

Projet **: Compilation**

Ecole Nationale Supérieure d’Informatique et d’Analyse des Systèmes

Encadré par :

Pr. Y.Tabii

Pr.R.Oulad lhaj Thami

Année universitaire : 2019-2020

Réalisé par :

AABBAR Adnane

ABED Abir

AIT RAMI Mohamed

ANDAME Younes

Réalisation d’un compilateur du langage R

Remerciements

Avant tout développement sur cette expérience, il parait opportun de commencer ce rapport par des remerciements adressés en toute reconnaissance à notre respectueux professeurs Pr Y.Tabii et Pr R.Oulad lhaj thami pour avoir su nous éclairer sur les bonnes solutions lors des débuts délicats de notre projet, pour les conseils qu’ils ont pu nous prodiguer tout au long de cette période, et pour leur suivi continu de notre travail.

Nous tenons également à remercier nos professeurs à l’École Nationale Supérieure d’Informatique et d’Analyse des Systèmes pour leur formation qui nous a permis de développer les compétences nécessaires à la réalisation de ce travail.

Contexte général du projet

Lors de notre 2ème année, nous sommes menés à réaliser un projet de compilation afin de bien maîtriser les connaissances acquises durant les 14 semaines de cours ainsi que les travaux pratiques de compilation et d’enrichir notre bagage informatique avec plusieurs notions de développement.

Avant d’en venir à la présentation de ce rapport, tâchons de préciser quelques points au sujet de ce projet. Le sujet de notre projet consiste à mettre en place un compilateur du langage R en utilisant le langage C.

Pour se faire, il fallait commencer, en premier lieu, par comprendre et construire la grammaire ainsi que la table des symboles. En deuxième lieu, on a réalisé les différents analyseurs en commençant par l’analyseur lexical ensuite le syntaxique et sémantique.

En dernier lieu, nous nous sommes concentrés sur la phase de la réalisation du générateur du pseudo-code. Dans le cadre de cette phase, nous avons commencé par une étape d’étude du jeu de code machine qui seront utilisés ensuite pour réaliser notre générateur.

Enfin, des tests assurant le bon fonctionnement des modules développés ont été faits.

Le langage R

R est un langage de programmation dont le but est de pouvoir traiter et organiser des jeux de données afin de pouvoir y appliquer des tests statistiques plus ou moins complexes et se représenter ces données graphiquement à l'aide d'une grande variétés de graphiques disponibles.

Développé initialement au début des années 1990 par Ross Ihaka et Robert Gentleman, ce langage basé sur un autre langage statistique appelé **S** (oui, les informaticiens ont beaucoup d'imagination), est désormais maintenu et mis à jour par une équipe de développeurs au sein du [R Project](https://www.r-project.org/). Cette structure garantit des mises à jour fréquentes et une communauté importante d'utilisateurs apporte son aide pour notamment développer de nouvelles fonctionnalités (des librairies, nous en parlerons plus tard) au projet.

Le langage R est en effet très implanté dans la communauté scientifique et universitaire et ses nombreux utilisateurs organisent même [des conférences](http://user2010.org/) pour discuter des nouveautés disponibles sous R.

Malgré quelques critiques généralement portées, ce langage est très utilisé pour diverses raisons:

* il permet d'organiser et traiter des volumes importants de données de manière rapide et flexible. Comme il s'agit d'un langage de programmation, vous pouvez (à peu près) en faire tout ce que vous voulez.
* il permet d'assez facilement créer des graphiques paramétrables afin de pouvoir mieux visualiser le résultat de vos analyses.
* il est entièrement gratuit et sous licence GPL, ce qui signifie que vous pouvez en télécharger les sources et les modifier. C'est notamment grâce à cela qu'une active communauté peut améliorer le langage en permanence.
* il est multi-plateforme. Vous pourrez en effet l'utiliser sous Windows, Mac OS ou Linux.

R est un langage de programmation mais son utilisation est fortement orientée vers les analyses de données et statistiques. Ainsi, si vous souhaitez créer des jeux vidéos, un lecteur vidéo ou MP3 ou tout autre style d'application utilitaire vous faites certainement fausse route.

La grammaire

|  |  |  |
| --- | --- | --- |
| **INST** | ::= | INST | AFFEC | SI | TANTQUE | epsilon |
| **AFFEC** | ::= | ID <-EXPR|ID=EXP |
| **SI** | ::= | if ( <COND> ){INST} |
| **TANTQUE** | ::= | while (COND) {INST} |
| **POUR** | ::= | for (VAR IN SEQ) {INST} |
| **REPEAT** | ::= | Repeat {INST if(COND) break} |
| **COND** | ::= | EXPR RELOP EXPR |
| **RELOP** | ::= | = | != | < | > | <= | >= |
| **ADDOP** | ::= | + | - |
| **MULOP** | ::= | \* | / |
| **EXPR** | ::= | TERM { ADDOP TERM } |
| **TERM** | ::= | FACT { MULOP FACT } |
| **FACT** | ::= | ID | NUM | ( EXPR ) |
| **ID** | ::= | lettre {lettre | chiffre} |
| **NUM** | ::= | chiffre {chiffre} |
| **Chiffre** | ::= | 0|..|9 |
| **Lettre** | ::= | a|b|..|z|A|..|Z |

Les métarègles

Un commentaire est une suite de caractères commençant par #

Un séparateur est un caractère séparateur (espace blanc, tabulation, retour chariot) ou un commentaire ;

Deux ID ou mots clés qui se suivent doivent être séparés par au moins un séparateur ;

Distinction entre minuscule et majuscule ;

Une séquence est deux nombres séparés par deux points ;