', 'ait', 'ée', 'ées', ...), faites-vous une fiche des principaux cas possibles, pour chacun des cas donnez-vous des exemples, et faites des *flashcards*.
- d- Répertorier et utiliser à bon escient les marques du pluriel, ex. sujet pluriel -> verbe pluriel, etc. Pareil pour les marques du féminin.
- e- Classifier et utiliser à bon escient 'a', 'as' et 'à'. Faites-vous une fiche et des *flashcards*.
- f- Classifier et utiliser à bon escient 'se' et 'ce'. Même méthode qu'au-dessus.
- g- Classifier et utiliser à bon escient 'sa' et 'ça'. Même méthode qu'au-dessus.
- h- Classifier et utiliser à bon escient 'son' et 'sont'. Même méthode qu'au-dessus.
- i- Classifier et utiliser à bon escient 'ses', 'ces', 's'est' et 'c'est'. Même méthode qu'au-dessus.
- j- Classifier et utiliser à bon escient 'ai', 'ait', 'et' et 'est'. Même méthode qu'au-dessus.
- k- Classifier et utiliser à bon escient 'fait' et 'faite'. Même méthode qu'au-dessus.
- l- Classifier et utiliser à bon escient 'dans' et 'd'en', 'la', 'là' et 'l'a', 'quel', 'quels', 'quelle', 'quelles', 'qu'elle' et 'qu'elles', 'quelque', 'quel que' et 'quelle que', 'non', 'n'ont' et 'nom'. Même méthode qu'au-dessus.
- m- Faire une liste de mots qui ont un cousin anglais avec une orthographe légèrement différente : ex. 'adresse' et 'address', 'littérature' et 'literature', 'accès' et 'access'.
- n- Faire une liste de mots compliqués avec leur orthographe et leur signification : ex. sophistiqué et pas sofistiqué. Facultatif ! À faire seulement si vous voulez utiliser ces mots. Si vous voulez juste les comprendre, utilisez un dictionnaire. Toujours avoir un dictionnaire pas trop loin de là où on travaille.

C'est tout, mais pour certains, ça sera déjà un progrès considérable. D'autres auront repéré que ce point 3 n'est qu'une variante du point 3 des éléments de réponse du 1^{er} CC de l'an dernier. C'est que le problème perdure, mais je peux vous assurer que votre futur environnement professionnel tient à la rigueur orthographique. En fait, il tient à la rigueur tout court, mais la rigueur orthographique en est un bon indicateur.

En circulant entre les tables lors du CC, j'ai pu voir d'excellents aide-mémoires. Excellents au sens où ils présentaient de façon très visuelle une structure complexe. Ça m'a donné envie de collecter ces aide-mémoires, mais je ne le ferai pas car je considère qu'il s'agit d'un outil d'apprentissage complètement personnel. Et surtout ne soyez pas déçu si votre aide-mémoire ne vous a pas servi à grand chose pendant l'épreuve ; c'est juste que le constituer vous a aidé à mémoriser plus de chose.

Un mot sur l'assiduité. Le diagramme qui suit montre le lien en l'assiduité (en ordonnée) et les résultats obtenus (en abscisse, rangés des meilleurs à gauche au moins bon à droite). Il y a plein de précautions à prendre (je n'ai mesuré l'assiduité qu'une fois, et j'ai fait un lissage), mais la tendance est claire.

Assiduité

