Langages et expressions régulières

Florian Bourse

Expressions régulières

Question 1. La commande `grep -E 'regexp' fichier` retourne les lignes d'un fichier où l'on reconnaît l'expression régulière `regexp`. Une expression régulière est formée à partir de lettres de l'alphabet (chaque lettre constituant un élément) et des symboles :

`*` qui indique que l'élément précédent apparaît un nombre quelconque de fois

`+` qui indique que l'élément précédent apparaît au moins une fois

`(` et `)` qui entourent une expression régulière pour former un élément

`|` qui indique une disjonction entre deux expressions régulières (on reconnaît l'une ou l'autre)

`?` qui indique que l'élément précédent est facultatif (i.e., il apparaît au plus une fois)

Une expression régulière représente donc sous forme condensée un ensemble de mots (c'est-à-dire un langage). Soit `tutu` le fichier contenant :

tutu
```
caa
cbbab
cabab
caaabba
```

1. Que répond la commande `grep -E 'regexp' tutu` – et pourquoi ? – lorsque `regexp` vaut respectivement :

   `ca+`; `c(ab)+`; `ca+(a*b*)`; `(aa+)|(bb+)`; `baa?b`; `a*`.

2. Décrire sous forme ensembliste le langage correspondant aux expressions régulières précédentes.

Question 2. Déterminer les mots de longueur maximale 4 qui appartiennent au langage dénoté par chacune des expressions régulières suivantes :

1. `(b|ba)*`
2. `ab*|b`
3. `(a|b)*abb`
4. `(x|`ε`)*dd*`
5. `(xd|`ε`)*d*`
6. `a*(b|c)d*`

Question 3. Donner une description en français des langages donnés par les expressions régulières suivantes :

1. `(a|b)*`
2. `a(a|b)*`
3. `(a|b)*a`
4. `(b|ab)*(a|ε)`
5. `a*|b*`
6. `(aa|b)*`
7. `(ab*a|b)*`
8. `(a|b)(a|b)`
9. `(εa|b)(εa|b)`
10. `((a|b)(a|b))*`
11. `(a|b)*a(a|b)*`
12. `(a|b)*ab(a|b)*`
13. `(a|b)*a(a|b)*b(a|b)*`
14. `(ab)*`

Question 4. Prouver les équivalences suivantes :

1. `(a|b)|(a|b)(a|b)*|∅*≡(a|b)*`
2. `a(c+|)∅)|(a|b)(c*|(c*)*)≡(a|b)c*`
3. `(x|y)∅|(x|y)∅*+((x|y)*)∅*)*≡(x|y)*`
4. `0(ε|00)*(1|01)|1≡0*1`

Question 5. Pour chacun des langages suivants, donner une expression régulière représentant son complément :

1. `(a|b)*b`
2. `((a|b)(a|b))*`

Langages

Question 6. Déterminer les facteurs, les préfixes et les suffixes du mot $u = abac$.

Question 7.

1. Compter les occurrences des lettres a et b dans les mots suivants : a^3cbbca, $aabgjdd$, $titi$, $babc$.
2. Donner l'ensemble des couples (u, v) tels que $uv = abaac$.
3. Calculer LM pour les ensembles suivants :
   — $L = \{a, ab, bb\}$ et $M = \{\varepsilon, b, a^2\}$;
   — $L = \emptyset$ et $M = \{a, ba, bb\}$;
   — $L = \{\varepsilon\}$ et $M = \{a, ba, bb\}$;
   — $L = \{aa, ab, ba\}$ et $M = \{a, b\}^*$.

Question 8. Prouver les assertions suivantes, où V est un alphabet, $a, b \in V$ et $u, v, x, y \in V^*$.

— $au = bv \Rightarrow a = b \wedge u = v$;

— $xu = xv \Rightarrow u = v$;

— $(xu = yv \wedge |x| = |y|) \Rightarrow u = v$;

— $(xu = yv \wedge |x| \leq |y|) \Rightarrow$ (x est préfixe de y et v est suffixe de u).

Question 9. Soient u et v deux mots. Montrer que $uv = vu$ si et seulement si il existe $\gamma \in V^*$ et $p, q \in \mathbb{N}$ tels que $u = \gamma^p$ et $v = \gamma^q$.

Question 10. Montrer que :

1. Il n'existe pas de mot $x \in \{a, b\}^*$ tel que $ax = xb$.
2. Il n'existe pas de mots $x, y \in \{a, b\}^*$ tels que $xay = ybx$.

Question 11. Soit A, B, C trois langages sur un même alphabet. Prouver les propriétés suivantes :

1. $(AB)C = A(BC)$;
2. $(A^*)^* = A^*$;
3. $A \subset B \Rightarrow A^* \subset B^*$;
4. $A(B \cup C) = AB \cup AC$;
5. $A(B \cap C) \subset AB \cap AC$, et l'inclusion réciproque est fausse en général ;
6. $(A \cup B)^* = (A^*B^*)^*$.

Question 12. Soient A, B deux langages sur un même alphabet.

1. Comparer $(A \cup B)^*$ et $A^* \cup B^*$.
2. Comparer $(A \cap B)^*$ et $A^* \cap B^*$.
3. Comparer $(AB)^*$ et A^*B^*.

Question 13. Deux mots u et v sont dits *conjugués* s'il existe deux mots w_1 et w_2 tels que $u = w_1w_2$ et $v = w_2w_1$. En d'autres termes, v s'obtient à partir de u par permutation cyclique de ses lettres.

1. Montrer que la conjugaison est une relation d'équivalence. (réflexive, symétrique, transitive)
2. Montrer que u et v sont conjugués si et seulement s'il existe un mot w tel que $uw = wv$.

Question 14. On appelle *code* sur un alphabet V tout langage X sur V tel que pour tous $x_1, \ldots, x_p \in X$ et pour tous $y_1, \ldots, y_q \in X$: $x_1 \ldots x_p = y_1 \ldots y_q \Rightarrow (p = q et \forall i \leq p : x_i = y_i)$. Autrement dit, X est un code ssi tout élément de X^* se factorise de manière unique sur X.

1. Les langages suivants sont-ils des codes ?
   — $X_1 = \{ab, baa, abba, aabaa\}$;
   — $X_2 = \{b, ab, baa, abba, aaaa\}$;
   — $X_3 = \{aa, ab, aab, bba\}$;
   — $X_4 = \{a, ba, bba, baab\}$.
2. Soit $u \in V^*$. Montrer que le singleton $\{u\}$ est un code si et seulement si $u \neq \varepsilon$.
3. Soit u et v deux mots distincts sur V. Montrer que la paire $\{u, v\}$ est un code si et seulement si u et v ne commutent pas.
4. Soit X une partie de V ne contenant pas ε et telle qu'aucun mot de X n'est préfixe propre d'un autre mot de X. Montrer qu'alors X est un code. (Un tel code est appelé *code préfixe*.)

Langages réguliers

Question 15. On rappelle que deux ensembles X et Y sont équipotents s'il existe une bijection de l'un vers l'autre. On note alors $X \sim Y$. En particulier, un ensemble équipotent à $\mathbb{N}$ est dit dénombrable.

1. Prouver qu'un ensemble n'est jamais équipotent à l'ensemble de ses parties.
2. Soit Σ un alphabet. Montrer que l'ensemble des *mots* sur Σ est dénombrable alors que l'ensemble des *langages* sur Σ ne l'est pas.
3. En déduire qu'il existe des langages non réguliers.

Question 16. Pour chacune de ces affirmations, dire si elle est vraie ou fausse en argumentant brièvement.

1. Tout langage régulier est infini.
2. Tout langage non régulier est infini.
3. Il y a une infinité des langages réguliers.
4. Il y a une infinité de langages non réguliers.
5. Tout langage inclus dans un langage régulier est régulier.
6. Il y a toujours une infinité d'expressions régulières pour décrire un langage régulier.