
Correction des exercices de TD

Table des matières

1	Analyse Lexicale	2
1.1	TD/TP 1	2
1.2	TD/TP 2	6
1.3	TD/TP 3	7
1.4	TD/TP 4	9
2	Analyse descendante récursive	16
2.1	TD/TP 5	16
2.2	TD/TP 6	17
2.3	TD/TP 7	18
3	Analyse ascendante LR	22
3.1	TD/TP 8	22
3.2	TD/TP 9	22
3.3	TD/TP 10	23

1 Analyse Lexicale

1.1 TD/TP 1

Exercice 1 (théorie des langages)

1. Comment peut-on caractériser un langage rationnel (régulier) ?
2. Les langages de programmation (C, Java, Python) sont-ils réguliers ? Pourquoi ?
3. Comment peut-on caractériser un langage algébrique ?
4. Soit l'extrait suivant de grammaire EBNF de requête de consultation SQL :

```
SELECT
  [ALL | DISTINCT | DISTINCTROW ]
  select_expr [, select_expr] ...
  [into_option]
  [FROM table_references
    [PARTITION partition_list]]
  [WHERE where_condition]
  [GROUP BY {col_name | expr | position}
    [ASC | DESC], ... [WITH ROLLUP]]
  [HAVING where_condition]
  [ORDER BY {col_name | expr | position}
    [ASC | DESC], ...]
  [LIMIT {[offset,] row_count | row_count OFFSET offset}]
;
```

Expliquer les différentes conventions de notation utilisées !

5. Les requêtes suivantes sont-elles correctes ?

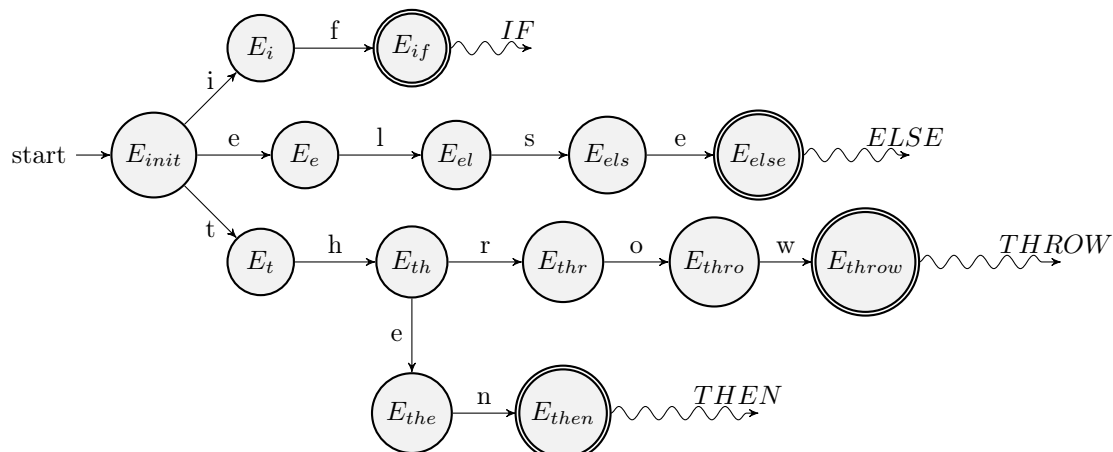
```
select 5 ;
SELECT distinct nom from etudiant
select ref from articles order by rayon asc, ref asc;
```

1. Tout langage reconnu par un automate d'état fini déterministe est un langage rationnel (régulier).
2. Les langages de programmation ne sont pas des langages réguliers car ils contiennent une structure emboîtée non bornée (aucun automate d'état fini déterministe ne peut être construit pour le langage $a^n b^n$).
3. Langages a association de symboles (structure emboîtée non bornée), reconnus par un automate à pile avec un arbre de dérivation constructible par induction.
4. EBNF = Extended Backus-Naur Form. Les `[]` sont utilisés pour désigner une option facultative, `|` pour le « ou », `...` pour l'opération de répétition de n caractères ($n \geq 0$) et `{}` pour indiquer un choix non optionnel.
5. Seule la seconde requête n'est pas correcte (il manque le `;`).

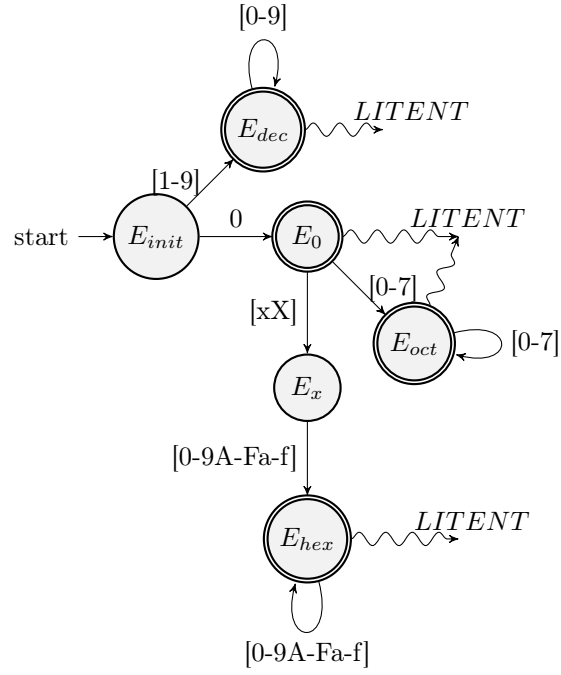
Exercice 2 (automate)

- Dessiner un Automate à états Finis Déterministe (AFD) distinct pour chacun des langages suivants :
 1. Langage de certains mots-clés du C : $L_{key} = \{if, then, else, throw\}$ (sensible à la casse).
 2. Langage des littéraux numériques entiers du C (ou C++, ou Java), décimaux L_{c10} , octaux L_{c8} , hexadécimaux L_{c16} .
 3. Langage L_{id} des identificateurs composés d'une lettre au moins, éventuellement suivie de chiffres, de lettres et de "_".
 4. Langage des littéraux numériques flottants décimaux L_f ; la suite de chiffres à gauche ou bien à droite du point décimal pouvant être vide ; l'exposant entier n'est pas obligatoire. Exemples : 13., 1.7e23, .89E-34.
 5. Langage L_{sep} des séparateurs composés de blancs (espace, \t, \n), des commentaires à la C et à la C++.
- Dessiner un unique AFD à jeton reconnaissant une partie de ces langages. Vous reconnaîtrez notamment : le mot-clé if, les identificateurs, les entiers décimaux, les flottants sans exposant, les séparateurs. Utiliser des jetons négatifs pour les lexèmes à filtrer (séparateurs).

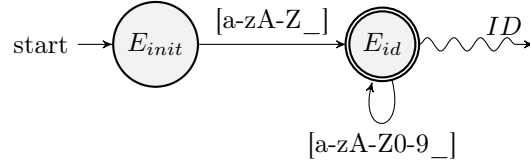
- 1. Automate fini déterministe L_{key} :



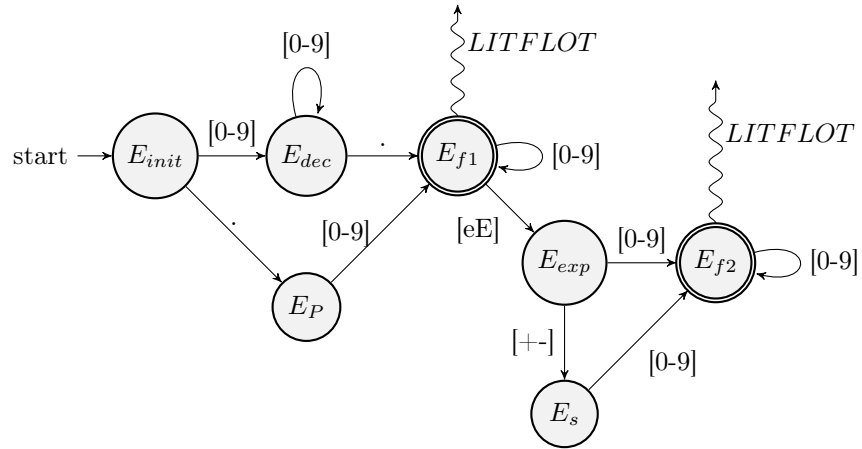
2. Automate fini déterministe de L_{c10} , L_{c8} et L_{c16} :



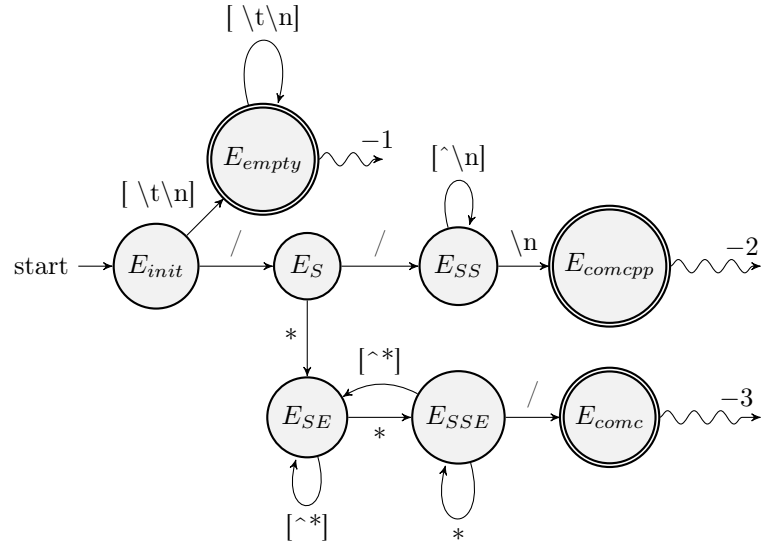
3. Automate fini déterministe de L_{id} :



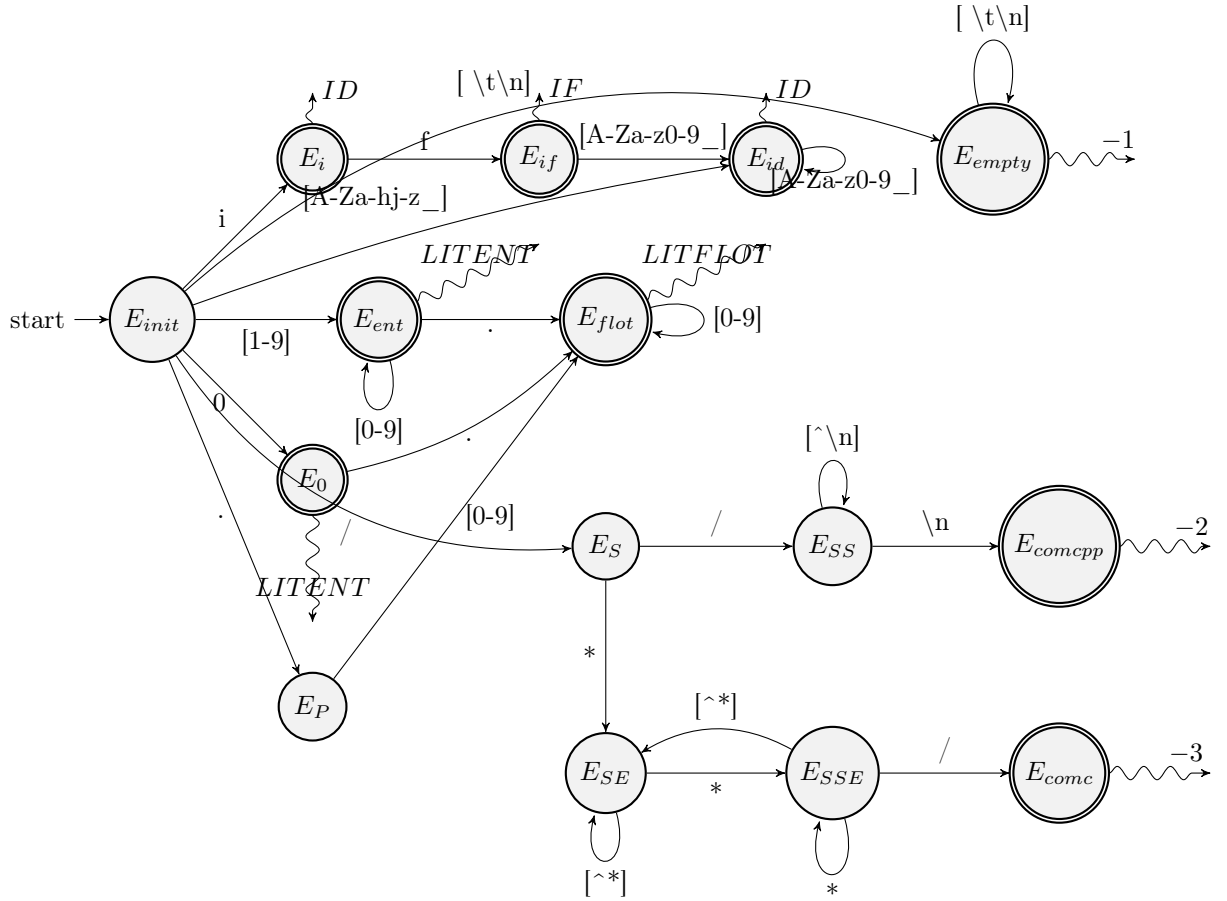
4. Automate fini déterministe de L_f :



5. Automate fini déterministe de L_{sep} :



— AFD de L_{if} , L_{id} , L_{c10} , L_f et L_{sep} :



Exercice 3 (AFD en C)

On utilisera dans ce TP, la fonction `analex` définie dans le cours et qui doit être récupérée sur l'ENT. Trois fichiers sont à télécharger :

- `afd.h` qui contient la définition d'un automate ;
- `analex.h` qui contient la définition de la fonction `analex()` ;
- `analex.c` qui contient un `main` appelant la fonction `analex()` itérativement ;

L'objectif de ce TP est d'implémenter l'unique AFD à jeton réalisé dans l'exercice précédent de façon à reconnaître une partie des catégories lexicales du langage C. Pour cela, on modifiera la fonction `creerAfd` du fichier `afd.h`.

1. Lisez et testez les 3 fichiers téléchargés afin d'en comprendre le fonctionnement ;
2. On aura besoin de créer plusieurs transitions d'un état à un autre étiqueté par une classe de caractères comme les minuscules ou les chiffres. Écrire le corps de la fonction suivante :

```
/** construit un ensemble de transitions de ed à ef pour un intervalle de char  
* @param ed l'état de départ  
* @param ef l'état final  
* @param cd char de début  
* @param cf char de fin  
**/  
void classe(int ed, int cd, int cf, int ef);
```

3. En utilisant cette fonction `classe()`, modifiez la fonction `creerAfd()` dans le fichier `afd.h` ;
4. Dans `analex.c`, modifiez le `main()` afin qu'il n'affiche plus d'invite itérativement `analex()` et affichera une chaîne correspondante à l'entier retourné ainsi que le lexème, ceci jusqu'à la fin du fichier.

► Voir TP1.

1.2 TD/TP 2

Exercice 4 (expressions régulières)

Écrire les expressions régulières correspondant aux langages réguliers suivants : L_{key} , L_{c10} , L_{c8} , L_{c16} , L_{id} , L_f , L_{sep} .

```
 $L_{key}$   if|else|th(en|row)  
 $L_{c10}$  ([1-9][0-9]*|0)  
 $L_{c8}$   0[0-7]*  
 $L_{c16}$  0[xX][0-9A-Fa-f]*  
 $L_{id}$   [a-zA-Z_][a-zA-Z0-9_]*  
 $L_f$    ([0-9]+.[0-9]*|. [0-9]+)([eEdD](\+-)?[0-9]+)?  
 $L_{sep}$  ([ \t\n\r\f] | ("/".*\n) | ("/*"([~*]*\n+[*~])*)[*~]* \n+\/))
```

Exercice 5 (flex)

Écrire un analyseur lexical reconnaissant l'ensemble des expressions régulières des exercices précédents à l'aide de flex. L'action associée à chaque lexème reconnu consiste à retourner le jeton correspondant au lexème. Toute autre expression d'un caractère retournera un jeton correspondant au code ASCII de ce caractère.

1. Écrivez l'analyseur lexical `analflex.1`
2. Transformez-le en exécutable
3. Testez-le à la ligne de commande puis sur un programme C redirigé. Continuez à améliorer ce programme jusqu'à ce que tous les lexèmes soient reconnus.

Testez cet analyseur

► Voir TP2.

Exercice 6 (supplément)

Améliorer l'analyseur lexical précédent à l'aide de flex en affectant à une variable globale `yyval` une valeur sémantique dépendant de la catégorie du lexème reconnu :

mots-clés rien ;
LITENT valeur entière longue (long int) ;
ID valeur chaîne de caractères ;
LITFLOT flottant double précision ;

► Voir TP2.

1.3 TD/TP 3

Exercice 7 (flex nettoyage de texte)

Écrire un source flex `delblancs.1` filtrant un fichier en :

- supprimant les lignes blanches (lignes vides ou remplies de blancs (espace et tabul.)),
- supprimant les débuts et fins de ligne blancs,
- remplaçant tous les blancs `\t<espace>` multiples par un seul espace,
- remplaçant les tabulations `\t` par un espace.

► Voir TP3.

Exercice 8 (wc en flex)

Écrire en flex un programme comptant le nombre de lignes, le nombre de mots et le nombre de caractères d'un fichier passé en argument (**man wc**). Un mot est une suite de caractères séparés par des blancs (tab, espaces, retour ligne). Vérifiez que vos résultats sont les mêmes que ceux de **wc**.

► Voir TP3.

Exercice 9 (convertisseur Markdown (.md) to HTML)

Markdown est un langage de balisage léger destiné à offrir une syntaxe facile à lire et à écrire utilisée par GitHub et GitLab (readme.md). Un document balisé par Markdown peut être converti facilement en HTML. Pour réaliser des listes non numérotées (``) :

- une ligne vide précède et suit la liste de 1er niveau
- chaque item est précédé d'une étoile (*) ou d'un tiret (-) en début de ligne
- on peut emboîter des listes en indentant avec au moins 2 espaces de plus que le niveau parent.

L'exemple suivant :

```
* 1
  * 1, 1
    - 1, 1, 1
      * 1, 1, 1, 1
      * 1, 1, 1, 2
    - 1, 2
* 2
```

produira :

```
<ul><li>1</li>
<ul><li>1, 1</li>
<ul><li>1, 1, 1</li>
<ul><li>1, 1, 1, 1</li>
<li>1, 1, 1, 2</li>
</ul></ul><li>1, 2</li>
</ul><li>2</li>
</ul>
```

Écrire le source flex permettant d'effectuer la traduction.

► Voir TP3.

1.4 TD/TP 4

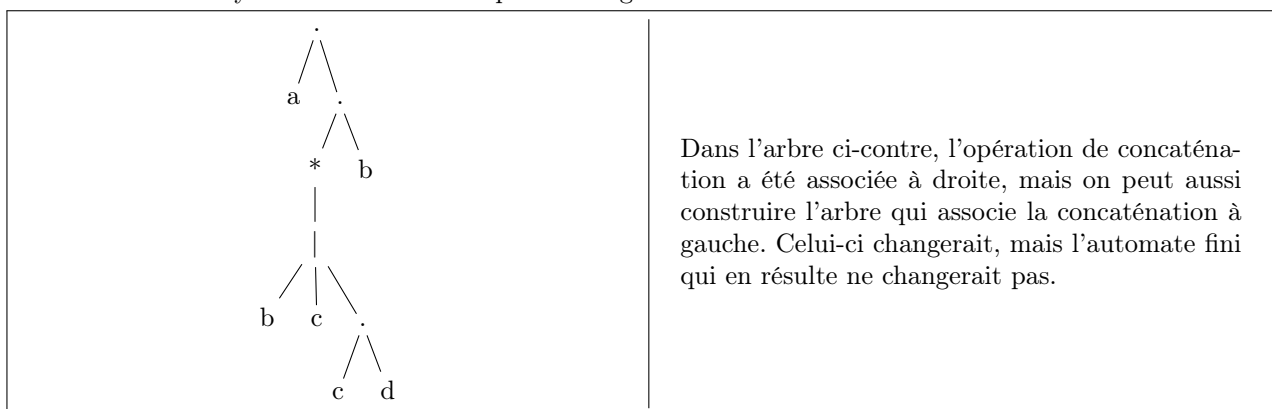
Exercice 10 (algorithmique des automates d'états finis)

Soit l'expression régulière e suivante : $e = a((b|c)^*|cd)^*b$

1. Dessiner un automate fini équivalent à e .
2. Cet automate est-il déterministe ? Si oui, indiquez pour quelle raisons, sinon, déterminez-le.
3. Minimisez cet AFD.

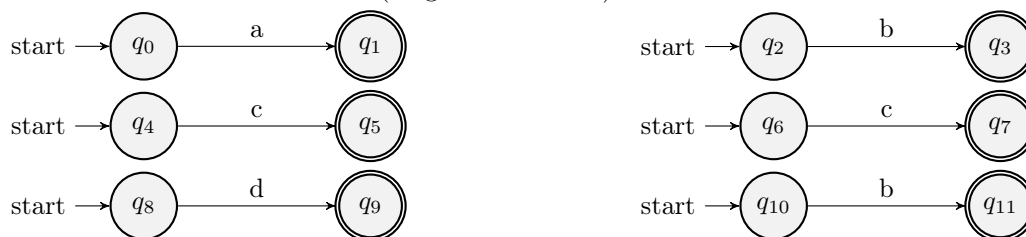
1. Pour commencer, simplifions l'expression régulière e . On sait que $r^* = r^n$ donc $r^{**} = r^{n^m} = r^{nm} = r^*$. Ainsi, on a $(b|c)^{**} = (b|c)^*$. De plus, $(r * |s)^* = r^{**}|s^* = r^*|s^* = (r|s)^*$ donc $((b|c)^*|cd)^* = (b|c|cd)^*$. On peut réécrire e en $a(b|c|cd)^*b$.

Appliquons la méthode de Thompson pour construire l'automate fini équivalent à e . On construit tout d'abord l'arbre de syntaxe abstrait de l'expression régulière :

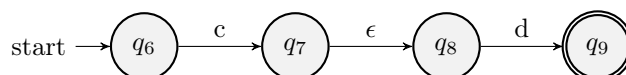


On peut maintenant construire itérativement l'automate fini équivalent à e , en commençant par les feuilles, et en remontant tout l'arbre :

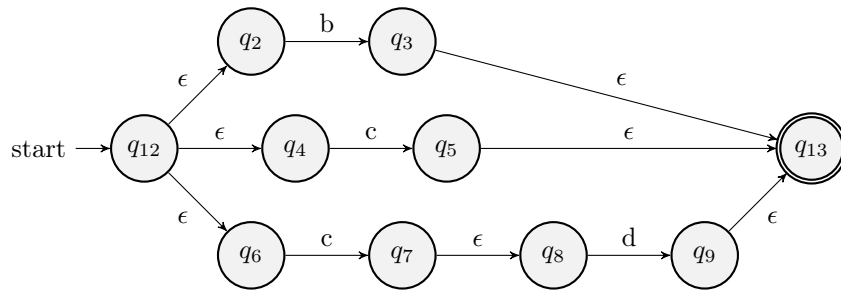
Étape 1 : création des noeuds des feuilles (de gauche à droite) :



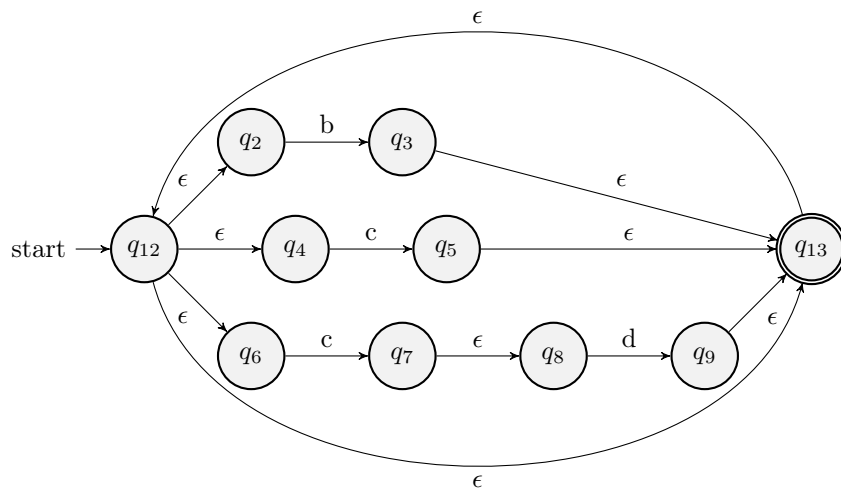
Étape 2 : concaténation de c et d :



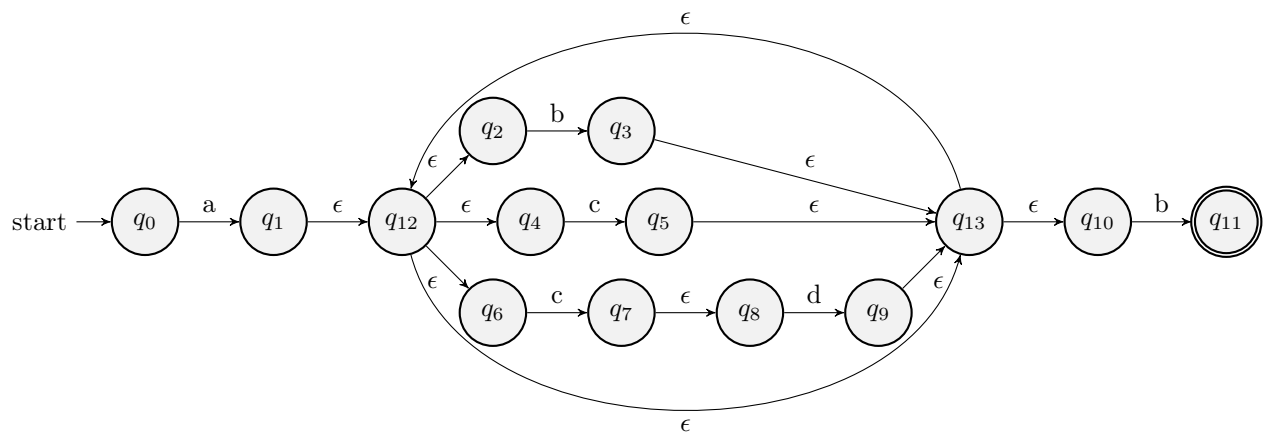
Étape 3 : application du $|$ sur les feuilles b , c et cd :



Étape 4 : application de l'étoile de Kleene :



Étape 5 : concaténation des deux feuilles restantes :



On a construit un automate fini non déterministe (car q_{12} va vers 3 états différents avec la même transition) équivalent à e .

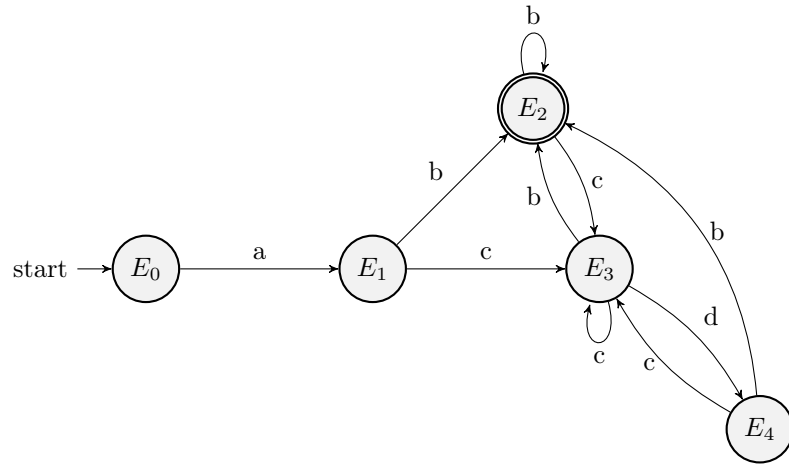
2. Pour le déterminer, il faut supprimer toutes les ϵ -transitions. Pour ce faire, on va déterminer les différents états de l'automate :

- Depuis q_0 , les états atteignables avec a sont : $q_1, q_{12}, q_2, q_4, q_6, q_{13}, q_{10}$.
- Depuis ces derniers, on peut atteindre :
 - $q_2, q_3, q_4, q_6, q_{10}, q_{11}, q_{12}, q_{13}$ avec b ,
 - $q_2, q_4, q_5, q_6, q_7, q_8, q_{10}, q_{12}, q_{13}$ avec c .
- Si on a ab , on peut atteindre :
 - $q_2, q_3, q_4, q_6, q_{10}, q_{11}, q_{12}, q_{13}$ avec b . On remarque que ce sont les mêmes que les précédents, il y a donc une boucle.
 - $q_2, q_4, q_5, q_6, q_7, q_8, q_{10}, q_{12}, q_{13}$ avec c . On remarque que ce sont les mêmes que les précédents, on pourra donc relier le noeud avec le précédent.
- Si on a ac , on peut atteindre :
 - $q_2, q_3, q_4, q_6, q_{10}, q_{11}, q_{12}, q_{13}$ avec b . On remarque que ce sont les mêmes que les précédents, on pourra donc relier le noeud avec celui-ci.
 - $q_2, q_4, q_5, q_6, q_7, q_8, q_{10}, q_{12}, q_{13}$ avec c . On remarque que ce sont les mêmes que les précédents, on pourra donc relier le noeud avec le précédent.
 - $q_2, q_4, q_6, q_9, q_{10}, q_{13}, q_{12}$ avec d .
- Si on a acd , on peut atteindre :
 - $q_2, q_3, q_4, q_6, q_{10}, q_{11}, q_{12}, q_{13}$ avec b . On remarque que ce sont les mêmes que les précédents, on pourra donc relier le noeud avec celui-ci.
 - $q_2, q_4, q_5, q_6, q_7, q_8, q_{10}, q_{12}, q_{13}$ avec c . On remarque que ce sont les mêmes que les précédents, on pourra donc relier le noeud avec le précédent.

On trouve 5 états différents en déterminisant cet algorithme :

- E_0 , l'état de départ,
- $E_1 = q_1, q_{12}, q_2, q_4, q_6, q_{13}, q_{10}$ le noeud a ,
- $E_2 = q_2, q_3, q_4, q_6, q_{10}, q_{11}, q_{12}, q_{13}$ l'état final,
- $E_3 = q_2, q_4, q_5, q_6, q_7, q_8, q_{10}, q_{12}, q_{13}$ l'état où c boucle,
- $E_4 = q_2, q_4, q_6, q_9, q_{10}, q_{12}, q_{13}$ le noeud avec d .

On construit l'automate fini déterministe suivant :



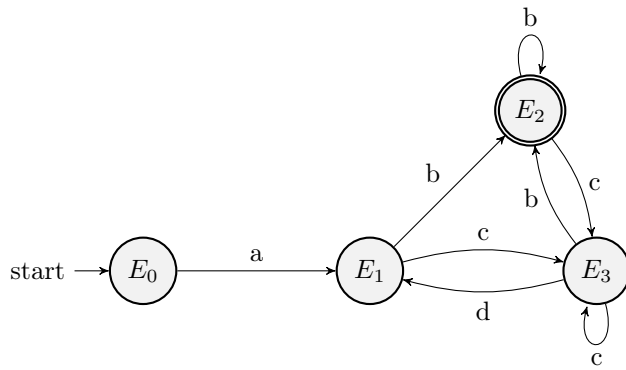
3. Pour savoir si l'AFD précédent est minimisable, on pose les transitions suivantes :

- $E_0 : a,$
- $E_1 : b, c,$
- $E_2 : b, c,$
- $E_3 : b, c, d,$
- $E_4 : b, c.$

On peut ensuite catégoriser chaque état :

- $\boxed{E_0, E_1, E_3, E_4, p} \parallel \boxed{E_2}$ E_2 se distingue des autres car c'est l'état final. p est le puit qui a été ajouté pour reproduire un automate complet.
- $\boxed{E_0} \boxed{E_1, E_3, E_4, p} \parallel \boxed{E_2}$ on distingue E_0 par a .
- $\boxed{E_0} \boxed{E_3} \boxed{E_1, E_4, p} \parallel \boxed{E_2}$ on distingue E_3 par d .
- $\boxed{E_0} \boxed{E_3} \boxed{E_1, E_4} \boxed{p} \parallel \boxed{E_2}$ on distingue E_1 et E_4 par b .

On ne peut pas distinguer E_1 et E_4 . Ces deux états peuvent donc être fusionnés, pour nous donner l'AFD minimal suivant :

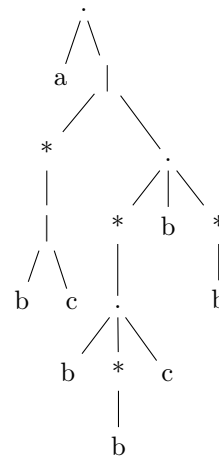
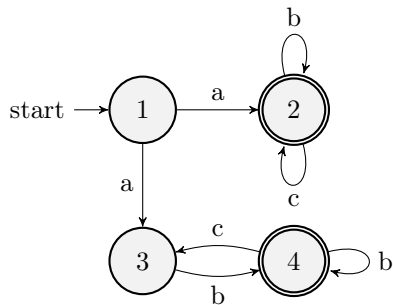


Exercise 11

Soit l'automate fini $B = (\{a, b, c\}, \{1, 2, 3, 4\}, \{1\}, \{2, 4\}, \{1a2, 1a3, 2b2, 2c2, 3b4, 4c3, 4b4\})$.

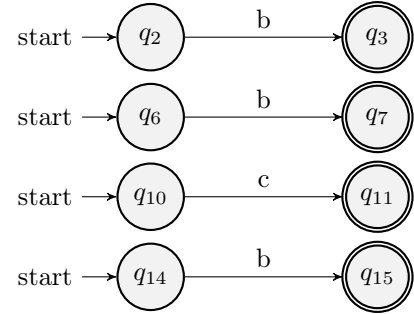
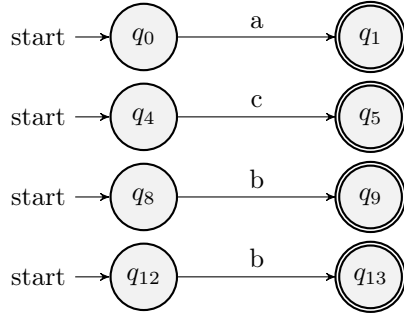
1. Dessiner un automate fini déterministe équivalent à B .
2. Minimisez cet AFD.

1. On peut construire l'automate fini suivant (à gauche) avec la définition de B :

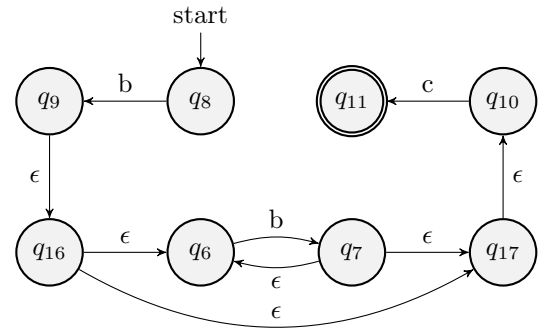
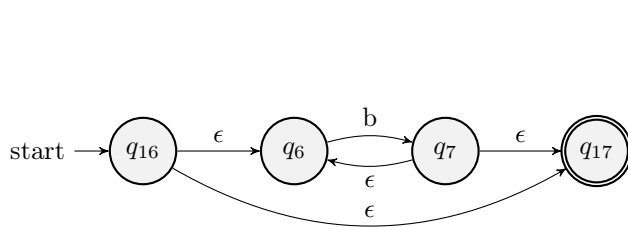


C'est un automate fini non déterministe, qui reconnaît l'expression régulière $e = a((b|c)^*(bb^*c)^*bb^*)$. L'arbre syntaxique de cette expression régulière est à droite de l'automate. En suivant le même procédé que tout à l'heure, on peut normaliser cet automate puis le déterminer.

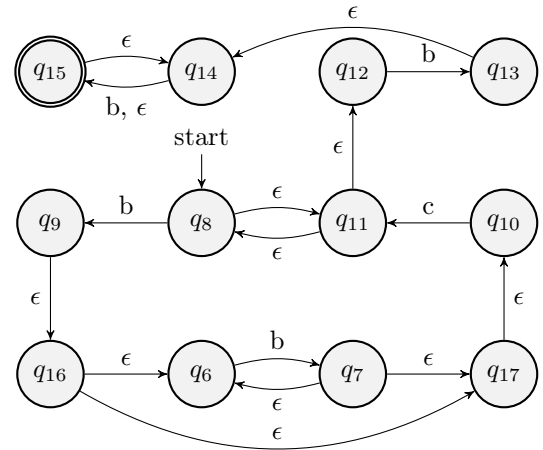
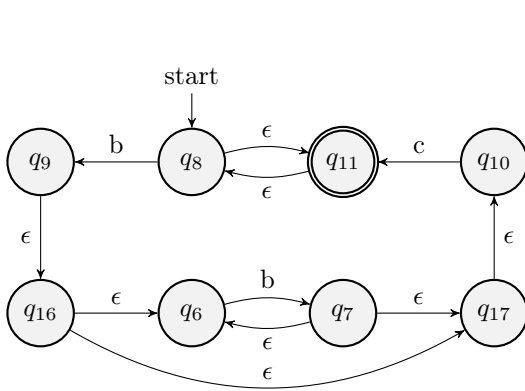
Étape 1 : création des noeuds des feuilles (de gauche à droite) :



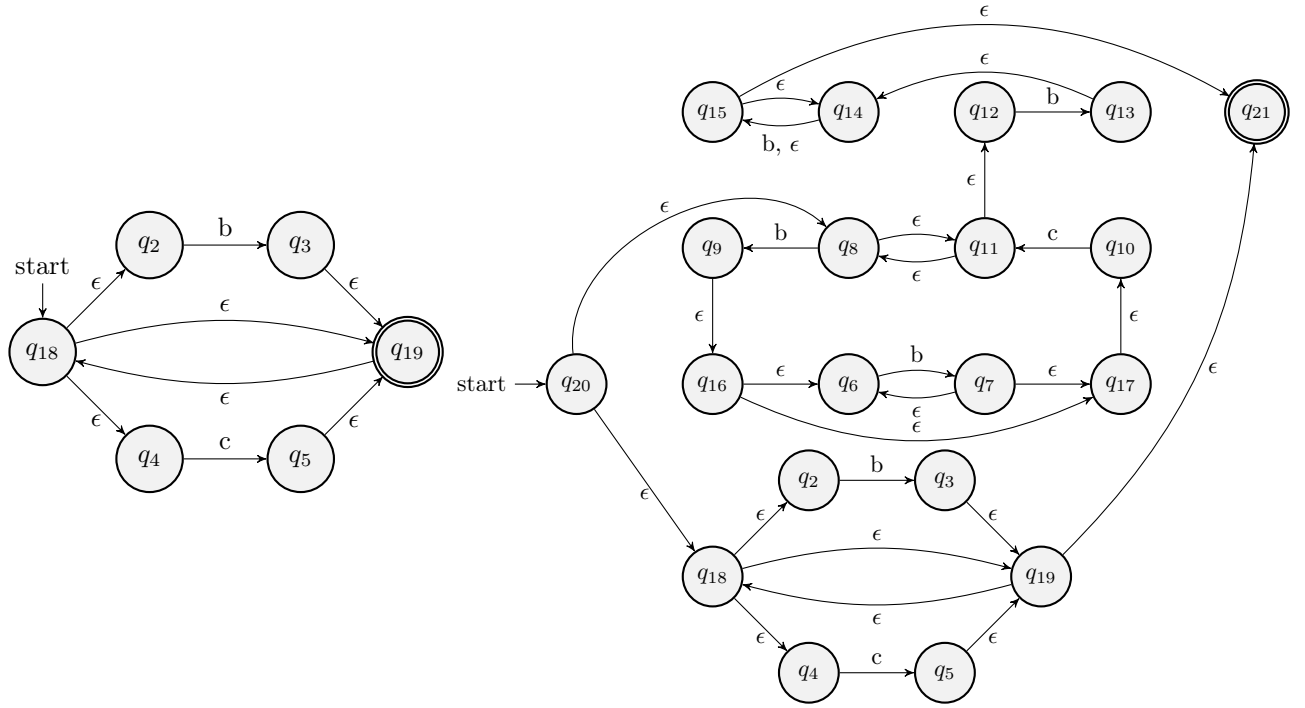
Étape 2 : création de b^* et concaténation de bb^*c :



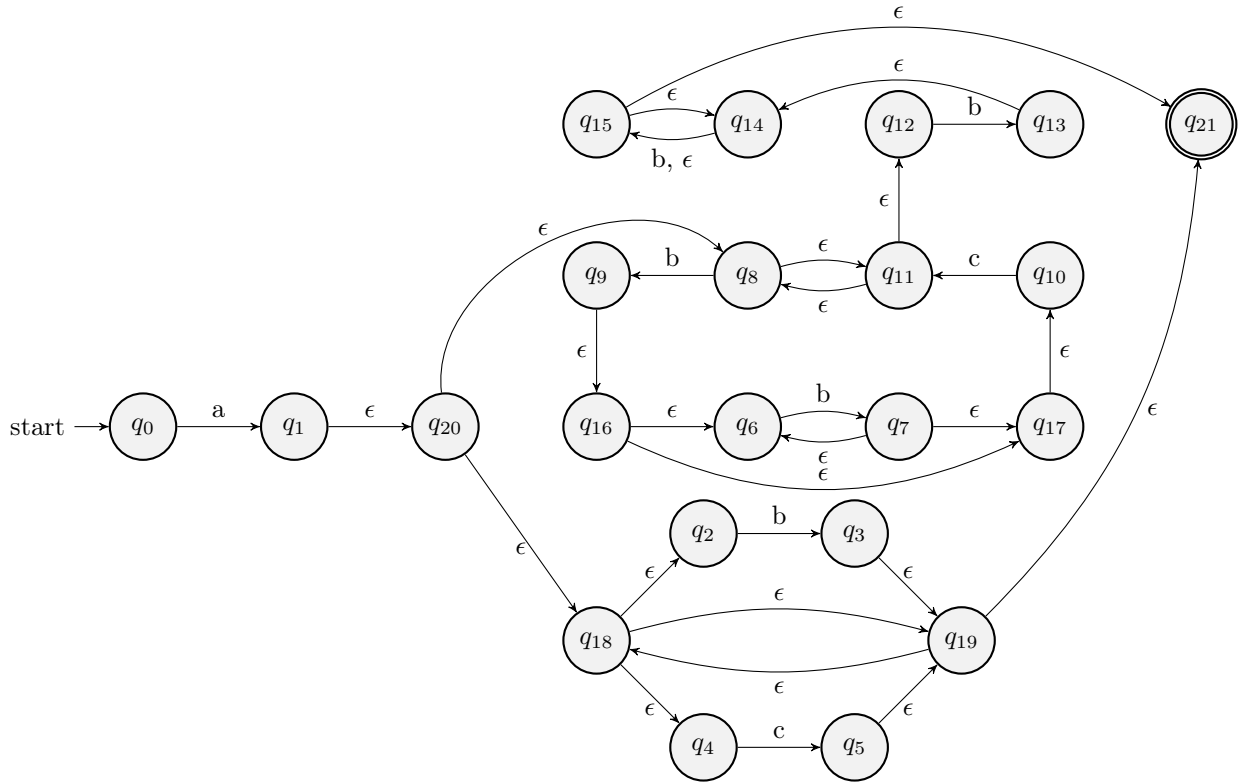
Étape 3 : création de $(bb^*c)^*$ et concaténation à bb^* pour donner $(bb^*c)^*bb^*$:



Étape 4 : $(b|c)^*$ et choix entre $(b|c)^*$ et $(bb^*c)^*bb^*$:

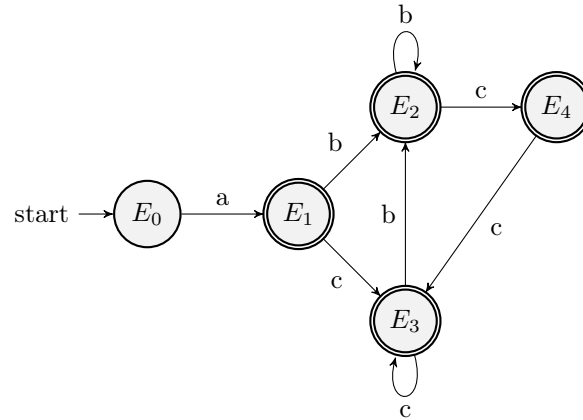


Dernière étape : concaténation avec a :



On a construit l'AFN depuis l'expression régulière décrite par B . Il n'y a plus qu'à appliquer l'algorithme

de détermination pour avoir un AFD équivalent à B :



Pour le réaliser, on a remarqué qu'il y a 5 états différents :

- $E_0 = \{q_0\}$,
- $E_1 = \{q_1, q_2, q_4, q_8, q_{11}, q_{12}, q_{18}, q_{19}, q_{20}, q_{21}\}$,
- $E_2 = \{q_2, q_3, q_4, q_6, q_7, q_8, q_{10}, q_{12}, q_{13}, q_{14}, q_{15}, q_{16}, q_{17}, q_{18}, q_{19}, q_{21}\}$,
- $E_3 = \{q_2, q_4, q_5, q_{18}, q_{19}, q_{21}\}$,
- $E_4 = \{q_2, q_4, q_5, q_8, q_{11}, q_{12}, q_{18}, q_{19}, q_{21}\}$.

2. On pose les transitions suivantes :

- $E_0 : a$,
- $E_1 : b, c$,
- $E_2 : b, c$,
- $E_3 : b, c$,
- $E_4 : c$.

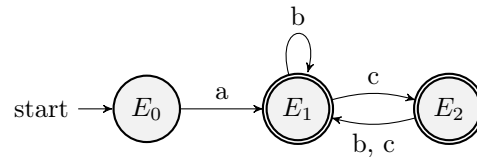
On peut ensuite catégoriser chaque état :

- | | |
|----------|----------------------|
| E_0, p | E_1, E_2, E_3, E_4 |
|----------|----------------------|
- | | | |
|-------|-----|----------------------|
| E_0 | p | E_1, E_2, E_3, E_4 |
|-------|-----|----------------------|

 on distingue E_0 de p par a .
- | | | | |
|-------|-----|-----------------|-------|
| E_0 | p | E_1, E_2, E_3 | E_4 |
|-------|-----|-----------------|-------|

 on distingue E_1, E_2, E_3 de E_4 par b .

On ne peut pas distinguer E_1, E_2 et E_3 . Ces trois états peuvent donc être fusionnés, pour nous donner l'AFD minimal suivant :



Exercice 12 (amélioration des TPs précédents)

Vous pouvez reprendre les exercices précédents afin d'améliorer la reconnaissance des jetons du langage C.

► Voir TP3.

2 Analyse descendante récursive

2.1 TD/TP 5

Exercice 13 (analyse descendante récursive)

Soit la grammaire non récursive à gauche vue en cours $G_{ENR} = (\{0, 1, \dots, 9, +, *, (,)\}, \{E, R, T, S, F\}, X, E)$ avec les règles X suivantes :

$$\begin{aligned} E &\rightarrow TR \\ R &\rightarrow +TR|\varepsilon \\ T &\rightarrow FS \\ S &\rightarrow *FS|\varepsilon \\ F &\rightarrow (E)|0|1|\dots|9 \end{aligned}$$

Soit le programme C vérifiant un mot du langage $L(G_{ENR})$: **analdesc.c**.

1. Dessiner l'arbre de dérivation associé au mot : **1+2+3*4**. Dessiner l'arbre des appels récursifs des fonctions E, R, T, S, F lorsqu'on reconnaît ce même mot. Que remarquez-vous ?
2. On souhaite implémenter l'évaluation de la valeur d'une expression arithmétique pendant sa vérification syntaxique. La multiplication sera prioritaire par rapport à l'addition et l'associativité des deux opérateurs sera à gauche (de gauche à droite). Annoter l'arbre obtenu à la question précédente pour indiquer où sont effectuées les 3 opérations et comment les fonctions transmettent leurs résultats.
3. Sur le modèle du vérificateur, écrire un programme évaluant la valeur d'une expression arithmétique. Par exemple, **evaldesc** sur la chaîne **1+2+(2+1)*3** retournera 12. Attention, on précisera pour chaque opérateur le type d'associativité, gauche ou droite, employée dans le programme.
4. Sur le modèle du vérificateur, écrire un programme traduisant une expression arithmétique en sa forme postfixée (polonaise inversée). Par exemple, **postdesc** sur la chaîne **1+2+(2+1)*3** retournera **12+21+3*+**. On utilisera l'associativité à gauche pour l'addition et la multiplication.
5. Sur le modèle du vérificateur, et en utilisant le type abstrait **Arbin** implémenté en C, écrire un analyseur syntaxique produisant et affichant l'arbre abstrait correspondant à une expression. On utilisera l'associativité à gauche pour l'addition et la multiplication. Par exemple, **arbindesc** sur la chaîne **1+2+(2+1)*3** affichera :

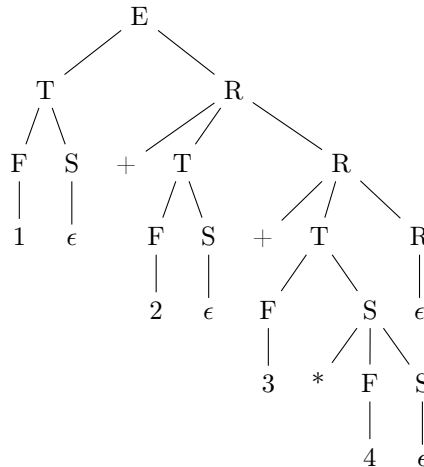
```

+
  +
    1
    2
  *
    +
      2
      1
    3

```

► Voir TP4 (pour les questions 3, 4 et 5).

1. L'arbre de dérivation et l'arbre des appels récursif est plus ou moins similaire :



On remarque que la priorité des opérations est respectée d’emblée dans cet arbre.

2. Il suffit de faire renvoyer des entiers à toutes les fonctions, et à faire prendre un paramètre entier à R et S (pour que ces fonctions puissent effectuer le calcul dans leur corps).
3. ► Voir TP4-5
4. ► Voir TP4-5
5. ► Voir TP4-5

2.2 TD/TP 6

Exercice 14 (Calculatrice CLI)

Écrire un programme évaluant la valeur d’une expression arithmétique composée de nombres flottants non signés et des opérateurs $()$, $^$, $*$, $/$, $+$, $-$. Attention à respecter les associativités des opérateurs! Vous utiliserez flex pour la partie analyse lexicale. Par exemple, `calcdesc` sur la chaîne `5.12-2.56-2^2^3/100` retournera 0.

► Voir TP6

2.3 TD/TP 7

Exercice 15 (TD et TP Langage des expressions régulières)

On désire écrire un traducteur régulière en automate d'état fini non déterministe par l'algorithme de Thompson. Une expression régulière de base est constituée d'une lettre minuscule comprise entre a et z (symboles terminaux), du symbole '@' pour représenter epsilon ε , du symbole '0' pour représenter le langage vide \emptyset . Une expression régulière (er) est de base ou est obtenue par concaténation (implicite) de deux er, par union de deux er (symbole |), par l'opération * appliquée à une er. Le parenthésage est permis pour modifier l'ordre de priorité des opérateurs qui est du plus petit au plus grand : |, concaténation, *, ().

1. Écrire une grammaire "naturelle" engendrant des expressions régulières et prouver son ambiguïté.
2. Écrire une grammaire récursive à gauche et non ambiguë des expressions régulières (S, E, T, F).
3. Écrire une grammaire non récursive à gauche et non ambiguë des expressions régulières : dérécursiver la grammaire précédente (S, X, E, R, T, Y, F).
4. Écrire un analyseur récursif descendant construisant l'arbre abstrait correspondant à une expression régulière.
5. Optimiser l'arbre abstrait construit en tenant compte des règles suivantes, e étant une er quelconque :
 - $e** = e*$,
 - $@* = @ = 0*$,
 - $@e = e@ = e$, $0e = e0 = 0$,
 - $0|e = e|0 = e$.
6. Fabriquer la table d'analyse de l'automate à pile en analyse descendante.
7. Dessiner les états successifs de la pile lors de la reconnaissance de l'expression $ab*|c$.
8. Écrire la fonction `nombreEtat(Arbin a)` calculant le nombre d'état de l'automate fini non déterministe obtenu par l'algorithme de Thompson.
9. Programmer l'algorithme de Thompson en implémentant l'automate non déterministe par un tableau dynamique d'entiers de `nombreEtat(a)` lignes et 3 colonnes. Chaque ligne correspond à un état. La première colonne de ce tableau contient le symbole (@a-z) de la ou les transitions sortantes. Les deux autres colonnes contiennent le ou les états de destination de la ou des deux transitions possibles.

1. On définit la grammaire de manière inductive :

Base a, b, ..., z, ε et \emptyset sont des expressions régulières.

Règles Si e et r sont deux expressions régulières :

- $(e|r)$ est une expression régulière,
- (er) est une expression régulière,
- $(e*)$ est une expression régulière.

On peut déduire la grammaire associée suivante : $E \rightarrow a|b|\dots|z|(E'|E)|(EE)|(E*)|@|0$. Comme les parenthèses peuvent être redondantes, elles peuvent être enlevées : $a|bc* = a|(b(c*))$. Ainsi, on fini sur la grammaire intuitive suivante : $E \rightarrow a|b|\dots|z|E+E|EE|E*|(E)|@|0$. Son ambiguïté est assez

évidente. Prenons l'exemple $a|bc^*$, qui peut être obtenue par la dérivation suivante :

$$\begin{aligned} E &\rightarrow E|E \\ &\rightarrow a|EE \\ &\rightarrow a|bE^* \\ &\rightarrow a|bc^* \end{aligned}$$

Mais aussi la dérivation suivante :

$$\begin{aligned} E &\rightarrow EE \\ &\rightarrow EE^* \\ &\rightarrow E|EE^* \\ &\rightarrow a|EE^* \\ &\rightarrow a|bE^* \\ &\rightarrow a|bc^* \end{aligned}$$

2. Soit la grammaire S, E, T, F suivante :

$$\begin{aligned} S &\rightarrow S'|E|E \\ E &\rightarrow ET|T \\ T &\rightarrow T^*|F \\ F &\rightarrow (S)|a|b|\dots|z|@|0 \end{aligned}$$

3. Soit la grammaire S, X, E, R, T, Y, F non-réursive à gauche suivante :

$$\begin{aligned} S &\rightarrow EX \\ X &\rightarrow '|EX|\varepsilon \\ E &\rightarrow TR \\ R &\rightarrow TR|\varepsilon \\ T &\rightarrow FY \\ Y &\rightarrow *Y|\varepsilon \\ F &\rightarrow (S)|a|b|\dots|z|@|0 \end{aligned}$$

4. ► voir TP7.

5. ► voir TP7.

6. Pour fabriquer la table d'analyse de l'automate à pile en analyse descendante, il faut déjà calculer les premiers et les suivants de chaque règles. Les **premiers** sont obtenus à partir de la dérivation en symboles terminaux de la première règle à droite. Si cette règle produit ε , on prend les premiers de la deuxième règle à droite, et ainsi de suite.

$$\begin{aligned} \text{premiers}(F) &= \{ (, a, b, \dots, z, @, 0 \} \\ \text{premiers}(Y) &= \{ *, \varepsilon \} \\ \text{premiers}(T) &= \{ (, a, b, \dots, z, @, 0 \} \\ \text{premiers}(R) &= \{ (, a, b, \dots, z, @, 0, \varepsilon \} \\ \text{premiers}(E) &= \{ (, a, b, \dots, z, @, 0 \} \\ \text{premiers}(X) &= \{ |, \varepsilon \} \\ \text{premiers}(S) &= \{ (, a, b, \dots, z, @, 0 \} \end{aligned}$$

Les **suivants** sont les symboles qui arrivent après les premiers. L'algorithme de calcul des **suivants** est :

- (a) Ajouter un marqueur de fin (\$) à l'ensemble des suivants de l'axiome.
- (b) Pour chaque production $A \rightarrow \alpha B \beta$ où B est un non-terminal, ajouter le contenu de $\text{premiers}(\beta)$ dans $\text{suivants}(B)$, sauf ε .
- (c) Pour chaque production $A \rightarrow \alpha B$, ajouter $\text{suivants}(A)$ à $\text{suivants}(B)$.
- (d) Pour chaque production $A \rightarrow \alpha B \beta$ avec $\varepsilon \in \text{premiers}(\beta)$, ajouter $\text{suivants}(A)$ à $\text{suivants}(B)$.
- (e) Recommencer à partir de l'étape (c), jusqu'à ce que les ensembles restent inchangés.

On trouve donc le tableau des suivants :

S	\$,)
X	
E	
R	
T	(, a, b, ..., z, @, 0
Y	
F	*

Étapes (a) et (b) pour chaque règle

S	\$,)
X	\$,)
E	
R	
T	(, a, b, ..., z, @, 0
Y	
F	*

Étapes (c) pour chaque règle

S	\$,)
X	\$,)
E	, \$,)
R	
T	(, a, b, ..., z, @, 0
Y	
F	*

Étapes (d) pour chaque règle

S	\$,)
X	\$,)
E	, \$,)
R	, \$,)
T	(, a, b, ..., z, @, 0, , \$,)
Y	(, a, b, ..., z, @, 0, , \$,)
F	*, (, a, b, ..., z, @, 0, , \$,)

Répétitions des étapes (c) et (d) pour chaque règle

On peut maintenant fabriquer la table d'analyse de l'automate à pile en analyse descendante. Pour ce faire, on prend tous les premiers de chaque règles, et on ajoute la règle de production dans la case. Si les $\varepsilon \in \text{premiers}$, on prend tous les suivants, et ils produisent ε :

	a b ... z @ 0	()	*		ε	\$
S	$S \rightarrow EX$	$S \rightarrow EX$					
X			$X \rightarrow \varepsilon$		$X \rightarrow EX$		$X \rightarrow \varepsilon$
E	$E \rightarrow TR$	$E \rightarrow TR$					
R	$R \rightarrow TR$	$R \rightarrow TR$	$R \rightarrow \varepsilon$		$R \rightarrow \varepsilon$		$R \rightarrow \varepsilon$
T	$T \rightarrow FY$	$T \rightarrow FY$					
Y	$Y \rightarrow \varepsilon$	$Y \rightarrow \varepsilon$	$Y \rightarrow \varepsilon$	$Y \rightarrow *Y$	$Y \rightarrow \varepsilon$		$Y \rightarrow \varepsilon$
F	$F \rightarrow a b ... z @ 0$	$F \rightarrow (S)$					

7. En se servant de la table d'analyse descendante, voici le tableau qui résume l'état de la pile qui reconnaît l'expression régulière $ab^*|c$:

Pile	Expression régulière	Action
\$S	$ab^* c$	$S \rightarrow EX$
\$XE	$ab^* c$	$E \rightarrow TR$
\$XRT	$ab^* c$	$T \rightarrow FY$
\$XRYF	$ab^* c$	$F \rightarrow a b \dots z @ 0$
\$XRYa	$ab^* c$	
\$XRY	$b^* c$	$Y \rightarrow \varepsilon$
\$XR	$b^* c$	$R \rightarrow TR$
\$XRT	$b^* c$	$T \rightarrow FY$
\$XRYF	$b^* c$	$F \rightarrow a b \dots z @ 0$
\$XRYb	$b^* c$	
\$XRY	$* c$	$Y \rightarrow *Y$
\$XRY*	$* c$	
\$XRY	$ c$	$Y \rightarrow \varepsilon$
\$XR	$ c$	$R \rightarrow \varepsilon$
\$X	$ c$	$R \rightarrow EX$
\$XE	$ c$	
\$XE	c	$E \rightarrow TR$
\$XRT	c	$T \rightarrow FY$
\$XRYF	c	$F \rightarrow a b \dots z @ 0$
\$XRYc	c	
\$XRY	$\$$	$Y \rightarrow \varepsilon$
\$XR	$\$$	$R \rightarrow \varepsilon$
\$X	$\$$	$X \rightarrow \varepsilon$
\$	$\$$	

8. ► voir TP7.

9. ► voir TP7.

3 Analyse ascendante LR

3.1 TD/TP 8

Exercice 17 (TD/TP Bison)

On reprend l'exercice 15 en analyse ascendante.

1. En utilisant la grammaire récursive à gauche non ambiguë, écrire un source bison qui produise et affiche l'arbre abstrait associé à une expression régulière (sans utiliser flex)?
2. En prenant le source bison suivant (attention, à la concaténation qui est implicite!) :

```
%left '|'
%left CONCAT
%left '*'
%%
expr: '(' expr ')' { $$ = $2; }
    | expr expr %prec CONCAT { $$ = ab_construire('.', $1, $2); }
    | expr '|' expr { $$ = ab_construire('|', $1, $3); }
    | expr '*' { $$ = ab_construire('*', $1, ab_creer()); }
    | SYMBOLE { $$ = ab_construire($1, ab_creer(), ab_creer()); }
```

La compilation bison avertit : »4 conflits par décalage/réduction « ! Examinez le fichier y.output afin de résoudre ce problème !

3. En utilisant la grammaire »intuitive « précédente et les règles de priorités et associativités trouvées à la question précédente, écrire un source bison traduisant une expression régulière en son arbre abstrait.

► Voir TP8.

3.2 TD/TP 9

Exercice 18 (Calculatrice logique)

En reprenant l'architecture de la calculatrice (calc.l et calc.y) vue en cours, écrire une calculatrice logique d'ordre 0. Une expression logique de base est constituée des constantes 0 ou false ou faux, ou 1 ou true ou vrai. Une expression logique est de base ou est obtenue par conjonction (&), disjonction (|), implication (->), équivalence (==), ou exclusif (^) de deux expressions logiques, ou encore par négation (!) d'une expression logique. Le parenthésage est permis pour modifier l'ordre de priorité des opérateurs qui est du plus petit au plus grand : { |, ->, ==, ^ }, &, !, (). Les expressions logiques sont évaluées de gauche à droite.

1. Écrire logic.l, logic.y pour évaluer une expression logique d'ordre 0.
2. Ajouter à cette calculatrice la fonctionnalité suivante : 26 variables logiques, nommées a-z, permettent de conserver le résultat de calculs précédents. L'affectation de ces variables et leur utilisation ressemblera à la syntaxe C. Par exemple, $a = 0|1 \rightarrow 0$ puis $b = a \rightarrow (0|1)$. Les variables seront impantées dans un tableau global de 26 entiers, réalisant une table des symboles rudimentaires.

► Voir TP9.

3.3 TD/TP 10

Exercice 19 (Collection canonique SLR)

Soit la grammaire $G_{ETF} = (\{0, 1, \dots, 9, +, *, (,)\}, \{E, T, F\}, R, E)$ avec les règles R suivantes :

$$\begin{aligned} E &\rightarrow E + T | T \\ T &\rightarrow T * F | F \\ F &\rightarrow (E) | 0 | 1 | \dots | 9 \end{aligned}$$

1. Calculer la collection canonique (AFD) des ensembles d'items LR(0) de la grammaire augmentée associée à G_{ETF} .
2. Construire les tables d'analyse Action et Successeur après avoir calculé les TabSuivants des non terminaux.
3. Analyser l'expression $(5 + 3) * 2\$$.

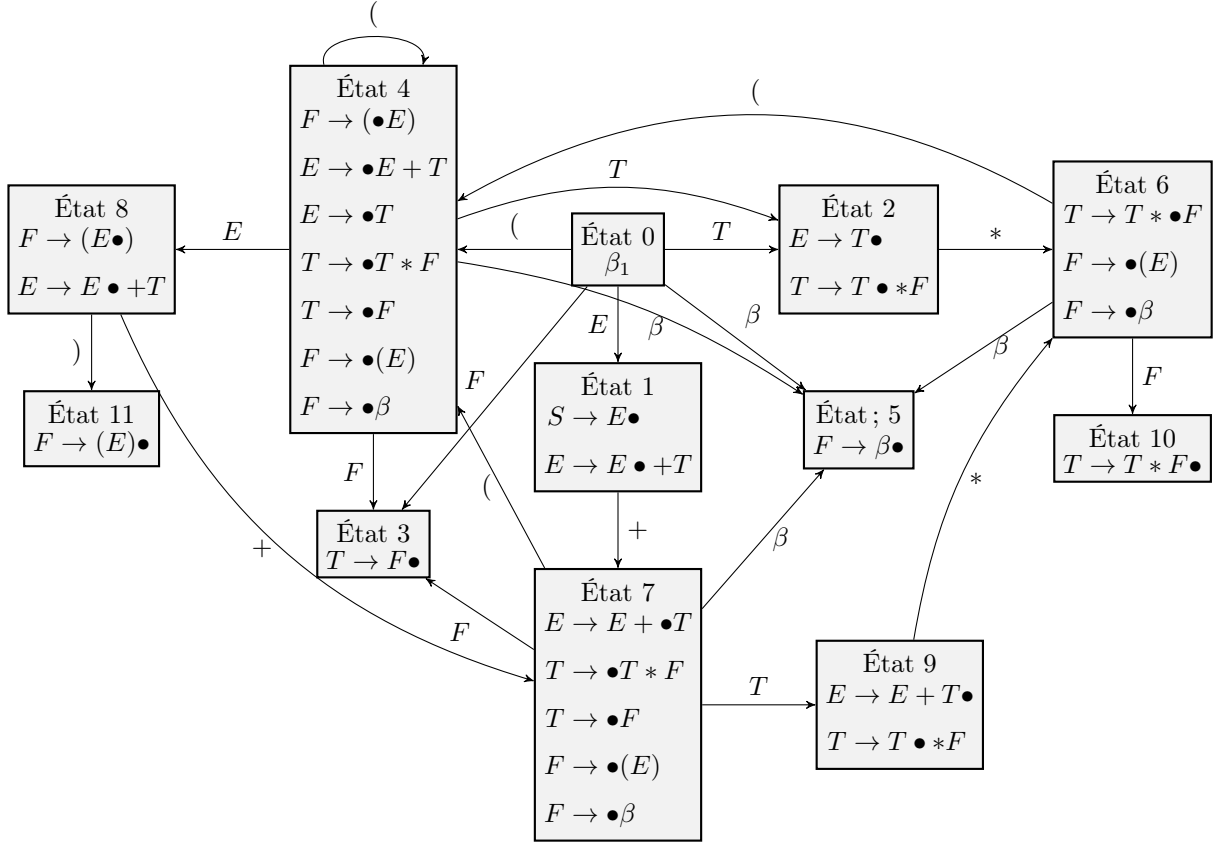
1. Construisons la collection canonique de la grammaire augmentée associée à G_{ETF} . Posons la grammaire augmentée $G'_{ETF} = (\{0, 1, \dots, 9, +, *, (,)\}, \{S, E, T, F\}, R, S)$ avec les règles R suivantes :

$$\begin{aligned} S &\rightarrow E \\ E &\rightarrow E + T | T \\ T &\rightarrow T * F | F \\ F &\rightarrow (E) | 0 | 1 | \dots | 9 \end{aligned}$$

Posons $\beta = 0 | 1 | \dots | 9$ pour simplifier la collection canonique. La fermeture β_1 de G'_{ETF} est la suivante :

$$\begin{aligned} S &\rightarrow \bullet E \\ E &\rightarrow \bullet E + T \\ E &\rightarrow \bullet T \\ T &\rightarrow \bullet T * F \\ T &\rightarrow \bullet F \\ F &\rightarrow \bullet (E) \\ F &\rightarrow \bullet \beta \end{aligned}$$

Construisons maintenant l'automate associé par calcul de fermetures successives (ajouter à l'état actuel les règles de productions de $\bullet\alpha$ avec α un symbole non terminal, puis faire toutes les transitions) :



2. Tout d'abord, on construit le tableau des suivants afin de déterminer les réductions possibles :

E	\$, +,)
T	*, \$, +,)
F	(, 0, ..., 9, *, \$, +,)

On peut maintenant contruire la table d'Action et Successeur suivante :

	Actions						Successeurs		
	β	$+$	$*$	$($	$)$	$\$$	E	T	F
0	S5			S4			1	2	3
1		S7				Accepter			
2		$R(E \rightarrow T)$	S6		$R(E \rightarrow T)$	$R(E \rightarrow T)$			
3		$R(T \rightarrow F)$	$R(T \rightarrow F)$		$R(T \rightarrow F)$	$R(T \rightarrow F)$			
4	S5			S4			8	2	3
5	$R(F \rightarrow \beta)$	$R(F \rightarrow \beta)$	$R(F \rightarrow \beta)$	$R(F \rightarrow \beta)$	$R(F \rightarrow \beta)$	$R(F \rightarrow \beta)$			
6	S5			S4					10
7	S5			S4				9	3
8		S7			S11				
9		$R(E \rightarrow E + T)$	S6		$R(E \rightarrow E + T)$	$R(E \rightarrow E + T)$			
10		$R(T \rightarrow T * F)$	$R(T \rightarrow T * F)$		$R(T \rightarrow T * F)$	$R(T \rightarrow T * F)$			
11	$R(F \rightarrow (E))$	$R(F \rightarrow (E))$	$R(F \rightarrow (E))$	$R(F \rightarrow (E))$	$R(F \rightarrow (E))$	$R(F \rightarrow (E))$			

3. Examinons l'expression $(5 + 3) * 2\$$. Il ne faut pas oublier d'ajouter l'état actuel dans la pile, et de supprimer cet état lors d'une réduction. Il faut alors regarder l'état précédent de la pile pour savoir quel est le successeur du symbole non-terminal de la réduction pour savoir quel sera le prochain état.

État	Flot d'entrée	Pile	Action
0	$(5 + 3) * 2\$$	[0]	S4
4	$5 + 3) * 2\$$	[0, 4(]]	S5
5	$+3) * 2\$$	[0, 4(], 5(5)]	$R(F \rightarrow \beta)$
3	$+3) * 2\$$	[0, 4(], 3(F)]	$R(T \rightarrow F)$
2	$+3) * 2\$$	[0, 4(], 2(T)]	$R(E \rightarrow T)$
8	$+3) * 2\$$	[0, 4(], 8(E)]	S7
7	$3) * 2\$$	[0, 4(], 8(E), 7(+)]	S5
5	$) * 2\$$	[0, 4(], 8(E), 7(+), 5(3)]	$R(F \rightarrow \beta)$
3	$) * 2\$$	[0, 4(], 8(E), 7(+), 3(F)]	$R(T \rightarrow F)$
9	$) * 2\$$	[0, 4(], 8(E), 7(+), 9(T)]	$R(E \rightarrow E + T)$
8	$) * 2\$$	[0, 4(], 8(E)]	S11
11	$*2\$$	[0, 4(], 8(E), 11())]	$R(F \rightarrow (E))$
3	$*2\$$	[0, 3(F)]	$R(T \rightarrow F)$
2	$*2\$$	[0, 2(T)]	S6
6	$2\$$	[0, 2(T), 6(*)]	S5
5	$\$$	[0, 2(T), 6(*), 5(2)]	$R(F \rightarrow \beta)$
10	$\$$	[0, 2(T), 6(*), 10(F)]	$R(T \rightarrow T * F)$
2	$\$$	[0, 2(T)]	$R(E \rightarrow T)$
1	$\$$	[0, 1(E)]	Accepter

Exercice 20 (TD collection canonique grammaire intuitive des exp. rég.)

Soit la grammaire $G_{regexp} = (\{\beta, |, *, (,)\}, \{e\}, R, e)$ avec les règles de R suivantes (attention, le symbole $|$ faisant partie du vocabulaire terminal, on utilisera $+$ pour représenter l'alternative) :

$$e \rightarrow e + e | ee | e * |(e)|\beta$$

Le jeton β représente une lettre minuscule. On peut, grâce à cette grammaire ambiguë, générer des expressions régulières telles que : $ab|c*$, $(ab)*$, $(a(a(ab)*b)*)*$.

1. Dessinez la collection canonique de cette grammaire puis calculez les premiers et les suivants de e .
2. Calculez la table d'analyse SLR de cette grammaire.
3. Discutez des conflits et résolvez-les avec Bison afin que les opérateurs soient associatifs à gauche et que les priorités ascendantes soient : $|$, la concaténation, $*$.
4. Analysez le texte d'entrée $a|cd*$.

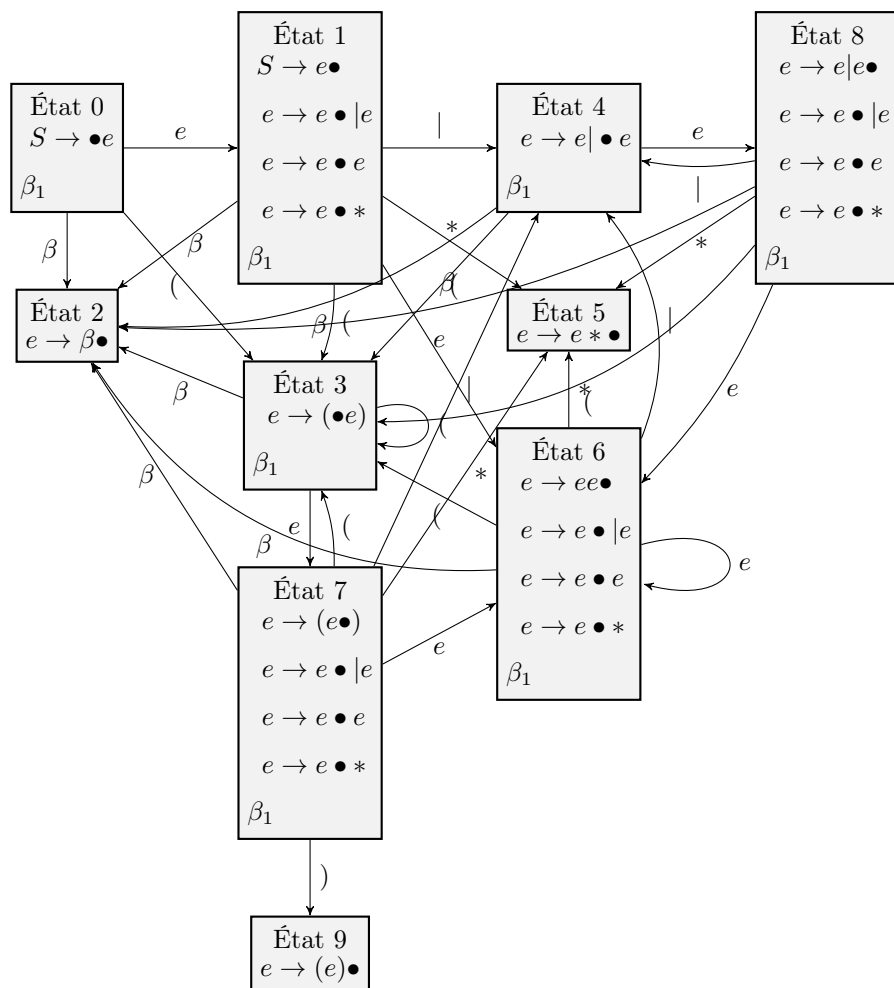
1. Soit la grammaire augmentée G'_{regexp} suivante :

$$\begin{aligned} S &\rightarrow e \\ e &\rightarrow e + e | ee | e * |(e)|\beta \end{aligned}$$

Soit β_1 la fermeture suivante :

$$\begin{aligned} e &\rightarrow \bullet e|e \\ e &\rightarrow \bullet ee \\ e &\rightarrow \bullet e* \\ e &\rightarrow \bullet(e) \\ e &\rightarrow \bullet\beta \end{aligned}$$

Dessignons la collection canonique de cette grammaire :



Comme c'est vraiment le bordel, résumons cet automate avec un tableau qui contient les arêtes et leur symbole de transition pour chaque état :

État	Arêtes
0	(1, e), (2, β), (3, $()$)
1	(2, β), (4, $ $), (5, $*$), (3, $()$), (6, e)
2	
3	(2, β), (3, $()$), (7, e)
4	(2, β), (8, e), (3, $()$)
5	
6	(2, β), (3, $()$), (4, $ $), (5, $*$), (6, e)
7	(2, β), (3, $()$), (4, $ $), (5, $*$), (6, e), (9, $)$)
8	(2, β), (3, $()$), (4, $ $), (5, $*$), (6, e)
9	

On a $suivants(e) = \{\beta, |, *, (), \$\}$.

2. On peut maintenant construire la table d'Action et Successeur :

	Actions						Successeurs
	β	$ $	$*$	$()$	$)$	$\$$	e
0	S2			S3			1
1	S2	S4	S5	S3		Accepter	6
2	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	
3	S2			S3			7
4	S2			S3			8
5	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	
6	S2	S4	S5	S3			6
7	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	
8	S2	S4	S5	S3	S9		6
9	S2	S4	S5	S3			6
8	$R(e \rightarrow e e)$	$R(e \rightarrow e e)$	$R(e \rightarrow e e)$	$R(e \rightarrow e e)$	$R(e \rightarrow e e)$	$R(e \rightarrow e e)$	6
9	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	

3. Il y a 8 conflits dans cette table. Pour les résoudre, il faut ajouter ces directives bison :

```
%left '|'
%left CONCAT SYMBOLE '('
%left '*'
%prec CONCAT
```

Ainsi, en ligne 6, comme la concaténation est prioritaire par rapport à β , $|$ et $()$, c'est la réduction qui sera choisie. Comme $*$ est prioritaire par rapport à la concaténation, c'est le shift qui sera choisi.

Pour la ligne 8, par le même principe, comme β , $*$ et $()$ sont plus prioritaires que $|$, les shifts seront choisis. De plus, comme l'associativité du connecteur est à gauche, c'est le réduit qui est choisi en case $|$. Ainsi, cela donne le tableau définitif suivant :

	Actions						Successeurs
	β		*	()	\$	e
0	S2			S3			1
1	S2	S4	S5	S3		Accepter	6
2	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	$R(e \rightarrow \beta)$	
3	S2			S3			7
4	S2			S3			8
5	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	$R(e \rightarrow e*)$	
6	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	S5	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	$R(e \rightarrow ee)$	6
7	S2	S4	S5	S3	S9		6
8	S2	$R(e \rightarrow e e)$	S5	S3	$R(e \rightarrow e e)$	$R(e \rightarrow e e)$	6
9	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	$R(e \rightarrow (e))$	

4. Passons maintenant à l'analyse de $a|cd*$:

État	Flot d'entrée	Pile	Action
0	$a cd * \$$	[0]	S2
2	$ cd * \$$	[0, 2(a)]	$R(e \rightarrow \beta)$
1	$ cd * \$$	[0, 1(e)]	S4
4	$cd * \$$	[0, 1(e), 4()]	S2
2	$d * \$$	[0, 1(e), 4(), 2(c)]	$R(e \rightarrow \beta)$
8	$d * \$$	[0, 1(e), 4(), 8(e)]	S2
2	$*\$$	[0, 1(e), 4(), 8(e), 2(d)]	$R(e \rightarrow \beta)$
6	$*\$$	[0, 1(e), 4(), 8(e), 6(e)]	S5
5	$\$$	[0, 1(e), 4(), 8(e), 6(e), 5(*)]	$R(e \rightarrow e*)$
6	$\$$	[0, 1(e), 4(), 8(e), 6(e)]	$R(e \rightarrow ee)$
8	$\$$	[0, 1(e), 4(), 8(e)]	$R(e \rightarrow e e)$
1	$\$$	[0, 1(e)]	Accepter