Rodent compiler structure

December 20, 2024

1 Abstract

Компилятор языка программирования Rodent был написан на C. А потом в страхе переписан на C++. Некоторые этапы переписывались с нуля много раз.

2 Лексический анализатор

Лексический анализатор был реализован очень нестандартным путём. Хотя многие связанные с ним вещи по типу парсинга операций или сохранения базовых типов из файла написаны достаточно банально, машина конечных состояний у лексера реализованна с помощью полиморфных классов. У нас есть классы состояния, и классы, отвечающие за переходы между состояниями.

3 Синтаксический анализатор

Синтаксический анализатор представляет из себя набор функций, которые рекурсивно вызывают друг друга. Например, функция functionDefinition() вызовет функцию body_(). Класс parser с подробным описанием всех его полей и методов представлен ниже:

Parser(const std::vector<Token>&, const std::string&); - Конструктор от набора токенов, или проще сказать распарсеной программы и имени файла который мы парсим

std::wstring filename_; - имя файла программы void program_(); - эта функция начинает обработку программы

void programThings_(); - тут программа начинает рассматриваться и делиться на три части : определение глобальных переменных, определение функций и подключение модулей void import_(); - эта функция рекурсивно вызывает сначала лексер а затем и парсер

void functionDefinition_(); - определение функции void arguments_(); - аргументы передаваемые функции void body_(); - тело функции void statement_(); - "утверждение", любая конструкция допустимая внутри функций: создание переменных, операторы и т.д.

void expression_(); - выражение

void expr_(); - inline выражение, в отличии от expression_() не заканчивается на ; эта функция рекурсивно вызывает expr0 - expr13 чтобы обработать выражение учитывая приоритет операций

void functionCall_(); - вызов функции void given_(); - аргументы функции void inline_body_(); - тело цикла, if, и прочих операторов void inline_expression(); - обёртка над expr_

bool get_(); - берёт следующую лексему, если лексема является концом файла то возвращает false void get(); - обёртка на get_ size_t now_ = -1; - индекс текущей лексемы

Иные поля и методы не представленные тут являются очевидными. Подробнее можно с ними ознакомится на github

4 Сематический анализатор