Особенности вычисления семантики встроенных языков

Автор: Иванов Андрей Васильевич
Научный руководитель:к.ф.-м.н. доц. Булычев Д.Ю.
Рецензент: ведущий программист
ООО “ПитерсофтвареХаус“ Полозов В.С.

Санкт-Петербургский государственный университет

27 мая 2016г.

Встроенные языки

- JavaScript в Java

```
String script = "function hello(name) {print'(Hello, ' +
    name);}";
engine.eval(script);
Invocable inv = (Invocable) engine;
inv.invokeFunction("hello", "Scripting!!!");
```

Обзор существующих методов и аналогов

- Статья “Static validation of dynamically generated HTML documents based on abstract parsing and semantic processing”
- Alvor – плагин к Eclipse IDE для проверки встроенного в Java SQL
- IntelliLang – поддержка встроенных языков в IntelliJ IDEA
- PhpStorm – IDE для PHP с поддержкой встроенных языков
- Varis – плагин к Eclipse IDE для поддержки JS и HTML в PHP: подсветка синтаксиса, навигация

- Плохо расширяемы
- Почти отсутствует семантический анализ

Схема анализа встроенного кода

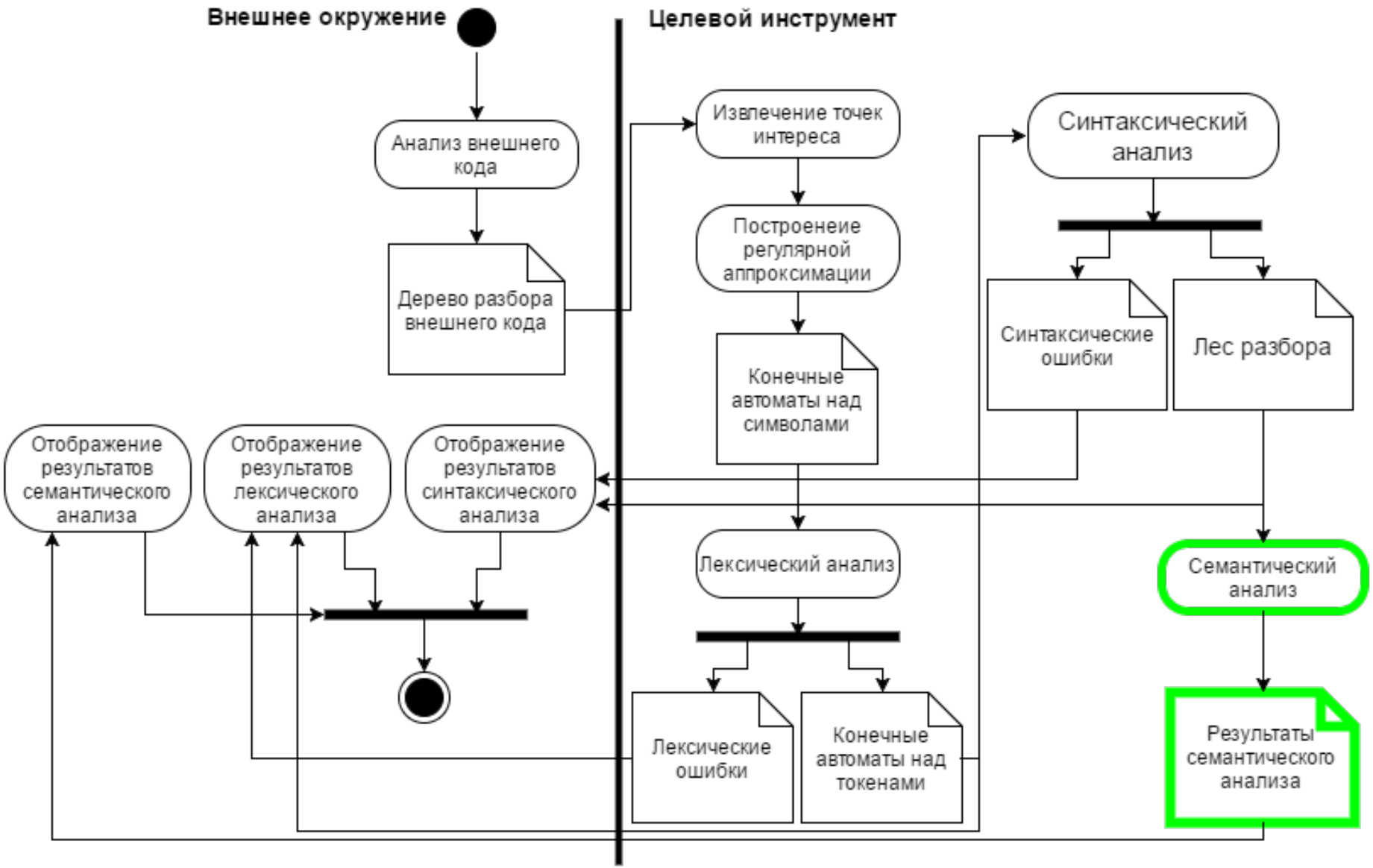
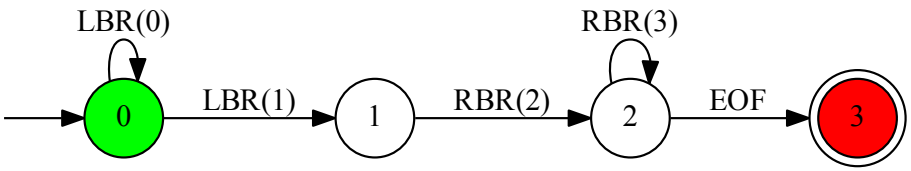


Схема анализа встроенного кода

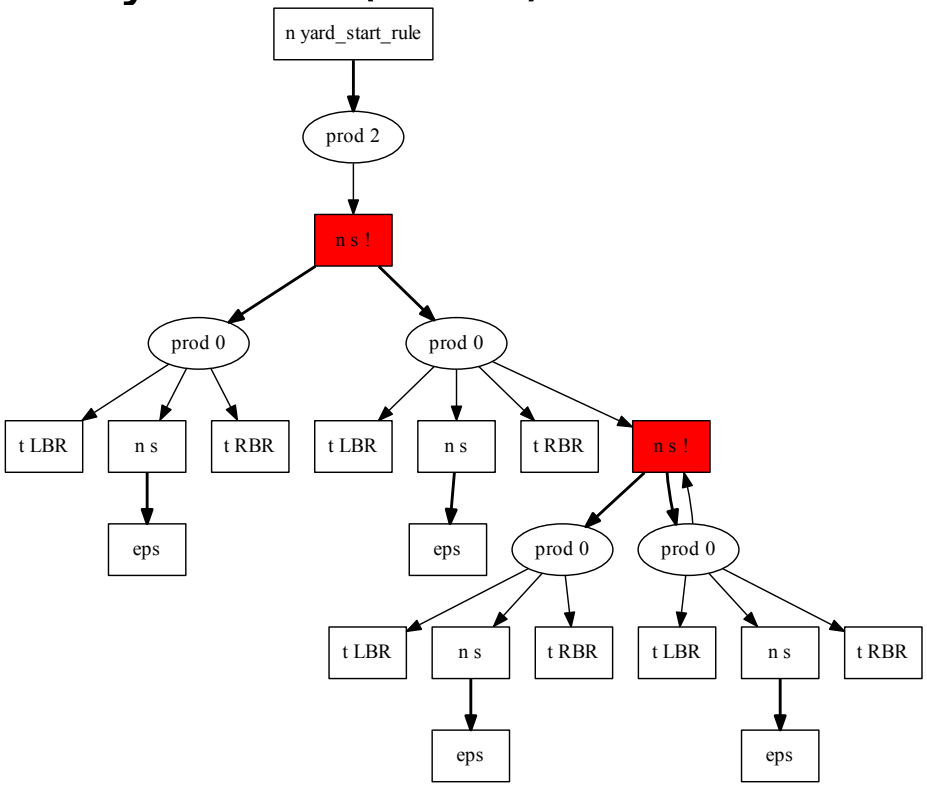
Грамматика:

$$
\begin{array}{lrcl}
(0) & \mathit{start\_rule} & ::= & s \\
(1) & s & ::= & \mathrm{LBR}\ s\ \mathrm{RBR} \\
(2) & s & ::= & \varepsilon
\end{array}
$$

Вход:



Результат (SPPF):

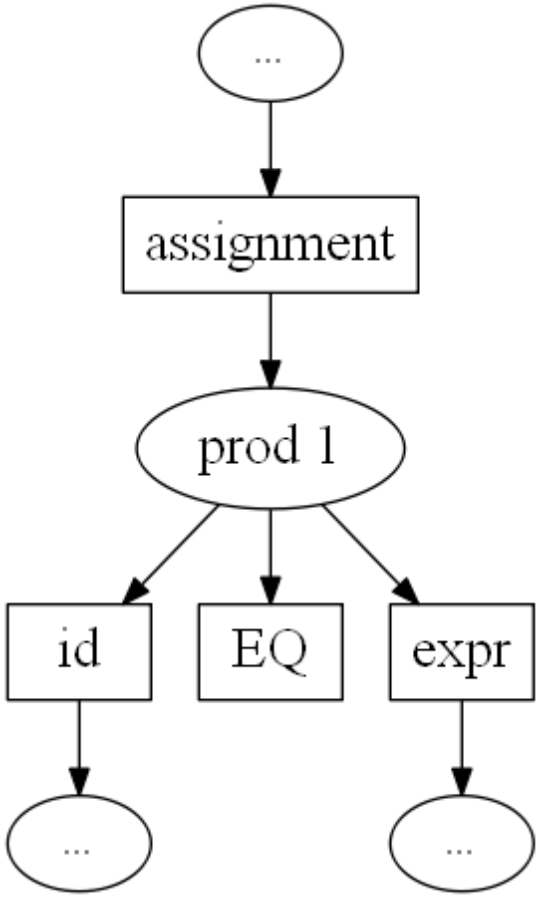


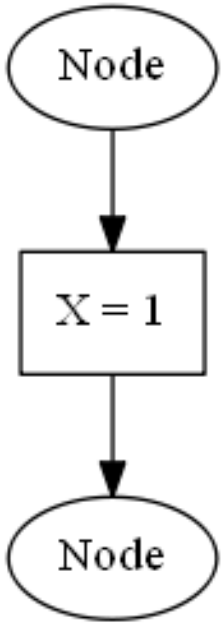
Постановка задачи

- Разработать алгоритм построения графа потока управления для встроенных языков, принимающий на вход лес разбора
- Решить задачу поиска хорошо определённых переменных при анализе встроенных языков
- Реализовать предложенные алгоритмы

Построение графа потока управления

- Обрабатывать лес целиком, без явного извлечения поддеревьев
- Вспомогательные функции-обработчики





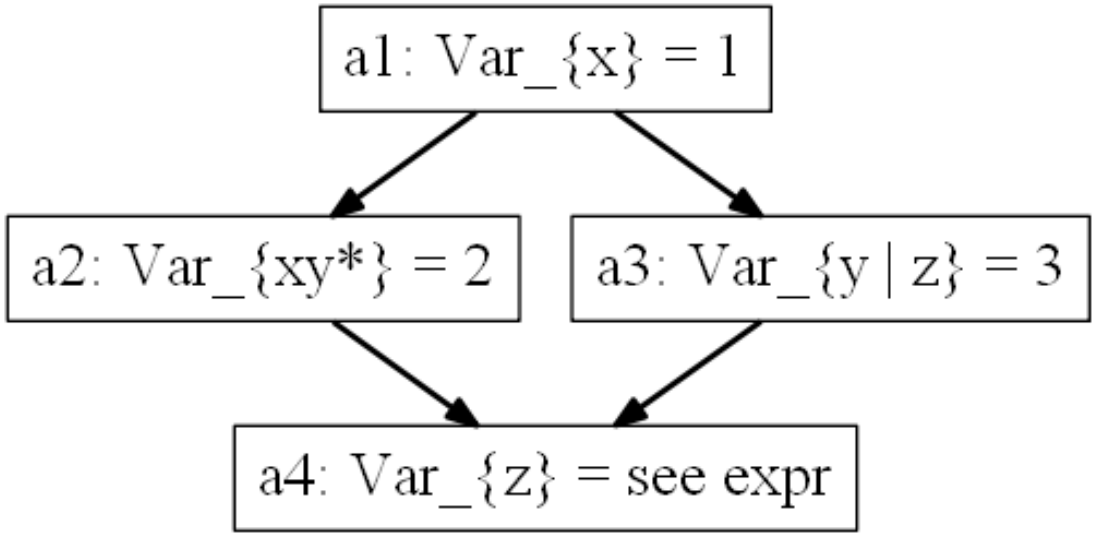
- Обработка неоднозначностей (в том числе и циклов)

Граф потока управления встроенного языка

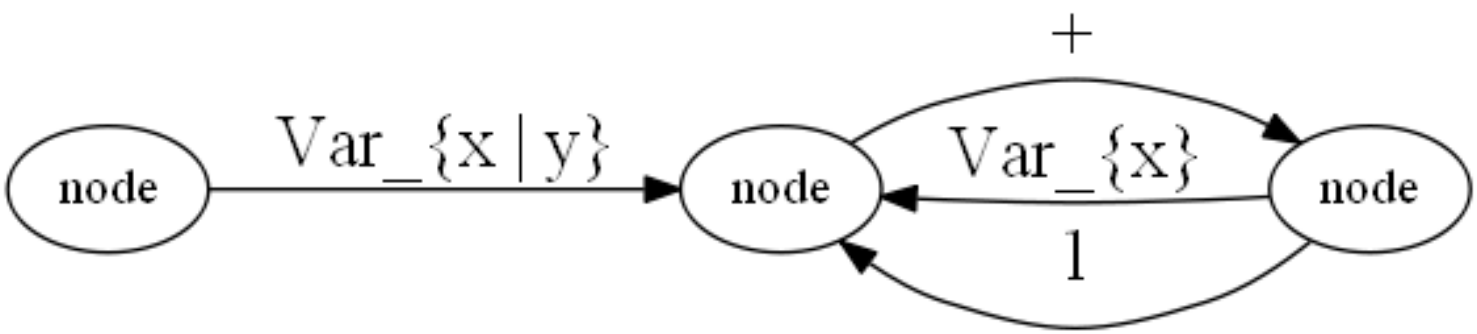
```
str = "x = 1;";
if (...)
{
    str += "x";
    while(...){
        str += "y";
    }
    str += " = 2;"
}
else str += (cond1? "y" : "z") + " = 3;";

str += "z = " + (cond2? "x" : "y");
while(...){
    str += "+";
    if (...) str += "x";
    else str += "1";
}
Eval(str);
```

Граф потока управления встроенного языка

- Граф потока управления



- Граф выражения для последнего блока



Анализ потоков данных

Стадии анализа

- Локальная стадия
  - Учитывает влияние отдельного оператора
  - Предполагается наличие решения *перед* оператором
- Глобальная стадия
  - Выделяет общую часть решений

Полурешётка $(L, \wedge)$: $\forall\, x, y, z \in L$

- Идемпотентность $x \wedge x = x$
- Коммутативность $x \wedge y = y \wedge x$
- Ассоциативность $(x \wedge y) \wedge z = x \wedge (y \wedge z)$

Поиск хорошо определённых переменных

Полурешётка (L, ∇)

- L – множество булевых формул
- Операция ∇

$$\Gamma_1 \nabla \Gamma_2 = \{\Phi_i \in \Gamma_1 : \Gamma_2 \to \Phi_i\} \cup \{\Phi_j \in \Gamma_2 : \Gamma_1 \to \Phi_j\}$$

- $\Gamma \nabla \Gamma = \Gamma$
- $\Gamma_1 \nabla \Gamma_2 = \Gamma_2 \nabla \Gamma_1$
- $(\Gamma_1 \nabla \Gamma_2) \nabla \Gamma_3 = \Gamma_1 \nabla (\Gamma_2 \nabla \Gamma_3)$

Поиск хорошо определённых переменных

- Локальная стадия
  - Известно, что один идентификатор определён, но неизвестно, какой именно
  - Например, $\exists\, w\ (\mathit{IsWord}_{xy*}(w)\ \wedge\ w\ \in\ \mathit{Defined})$
- Глобальная стадия
  - В соответствии с ∇

Поиск хорошо определённых переменных

- Составить формулу Φ для переменной
  - Должны быть определены все идентификаторы
  - Например, $\forall\, w\ (\mathit{IsWord}_{x|y}(w)\ \Rightarrow\ w\ \in\ \mathit{Defined})$
- Вывести $\Gamma\ \rightarrow\ \Phi$
  - Например,

$$\Gamma\ \rightarrow\ \forall\, w\ (\mathit{IsWord}_{x|y}(w)\ \Rightarrow\ w\ \in\ \mathit{Defined})$$

  - Если удалось, то переменная хорошо определена

Результаты

- Разработан алгоритм построения графа потока управления для встроенных языков, принимающий на вход лес разбора
- Решена задача поиска хорошо определённых переменных при анализе встроенных языков
- Выполнена реализация получившихся алгоритмов в рамках исследовательского проекта YaccConstructor
- Результаты работы вошли в статью “On Development of Static Analysis Tools for String-Embedded Languages” (CEE-SECR’15 Proceedings of the 10th Central and Eastern European Software Engineering Conference in Russia)