Colle d'informatique 1

Programme

Langages formels

- Alphabet, mot, préfixe, suffixe, facteur, sous-mot, concaténation
- Langage sur un alphabet, union, intersection, concaténation, exponentiation, étoilé

Langages réguliers

- Définition inductive
- Expression réguliere, définition inductive
- Langage dénoté par une expression régulière
- Expressions équivalentes

Automates finis

- DFA complet (automate fini déterministe complet) $\mathcal{A} = (Q, \Sigma, q_0, F, \delta)$
- Mot accepté/rejeté, chemin
- Langage d'un automate $\mathcal{L}(\mathcal{A})$
- Déterministe et complet, DFA incomplet, complétion, état puits
- États accessibles, co-accessibles, utiles, émondage
- NFA (automate fini non déterministe), intérêt et conséquences du non-déterminisme, chemin acceptant
- ε-NFA, chemin acceptant
- Équivalence DFA / NFA / ε-NFA, déterminisation (construction par sous-ensembles), ε-fermeture d'un état, ε-NFA vers DFA, ε-NFA vers NFA, élimination des ε-transitions par fermeture avant/arrière
- Langage reconnaissable L sur Σ : $L \subseteq \Sigma^*$ est reconnaissable ssi il existe un automate $\mathcal{A}$ tel que $L = \mathcal{L}(\mathcal{A})$.

Théorème de Kleene

Sur un alphabet Σ, on note $\mathrm{Rec}(\Sigma)$ l'ensemble des langages reconnaissables et $\mathrm{Reg}(\Sigma)$ celui des langages réguliers. Le *théorème de Kleene* établit que $\mathrm{Rec}(\Sigma) = \mathrm{Reg}(\Sigma)$. La démonstration s'appuie sur une double inclusion. La construction de l'automate de Glushkov, à l'aide de l'algorithme de Berry-Sethi, associe un automate à toute expression régulière. Par éliminition des états, on transforme un automate en un automate généralisé pour obtenir une expression régulière qui dénote le langage reconnu par ce dernier.

- Langage local, ensembles P, S, L et N, automate local
- Linéarisation d'une expression régulière
- Automate de Glushkov, algorithme de Berry-Sethi
- Algorithme de Thompson, construction d'un ε-NFA à partir d'une expression régulière
- Algorithme d'élimination des états d'un automate, automate généralisé
- Lemme d'Arden

Le lemme d'Arden est hors-programme mais il a été vu en cours et en exercices. Il est au programme de colle.

Stabilité des langages

- Rappel sur les langages réguliers, stabilité par les opérations régulières
- Stabilité par complémentation, une preuve à connaître
- Stabilité par intersection, une preuve à connaître
- Stabilité par miroir, preuve à connaître

Lemme de l'étoile

- Idée intuitive du lemme, intérêt
- Formulation du lemme, démonstration
- Exemple classique : $\{a^n b^n \mid n \in \mathbb{N}\}$ non régulier doit être connu et démontré

Commentaires

- Le cours doit être parfaitement connu. Une question de cours peut être posée directement ou intégrée dans un exercice. La non connaissance du cours implique une note inférieure à $10/20$.
- On s'assurera de la maîtrise des concepts, de l'exactitude du vocabulaire employé, d'un savoir-faire algorithmique. En fonction de l'attitude de l'étudiant, ne pas hésiter à prolonger le sujet pour approfondir un sujet.
- Outre le fond, la forme est également évaluée : présentation du tableau, aisance orale, réaction face aux questions, argumentation, justification, prise d'initiative.
- Ce programme de colle se prête à des exercices de type théorique. Dans la mesure du possible, des questions de code peuvent être posées pour manipuler des structures de données adaptées et mettre en œuvre certains algorithmes. Toutefois, la durée limitée de l'oral ne permettra que d'effleurer cet aspect du problème, abordé en TP et pourra l'être lors de la préparation aux oraux. OCaml a été plus largement utilisé en ce début d'année.

Extraits du programme officiel

Langages réguliers

On introduit les expressions régulières comme formalisme dénotationnel pour spécifier un motif dans le cadre d'une recherche textuelle.

Notions	Commentaires
Alphabet, mot, préfixe, suffixe, facteur, sous-mot.	Le mot vide est noté ε.
Langage comme ensemble de mots sur un alphabet. Opérations régulières sur les langages (union, concaténation, étoile de Kleene). Définition inductive des langages réguliers.	
Expression régulière. Dénotation d'un langage régulier.	On introduit les expressions régulières comme un formalisme dénotationnel pour les motifs. On note l'expression dénotant le langage vide $\varnothing$, celle dénotant le langage réduit au mot vide ε, l'union par $\vert$, la concaténation par juxtaposition et l'étoile de Kleene par une étoile.
Expressions régulières étendues.	Le lien est fait avec les expressions régulières de la norme POSIX, mais on ne développe aucune théorie supplémentaire à leur sujet et aucune connaissance au sujet de cette norme n'est exigible.

Automates finis

Les automates constituent un modèle de calcul puissant qui irrigue de nombreuses branches de l'informatique. On voit ici les automates comme un formalisme opérationnel efficace pour la recherche de motifs. On vérifie que le formalisme des automates coïncide exactement avec l'expressivité des expressions régulières.

Notions	Commentaires
Automate fini déterministe. État accessible, co-accessible. Automate émondé. Langage reconnu par un automate.	On insiste sur la richesse de systèmes dont le fonctionnement peut être modélisé par un automate.
Transition spontanée (ou ε-transition). Automate fini non déterministe.	
Déterminisation d'un automate non déterministe.	On fait le lien entre l'élimination des transitions spontanées et l'accessibilité dans un graphe. On aborde l'élimination des transitions spontanées et plus généralement les constructions d'automates à la Thompson sur des exemples, sans chercher à formaliser complètement les algorithmes sous-jacents.
Construction de l'automate de Glushkov associé à une expression régulière par l'algorithme de Berry-Sethi.	Les notions de langage local et d'expression régulière linéaire sont introduites dans cette seule perspective.
Passage d'un automate à une expression régulière. Élimination des états. Théorème de Kleene.	On se limite à la description du procédé d'élimination et à sa mise en œuvre sur des exemples d'automates de petite taille ; cela constitue la preuve du sens réciproque du théorème de Kleene.
Stabilité de la classe des langages reconnaissables par union finie, intersection finie, complémentaire.	
Lemme de l'étoile.	Soit L le langage reconnu par un automate à n états : pour tout $u \in L$ tel que $\vert u\vert \geq n$, il existe x, y, z tels que $u = xyz$, $\vert xy\vert \leq n$, $y \neq \varepsilon$ et $xy^*z \subseteq L$.