**Langages et Compilation**

***Mini projet***

TORTEL Paul – PILAUDEAU Matthieu – COURSAC Vincent

Groupe F – F7

Table des matières

[Bilan 1](#_Toc477959779)

[Fichier d’entrée 1](#_Toc477959780)

[Structures de données 2](#_Toc477959781)

[Algorithmes 2](#_Toc477959782)

[Traces d’exécutions 3](#_Toc477959783)

# **Bilan**

# **Fichier d’entrée**

Pour représenter une grammaire dans notre programme, il faut la décrire dans un fichier texte correspondant en respectant le format que nous avons choisis. Ainsi, nous avons un fichier par grammaire.

Chaque ligne du fichier texte décrit un symbole non terminal. Exemple :

Expreb = Expreb ou Termeb | Termeb

Cette ligne peut se décomposer de la manière suivante :

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Expreb** |  | **=** |  | **Expreb** |  | **ou** |  | **Termeb** |  | **|** |  | **Termeb** |

*Chaque expression ou operateur est séparé par un espace. Cela facilite la lecture du fichier et clarifie la représentation.*

Il y a deux *opérateurs* gérés dans notre format : **=** et **|**

**=** permet de représenter la flèche 🡪, qui a pour but d’indiquer quel est le non terminal décrit par les règles situées à droite de cet opérateur.

**|** permet de représenter l’opérateur logique « OU », qui permet de créer plusieurs règles pour un symbole non terminal. Dans notre exemple, le symbole Expreb possède donc deux règles :

**Expreb 🡪 Expreb ou Termeb OU Expreb 🡪 Termeb**

Deuxième règle

Première règle

**Exemple concret**

Représentation de la grammaire suivante :

**Représentation théorique**

S 🡪 (L) | a

L 🡪 L,S | S

**Représentation dans notre format texte**

S = ( L ) | a

L = L , S | S

***Représentation en tableaux***

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S** |  | **=** |  | **(** |  | **L** |  | **)** |  | **|** |  | **a** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **L** |  | **=** |  | **L** |  | **,** |  | **S** |  | **|** |  | **S** |

# **Structures de données**

Nous avons utilisé deux classes afin d’encapsuler les données fournies par le fichier d’entrée. Ces classes sont : *Analyzer* et *Symbol.*

La première, ***Analyzer***, nous permet de stocker par exemple, le nom de l’axiome et la grammaire donnée par le fichier d’entrée. Elle nous permet également de regrouper de nombreuses fonctions qui permettront de trouver les premiers et les suivants d’un symbole non-terminal, ainsi que de créer la table d’analyse.

La grammaire est représentée grâce à une map. Les indexes utilisés sont de type String. Ils représentent le nom d’une variable (par exemple A, ou bien même Expression). Le contenu indexé quant à lui est un pointeur vers une instance de la classe Symbol que nous verrons juste après. C’est à partir de cette map que l’on va pouvoir accéder aux symboles non-terminaux et à leurs règles.

**Analyzer**

String \_startSymbol

Map<String, Symbol\*> \_grammar

Vector<String> \_orderedSymbols

readGrammar(..)/displayGrammar()setAllFirst()/setAllFollow()  
createTable()  
*Fonctions utiles*

**Données :**

**Fonctions :**

Le Vector de String nommé *\_orderedSymbols* stocke uniquement les noms des symboles non-terminaux dans l’ordre que de celui du fichier.

Parlons maintenant des fonctions. *readGrammar(..)* est la fonction qui va lire le fichier et stocker les données lues dans nos structures de données, tandis que *displayGrammar()* va quant à elle afficher le contenu de la map \_*grammar* tout juste remplie par *readGrammar().* setAllFirst() et setAllFollow() vont, comme leur nom l’indique, trouver et enregistrer les premiers et les suivants de tous les symboles non-terminaux stockés dans la map \_*grammar*.

La fonction *createTable()* va se servir de la map \_*grammar* pour créer la table d’analyse à partir des premiers et des suivants de chacun des symboles non-terminaux. Elle va en fait, pour chaque symbole NT, définir leur table. L’ensemble des tables des symboles NT forment la table d’analyse.

A

B

a

b

$

…

…

…

…

Table de A

Table de B

Table d’analyse

Pour ce qui est des « fonctions utiles », nous avons implémentées un certain nombre de fonction qui nous sont essentielles dans la création de la table d’analyse (donc dans la recherche des premiers et des suivants). Pour en citer quelques-unes, nous avons implémenté une fonction détectant la présence ou non du mot vide epsilon (# dans le programme). Cette fonction est très utile à de nombreuses reprises notamment lorsque l’on cherche les premiers d’un symbole non-terminal. Nous avons également une fonction nommée *findFirstIfEpsilon* qui permet d’aller chercher les autres premiers d’un symbole non-terminal si ce dernier contient un epsilon dans ses premiers. Exemple : A 🡺 BC avec B 🡺 # | a alors Premier(A) = { a ; # } mais il faut dans ce cas aller chercher les premiers de C comme B peut donner le mot vide. C’est ce que cette fonction fait.

Voyons maintenant la classe ***Symbol***.

**Symbol**

String \_name

Map<String, Vector<String>> \_table

Vector<Vector<String>> \_rules

addRule(..) / addIntoTable()  
getter/setter  
*Fonctions utiles*

**Données :**

**Fonctions :**

Vector <String> \_first

Vector <String> \_follow

Un symbole non-terminal a son nom stocké dans la variable \_*name*. La Map \_*table* est indexée par des Strings, qui représentent les symboles terminaux, tandis que le Vector de String représente quant à lui la règle de remplacement correspondante. Exemple :

|  |  |  |
| --- | --- | --- |
|  | a | $ |
| A | A 🡺 BC | A 🡺 D |

La map \_*table* serra alors de la forme :

|  |  |  |  |
| --- | --- | --- | --- |
| a | |  |  | | --- | --- | | B | C | |
| $ | |  | | --- | | D | |

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | B | a | |
| |  |  |  |  | | --- | --- | --- | --- | | B | C | d | D | |
| |  | | --- | | # | |

Le Vector de Vector de String \_*rules* représente quant à lui les règles du symbole NT. Prenons   
A 🡺 Ba | bCdD | #. On aura alors \_*rules* de la forme :

Le Vector de String \_first et \_follow vont stocker respectivement les premiers et les suivants du symbole NT.

Comme bon nombre de classes, Symbol a un certain nombre de getter/setter. Elle possède notamment les fonctions addRule(), qui sera utilisée lorsque le programme lira le fichier d’entrée afin de définir les règles du symbole NT, ainsi que addIntoTable(), qui sera utilisée lorsque le programme créera la table d’analyse.

Comme la classe Analyzer, la classe Symbol possède des fonctions très pratiques, comme par exemple isIntoTable(…) qui va vérifier s’il existe une règle de remplacement pour un symbole NT lorsqu’il « rencontre » un symbole terminal donné (ex : on a $A (pile) a (mot à reconnaître), alors on va tester s’il existe une règle dans la table de A avec le symbole « a » ).

# **Algorithmes**

## **Lecture du fichier**

La lecture et l’interprétation du fichier texte repose sur un algorithme simple. Il suffit de lire ligne par ligne le fichier et d’analyser chaque ligne grâce à des méthodes simples.

L’analyse repose sur le principe de « split », c’est-à-dire transformer une chaine de caractères en un tableau de sous chaines de caractères en utilisant un symbole pour effectuer le découpage. Par exemple, si on split la chaine « A-B-C-D » par le symbole ‘ **-** ‘, on obtient un tableau de taille 4.



|  |  |  |  |
| --- | --- | --- | --- |
| A | B | C | D |

Split par ‘ – ‘

Grâce à ce procédé, on peut dégager des étapes pour effectuer l’analyse de la ligne.

Prenons comme exemple pour la suite de l’explication la ligne suivante

S = ( L ) | a

|  |  |
| --- | --- |
| **S** | **( L ) | a** |

**1. Split par ‘ = ‘**

*On a dans un tableau de taille 2 d’un côté le non terminal qui est en train d’être défini, et de l’autre ses règles qui attendent d’être analysées*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **(** | **L** | **)** | **|** | **a** |

**2. Split par ‘ ‘**

*On a split la partie de droite du tableau obtenu à l’étape1. On obtient un nouveau tableau contenant toutes les sous chaines de la ligne. Il faut maintenant traiter l’opérateur* ***|***

|  |
| --- |
| a |

|  |  |  |
| --- | --- | --- |
| ( | L | ) |

**3. Création des règles**

## **Elimination de la récursivité à gauche**

# **Traces d’exécutions**