



Диагностика синтаксических ошибок в динамически формируемом коде

В рамках проекта лаборатории JetBrains

Автор: Азимов Рустам Шухратуллович, 444 группа

Научный руководитель: м.и.т., ст.пр. С.В. Григорьев

Рецензент: программист "ИнтеллиДжей Лабс" Д.А. Авдюхин

Санкт-Петербургский государственный университет
Кафедра системного программирования

17 мая 2016г.

- Динамический SQL

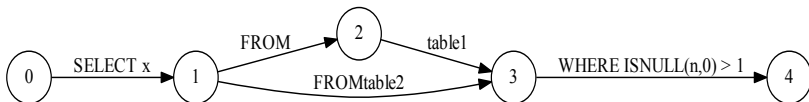
```
IF @X = @Y
    SET @TBL = ' table1 '
ELSE
    SET @TBL = 'table2 '
SET @S = 'SELECT x FROM' + @TBL + 'WHERE ISNULL(n,0) > 1'
EXECUTE (@S)
```

- Динамически формируемые выражения – код на некотором языке, который нужно соответствующим образом поддерживать и обрабатывать
 - ▶ Ошибки в динамически формируемых выражениях обнаруживаются лишь во время выполнения
 - ▶ Поддержка в IDE
 - ▶ Реинжиниринг ПО, разработанного с использованием встроенных языков

- Alvor, Java String Analyzer, PHP SA, Kyung-Goo Doh et al.
 - ▶ Имеется диагностика синтаксических ошибок, но плохо расширяемы и не строят структурного представления кода
- Ослабленный синтаксический анализ регулярной аппроксимации динамически формируемого кода в рамках проекта YaccConstructor
 - ▶ Строит деревья вывода, хорошо расширяем, но синтаксические ошибки игнорируются
 - ▶ **Вход:** эталонная ДКС-грамматика G и граф ДКА без ϵ -переходов над алфавитом терминалов G
 - ▶ **Выход:** конечное представление множества деревьев, соответствующих всем корректным цепочкам, принимаемым входным автоматом

Пример

- IF @X = @Y
 SET @TBL = ' table1 '
ELSE
 SET @TBL = 'table2 '
SET @S = 'SELECT x FROM' + @TBL + 'WHERE ISNULL(n,0) > 1'
EXECUTE (@S)
- Множество значений:
{ 'SELECT x FROM table1 WHERE ISNULL(n,0) > 1' ;
 'SELECT x FROMtable2 WHERE ISNULL(n,0) > 1' }
- Аппроксимация:



Синтаксические ошибки

- Алгоритмы LR семейства (LR, SLR, LALR)
 - ▶ Читают входную строку слева направо
 - ▶ Строят стек состояний
 - ▶ Синтаксическая ошибка на символе, делающем невозможным дальнейшее построение стека состояний по таблице переходов грамматики
- Обобщённый синтаксический анализ (GLR, RNLGR)
 - ▶ Читает входную строку слева направо
 - ▶ Строит организованный в виде графа стек (GSS)
 - ▶ Синтаксическая ошибка на символе, делающем невозможным ни в одном направлении дальнейшее построение GSS состояний по таблице переходов грамматики
- В обоих случаях синтаксическая ошибка на первом символе, делающем прочитанный префикс входной строки – некорректным для рассматриваемой грамматики

Постановка задачи

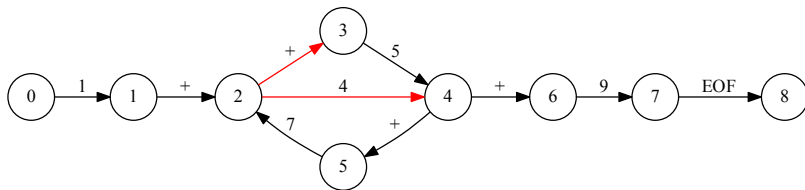
Целью работы является разработка механизма диагностики ошибок в ослабленном синтаксическом анализе регулярной аппроксимации динамически формируемого выражения

Задачи:

- Определить понятие синтаксической ошибки в терминах регулярной аппроксимации динамически формируемого выражения
- Разработать механизм диагностики ошибок в синтаксическом анализе регулярной аппроксимации динамически формируемого выражения
- Доказать корректность механизма
- Реализовать предложенный механизм на базе проекта YaccConstructor
- Провести экспериментальное исследование

Понятие синтаксической ошибки

- На вход вместо строки – граф ДКА
- Вместо лексем – ребра, нагруженные лексемами
- Часть прочитанного входа не префикс строки, а путь в графе из начальной вершины (далее префикс входного графа)
- Префикс входного графа корректен для рассматриваемой грамматики G , если строка, образованная лексемами данного префикса, является корректным префиксом в грамматике G
- Синтаксической ошибкой является ребро, делающее хотя бы один корректный префикс входного графа не корректным

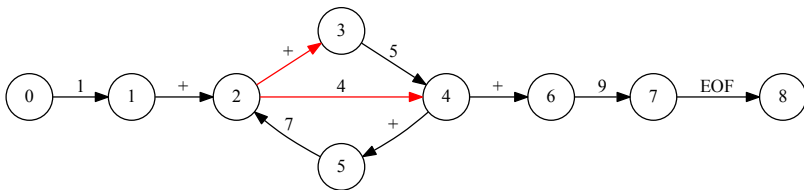


Механизм диагностики ошибок

- Алгоритм построения префиксов — модификация основного синтаксического анализа, дополнительно строящая для каждого GSS состояния все возможные корректные префиксы входного графа, ведущие в данное состояние
- Алгоритм диагностики ошибок — с использованием обхода в глубину по входному графу, сравнивает множество построенных префиксов
- В результате – создаются множества ребер *errors* и *probErrors*, где все ребра из множества *errors* являются ошибочными, а множество $errors \cup probErrors$ является аппроксимацией всех ошибочных ребер. Также для ребра *e* указываются множества префиксов входного графа $errors[e]$ и $probErrors[e]$, где все префиксы из $errors[e]$ являются корректными для ребра *e*, а множество $errors[e] \cup probErrors[e]$ является аппроксимацией всех таких префиксов

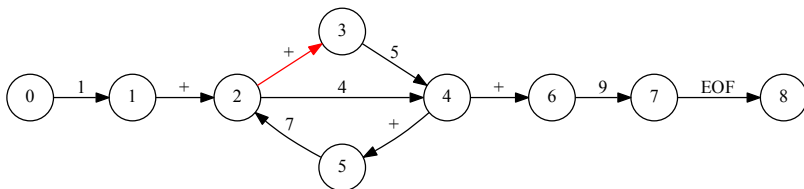
Пример работы алгоритма

- Грамматика:
 $s ::= e$
 $e ::= \text{Num} + e \mid \text{Num}$
- Входной граф ДКА:

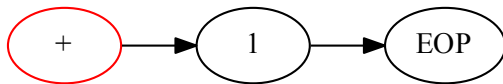


Пример работы алгоритма: префиксы

Синтаксическая ошибка:

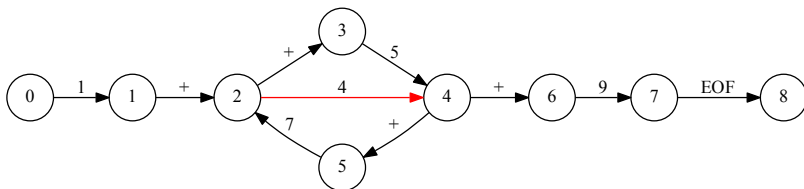


Префиксы:

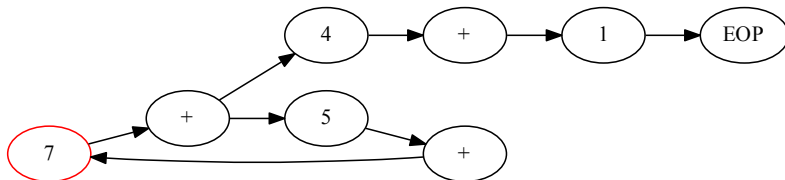


Пример работы алгоритма: префиксы

Синтаксическая ошибка:



Префиксы:



Корректность механизма диагностики ошибок

Теорема (Корректность алгоритма построения префиксов)

Для любого состояния GSS строятся все корректные префиксы входного графа, ведущие в данное состояние, и только они.

Теорема (Корректность алгоритма диагностики ошибок)

Все ребра из построенного множества $errors$ являются ошибочными, а множество $errors \cup probErrors$ является аппроксимацией всех ошибочных ребер. Также для каждого ребра e , все префиксы входного графа из $errors[e]$ являются корректными для ребра e , а множество $errors[e] \cup probErrors[e]$ является аппроксимацией всех таких префиксов

- Определено понятие синтаксической ошибки в терминах регулярной аппроксимации динамически формируемого выражения
- Разработан алгоритм диагностики ошибок в синтаксическом анализе регулярной аппроксимации динамически формируемого выражения
- Выполнена реализация предложенного алгоритма на языке программирования F# в рамках исследовательского проекта YaccConstructor
- Доказана корректность алгоритма
- Проведено экспериментальное исследование