

Programme de colle - Semaine 13 (8 décembre)

La démonstration des énoncés marqués d'une étoile est exigible

1 Langages réguliers

- Mots : alphabet, mot sur un alphabet, concaténation, puissance d'un mot, structure de monoïde.
- Langages : opérations ensemblistes sur les langages (union, intersection, différence, complémentaire), concaténation de deux langages, étoile de Kleene d'un langage
- Langages réguliers (aussi appelés langages rationnels): définition inductive de la classe des langages réguliers, **tout langage fini est régulier (*)**
- Expressions régulières : définition inductive, langage dénoté par une expression régulière (notation $\mu(e)$), **un langage est régulier si et seulement si il est dénoté par une expression régulière (* interroger sur un des deux sens seulement)**, expressions régulières équivalentes
- Expressions régulières étendues : aucune connaissance spécifique n'est exigible mais les étudiants doivent savoir trouver une expression régulière équivalente pour chacune des opérations introduites.

2 Automates finis

Les étudiants doivent savoir appliquer les constructions et algorithmes sur les automates.

- Automate fini déterministe (afd) : définition et calcul d'un afd, mots reconnus, langages acceptés
- Accessibilité et co-accessibilité; afd émondé
- Afd complet; complétion d'un afd
- Automate complémentaire
- Automate produit
- Classe des langages reconnaissables : stabilité par complémentaire, intersection et union finie
- Automate fini non déterministe (afnd) : définition et calcul d'un afnd, mots reconnus, langage accepté
- Automate fini non déterministe avec transitions instantanées (ε -afnd) : définition, ε -clôture, calcul d'un ε -afnd, mots reconnus, langage accepté
- Déterminisation d'un afnd ou d'un ε -afnd (automate des parties) : les classes de reconnaissabilité sont les mêmes pour tous les types d'automates finis.
- Langages non reconnaissables : **lemme de l'étoile (*)**, les exemples $\{a^n b^n \mid n \in \mathbb{N}\}$ et $\{u \mid |u|_a = |u|_b\}$ ont été vus en cours.

3 Théorème de Kleene

- Théorème de Kleene : la classe des langages réguliers est la même que la classe des langages reconnaissables.
- Passage d'une expression régulière à un automate :
 - Construction d'automates normalisés à la Thomson : les cas de base et les cas inductifs (union, concaténation, étoile) doivent être parfaitement connus, aucune formalisation précise de l'algorithme général n'est exigible mais les étudiants doivent savoir appliquer la méthode pour construire un automate.
 - Langage local : définition, savoir déterminer si un langage est local (ou non), automate de Glushkov reconnaissant un langage local.
 - Algorithme de Berry-Séthi : connaître les étapes de l'algorithme et savoir l'appliquer.
- Passage d'un automate à une expression régulière :
 - Automate généralisé avec des expressions régulières sur les transitions ne dénotant pas le langage vide.
 - Algorithme par élimination d'états sur un automate normalisé.
- Rappels sur les conséquences du théorème de Kleene sur les stabilités des langages réguliers/reconnaissables.

4 Apprentissage supervisé

- Principe de l'apprentissage supervisé
- Arbres binaires de décision : définition et écriture de programmes simples sur les arbres de décision.
- Éléments de théorie de l'information : définition de l'entropie, cas du pile ou face traité en cours, **étude de la fonction $H(X)$ pour X suivant une loi de Bernoulli de paramètre p** (*) calcul de l'entropie dans le cas d'un ensemble d'exemples d'apprentissage, calcul du gain d'information pour un attribut binaire uniquement.
- Algorithme ID3 : uniquement dans le cas d'apprentissage d'arbres binaires de décision.
- Algorithme du plus proche voisin. Algorithme des k plus proches voisins. Il faut savoir calculer les k plus proches voisins d'un ensemble de n points en $O(nk)$ c'est-à-dire sans trier la liste des points.
- Arbres k -dimensionnels : définition et exemples. Représentation du partitionnement de l'espace obtenu à l'aide d'un k -arbre dimensionnel. Insertion dans un arbre k -dimensionnel. Algorithme de recherche du plus proche voisin à l'aide d'un arbre k -dimensionnel.
- Révisions : le cas de la dimension 1 est correspond à celui de l'arbre binaire de recherche vu en MP2I. On pourra poser toute question de révision sur ce point (recherche, insertion, suppression). Pas de question sur les arbre rouge noir.

5 Complexité et décidabilité

- Notion de problème de décision.
- Problèmes décidables. La notion de machine de Turing est hors programme, une machine est un algorithme ou un programme écrit en C ou en OCaml s'exécutant sur une machine à mémoire infinie.
- Classe **P**.
- Principe de réduction polynomiale d'un problème A à un problème B noté $A \leq_P B$.
- Si $B \in \mathbf{P}$ et $A \leq_P B$ alors $A \in \mathbf{P}$ (*)
- Classe **NP** : définie comme la classe des problèmes qui peuvent être vérifiés en temps polynomial. Notion de certificat pour un problème de décision. **SUBSETSUM est dans NP** (*).