

ÉPREUVE ÉCRITE D'INFORMATIQUEPANTEL Marc

A. Remarques d'ordre général**1. Erreurs courantes**

Les correcteurs signalent que les notes attribuées aux copies rédigées en langage PASCAL restent plus faibles d'un à deux points par rapport à celles rédigées en langage CaML. Par contre, le nombre de copies en langage PASCAL (entre 10 et 15 % selon les informations transmises par les correcteurs) continue sa baisse par rapport à celles en langage CaML.

Les correcteurs ont indiqué que les candidats ne maîtrisent souvent pas la rédaction formelle de preuves. Par exemple, certaines récurrences/inductions sont remplacées par le traitement du cas d'arrêt, puis d'une étape de progression et enfin la notation « ... » pour indiquer « ainsi de suite ». Les arguments de preuve sont souvent exposés en vrac sans claire distinction entre les affirmations péremptoires et les déductions logiques. Les candidats semblent parfois avoir compris le problème et disposer des éléments de solution mais ne sont pas capables de rédiger correctement leur réponse. Cet aspect est pourtant fondamental dans le métier d'ingénieur : être capable d'argumenter scientifiquement et techniquement de la pertinence d'une solution.

Il semble aux correcteurs que les candidats hésitent parfois à faire une réponse formelle rigoureuse pour certaines questions, alors qu'ils montrent, pour d'autres questions, qu'ils en sont capables. Cela leur est en général nuisible car leur tentative informelle est parfois plus complexe et souvent fautive que la formalisation attendue.

Malgré les indications claires et sans ambiguïté, les candidats utilisent parfois les tables de vérité à la place de la résolution équationnelle avec les lois de De Morgan et réciproquement. Plus précisément, un plus grand nombre de candidats semblent mieux maîtriser le raisonnement équationnel que la construction des tables de vérité.

Les candidats ne sont pas du tout à l'aise avec les techniques d'estimation et de calcul de la complexité des algorithmes.

Les correcteurs signalent qu'un nombre non négligeable de candidats ont des difficultés dans la lecture ou l'interprétation de l'énoncé sans que ceci ne semble être lié à une difficulté spécifique dans l'énoncé lui-même.

Les correcteurs décrivent également des erreurs d'inattention dans la manipulation formelle des expressions logiques, régulières et des spécifications formelles des algorithmes.

2. Remarques sur le texte, sa compréhension

Le sujet proposé pour la session 2011 de l'épreuve optionnelle d'Informatique du Concours Commun Polytechnique est composé de trois parties indépendantes : la modélisation et la résolution de problèmes en logique des propositions ; les automates et langages ; et l'algorithmique et programmation en langage CaML ou PASCAL. Cette dernière partie se décompose en un exercice et un problème qui sont indépendants quoique manipulant la même structure de données : les arbres binaires de recherche. Cette décomposition de l'algorithmique et programmation est similaire dans l'esprit au sujet de la session 2010. Il s'agit, dans l'exercice, de faire étudier aux candidats un programme fourni en CaML ou PASCAL : le tri d'une séquence d'entiers en s'appuyant sur une structure d'arbre binaire, d'exécuter celui-ci symboliquement sur un exemple, de prouver sa correction,

sa terminaison et d'estimer sa complexité. Puis, dans le problème, de programmer plusieurs fonctions permettant de manipuler un vecteur creux représenté par un arbre binaire partiel. Ce problème repose également sur quelques preuves de propriétés de représentation des données qui guident les candidats dans la conception des algorithmes. Cette décomposition en deux sous-parties permet d'évaluer indépendamment les compétences des candidats en terme de compréhension de programmes, de preuves de propriétés (correction et terminaison) et de calcul de complexité d'une part et de programmation d'autre part. Les candidats ne traitent alors pas ces questions initiales sur leurs propres programmes mais sur celui fourni par le sujet, puis ils conçoivent ensuite leurs propres programmes. La qualité de l'évaluation en est renforcée car elle dépend moins fortement des capacités en conception des algorithmes en ce qui concerne les preuves de correction, de terminaison, les calculs de complexité et la compréhension des programmes. Cette forme de sujet se révèle plus scolaire mais distingue mieux la compréhension et la connaissance du cours d'une part et la capacité d'innovation d'autre part.

Le sujet est strictement conforme au programme (programmation exclusivement récursive, manipulation des structures de données restreinte aux listes et arbres). Il couvre la majeure partie du programme hors les circuits logiques et les langages non réguliers. Les questions proposées sont de difficultés croissantes.

Le sujet était très détaillé et n'a apparemment pas posé de problème majeur de compréhension.

3. Réactions (prévues ou non)

Alors que le sujet comporte de nombreux éléments proches de questions de cours et comporte des parties semblables dans la forme et le fond aux sujets posés les années précédentes au sein du concours, la part des candidats qui traitent celles-ci sans difficulté reste trop faible.

Les candidats utilisent régulièrement des affirmations telles : « c'est évident », « trivialement », ... comme réponses à des questions qui ne le sont pas toujours. Ces réponses sont bien sûr considérées comme erronées. Il est important que les candidats comprennent que chaque question posée mérite une réponse même si celle-ci est simple. Le but d'un concours est d'évaluer leurs capacités et de les classer les uns par rapport aux autres, il est donc nécessaire que les meilleurs candidats répondent également aux questions les plus simples même s'ils les trouvent évidentes.

Certains candidats ne semblent pas remettre en question les réponses qu'ils ont proposées en début de question ou lors de questions précédentes, alors que celles-ci les conduisent à des impasses. Un correcteur cite l'exemple d'une exécution symbolique de la fonction *trier* dont le résultat fourni par le candidat n'est pas trié.

4. Bilan

La modélisation, dans les parties logiques et automates, est en général très bien traitée. La résolution des problèmes logiques, la compréhension des programmes et la programmation des fonctions simples sont également bien traitées. Les preuves de correction et de terminaison sont plus satisfaisantes que lorsque les candidats devaient les réaliser eux-mêmes car ils ne semblaient pas comprendre le besoin de formaliser les programmes qu'ils avaient eux-mêmes rédigés. Par contre, la rédaction des preuves reste très approximative et indigne de candidats ayant une telle formation en mathématique. En ce qui concerne la programmation des fonctions plus complexes, les correcteurs ne peuvent pas déterminer si les candidats ont uniquement manqué de temps ou si celles-ci se sont révélées trop difficiles pour la plupart d'entre eux.

L'introduction de l'exercice en algorithmique et programmation a permis d'augmenter la qualité des copies par rapport aux années précédentes.

B. Rapport détaillé (Erreurs rencontrées, rédaction de certains candidats, questions sans problème)

Le pourcentage de réussite, c'est-à-dire la note moyenne obtenue par rapport au barème attribué, est indiqué pour chaque question.

Logique et calcul des propositions

Cette partie est globalement bien traitée (66 % de réussite) et un peu mieux réussie que les années précédentes. Étant donné son caractère systématique, il est encore surprenant de ne pas trouver une réussite à 100 % pour les questions I.1 à I.3 qui ne contiennent aucune innovation. Les candidats maîtrisent très bien le formalisme de la logique des propositions ainsi que les techniques de résolution syntaxique (formule de De Morgan) et sémantique (table de vérité).

- Modélisation sous la forme d'une formule du calcul des propositions (questions I.1-67 %, I.2-94 %, I.4-68 % et I.5-77 %)
Ces questions sont généralement très bien traitées. Certains candidats ont mal interprété le fait que « l'opérateur suivant doit également dire la vérité ». Certains candidats n'ont pas utilisé l'implication pour les questions I.1 et I.4 mais des formes normales disjonctives ce qui compliquait nettement les réponses.
- Résolution avec les formules de De Morgan et interprétation (question I.3-63 %)
Les candidats maîtrisent bien la résolution équationnelle.
Certains candidats utilisent des tables de vérité malgré l'indication explicite du sujet.
La résolution et l'interprétation sont correctes dans la grande majorité des cas quand la modélisation est correcte.
- Résolution avec les tables de vérité et interprétation (question I.6-47 %)
Les candidats font souvent des erreurs d'inattention dans la construction des tables de vérité.
Certains candidats utilisent la résolution équationnelle malgré l'indication explicite du sujet.
La résolution et l'interprétation sont correctes dans la grande majorité des cas quand la modélisation est correcte. Les correcteurs ont toutefois noté que la résolution équationnelle était mieux maîtrisée que la construction des tables de vérité.

Automates et langages

- Construction d'automates et d'expressions régulières (questions II.1-77 %, II.2-78 %, II.3-73 %)
Les candidats maîtrisent très bien la construction des expressions régulières décrivant le langage accepté par un automate, ainsi que la construction des automates. Ils ne simplifient pas les expressions régulières en général.
- Preuve par récurrence/induction et bilan (questions II.4-29 %, II.5-30 %, II.6-28 %)
Plus de la moitié des candidats qui traitent la question n'exploitent pas le principe de récurrence/induction mais détaillent une étape de calcul et indiquent que les autres se font de la même manière. Les autres candidats font des réponses correctes, ou ne traitent pas les questions.

Algorithmique et programmation en CaML/PASCAL

Le nombre de candidats qui utilisent les constructions impératives hors-programme, interdites par le sujet pour plus d'équité, est très faible cette année.

Question préliminaire : Tri d'une liste en exploitant un arbre binaire

Cette partie fournit un programme que les candidats doivent comprendre, puis prouver la correction et la terminaison et enfin, calculer la complexité. Ceci permet de couvrir une part importante de programme.

- Spécification formelle et calcul de propriétés (questions III.1 - 83 %, III.2 - 30 %, III.3 - 73 %, III.4 - 87 % et III.5 - 75 %)
La plupart de ces questions, très classiques dans la forme et le fond, ont été traitées par les candidats. Seule la seconde n'a pas donné lieu à la formalisation attendue dans de nombreux cas.
- Évaluation symbolique (question III.6 - 63 %)
- Preuve de correction (questions III.7 - 12 % et III.8 - 4 %)
La plupart des candidats qui abordent ces questions ne le font pas de manière formelle et tentent d'expliquer le comportement des fonctions en langue française au lieu de prouver mathématiquement leur correction.
- Preuve de terminaison (question III.9 - 31 %)
Les questions n'étaient pas difficiles mais de nombreuses réponses ne sont pas suffisamment rigoureuses.
- Calcul de complexité (question III.10 - 8 %)
Le calcul de la complexité d'une fonction semble poser des problèmes aux candidats qui traitent peu cette question et le font en général mal, en surestimant ou sous-estimant les coûts.

Problème : Codage d'un vecteur creux avec un arbre binaire

- Spécification formelle, preuve et calcul de propriétés (questions III.11 - 41 %, III.12 - 20 %, III.13 - 32 %, III.14 - 3 % et III.18 - 8 %)
Lorsque les candidats traitent ces questions, leurs réponses sont en général correctes. Par contre, de nombreux candidats n'abordent pas ces questions.
- Programmes (questions III.15 - 57 %, III.16 - 10 %, III.17 - 8 %, III.19 - 3 %, III.20 - 1 %, III.21 - 1 % et III.22 - 1 %)

C. Conclusion

La moyenne et l'écart type de l'épreuve sont respectivement de 10,11 et 2,64. Ces valeurs sont compatibles avec l'avis des correcteurs. Il est raisonnable que la moyenne soit proche de 10 ce qui permet d'obtenir le plus grand étalement et facilite le classement. Par contre, l'écart type reste trop faible et le sujet n'est donc pas suffisamment discriminant. En effet, celui-ci contient des questions simples qui permettent de valider la connaissance et la compréhension du cours lui-même sans évaluer les performances et la créativité des candidats (questions I.1 à I.6, II.1 à II.3, III.1 à III.6). La majorité des candidats semblent consacrer un temps trop long à ces questions-là et ne peuvent pas ensuite traiter suffisamment de questions permettant de les départager.