Correction du TP3, questions sur les langages

Paul Melotti (paul.melotti@ens.fr)

Avertissement : Les raisonnements sur les mots peuvent sembler arides et difficile à comprendre sans illustration ! Faites des dessins, des "blocs" pour visualiser les mots, les suffixes, etc.

Q.3.2.1 On numérote les états de notre automate par les entiers de $[0;|m|]$. Ainsi, l'état i correspond au préfixe de m de taille i.

Pour un mot t, on note $l(t)$ la longueur du plus grand mot qui est à la fois préfixe de m et suffixe de t :

$$l(t) = |\max(\mathrm{Suff}(t) \cap \mathrm{Pref}(m))|.$$

Alors on montre par récurrence sur $|t|$ la propriété : "Après avoir lu le mot t on se trouve dans l'état $l(t)$".

Pour $|t| = 0$, c'est clair.

Si la propriété est vraie pour les mots t de taille n, alors soit u un mot de taille $n+1$. On écrit $u = ta$ où a est une lettre. Par hypothèse de récurrence, après avoir lu t, on est à l'état $l(t)$, qui correspond à un mot p (p est le plus long préfixe de m qui soit aussi suffixe de t). Montrons qu'en lisant ta, on arrive à l'état $l(ta)$.

- Si pa est préfixe de m, l'automate nous dit d'aller en pa. Montrons qu'on a bien $|pa| = l(ta)$: ta est bien préfixe de m par hypothèse, et suffixe de ta puisque p était suffixe de t. Reste à montrer que c'est le plus grand, autrement dit, que tout mot qui a vérifie ces deux propriétés est plus petit que pa.
  si w est à la fois préfixe de m et suffixe de ta, soit c'est le mot vide et on a gagné, soit w n'est pas vide et donc finit par la lettre a. Il s'écrit alors $w = za$. Le mot z est alors préfixe de m et suffixe de t, il est donc plus court que p par hypothèse de récurrence. Donc $l(ta) = |pa|$.
- Si pa n'est pas préfixe de m, l'automate nous dit d'aller en $B(pa)$. Ce mot $B(pa)$ est un suffixe de pa par définition, donc c'est un suffixe de ta. C'est aussi un préfixe de pa différent de pa, donc c'est un préfixe (au sens large) de p, et comme p était préfixe de m, le mot $B(pa)$ est aussi préfixe de m. Reste à montre que c'est le plus grand :
  si w est à la fois préfixe de m et suffixe de ta, soit c'est le mot vide et on a gagné, soit il n'est pas vide et alors termine par a. Il s'écrit alors $w = za$ où z est préfixe de m et suffixe de t, donc $|z| \leq |p|$. De plus, pa n'est pas préfixe de m donc m commence par p suivi d'une lettre b différente de a. On en déduit que $|z| < |p|$ (sinon on aurait $z = p$ donc $w = pa$ qui serait préfixe de m, or m commence par pb). Ainsi, $w = pa$ est un préfixe et suffixe de pa différent de pa. Il est donc plus petit que $B(pa)$.

Finalement, un mot t est reconnu ssi $l(t) = |m|$,
ssi m est le plus grand préfixe de m qui soit suffixe de t,
ssi m est suffixe de t.

Donc l'automate reconnaît A^*m.

Q.3.2.2 Vu en TD.

Q.5.1 L'automate construit a $|m|+1$ états. C'est optimal, ce qui signifie qu'on a construit l'automate déterministe complet minimal. Pour prouver la minimalité, on peut utiliser les résiduels, ou plus simplement raisonner de la façon suivante.

Supposons par l'absurde qu'il existe un automate déterministe complet reconnaissant A^*m et ayant moins de $|m|$ états. Faisons passer le mot m dans cet automate. Comme le mot m a $|m|+1$ préfixes, il existe deux préfixes différents pour lesquels on se retrouve dans le même état, qu'on appelle i. Disons que ces deux préfixes sont p et q, et que $|p| < |q|$. Écrivons $m = qv$ où v est un mot éventuellement vide. Depuis l'état i, si je lis v, je dois arriver dans un état final puisque $m = qv$. Mais lorsque je lis p j'arrive aussi en i, donc le mot pv doit aussi être reconnu par l'automate. Or $|pv| < |qv| = m$ donc le mot pv n'a pas m comme suffixe, et n'a aucune raison d'être reconnu.