

О разработке инструментов статического анализа встроенных языков

Хабибуллин Марат

Санкт-Петербургский Академический Университет
НОЦНТ РАН

22 октября 2015г.

```
String sql;  
sql = "SELECT id, age FROM Employees";  
ResultSet rs = stmt.executeQuery(sql);
```



SQL

(Встроенный)



Java

(Основной)

Примеры встраивания языков

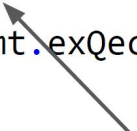
- SQL → Java (JDBC)
- JavaScript → Java (Scripting Engines)
- SQL → PHP (PDO)
- JavaScript, HTML, CSS → PHP
- Генераторы кода

```
String sql;  
sql = "SELETC id,. age FROM Employees";  
ResultSet rs = stmt.executeQuery(sql);
```

Опечатка



Лишний символ



- Поддержка в IDE
 - ▶ Подсветка синтаксиса
 - ▶ Автодополнение
 - ▶ Статический поиск ошибок
- Реинжиниринг
 - ▶ Сбор статистики по коду на встроенном языке
 - ▶ Автоматизированная трансформация

Особенности задачи

```
Entries getEntries() {  
    String usual = "Hello!";  
    String sql = "SELECT * FROM ";  
    if(cond)  
        sql = sql + "Table_1";  
    else  
        sql = sql + "Table_2";  
    return db.exec(sql);  
}
```

Тут нет
кода

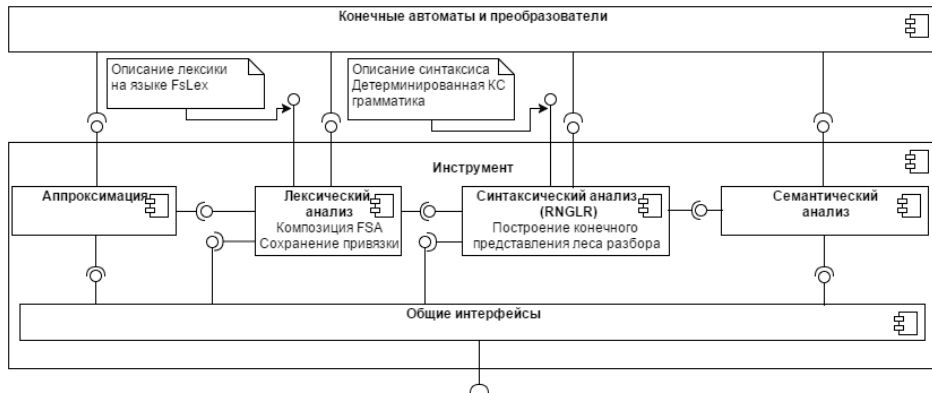
Эта переменная
содержит код

"SELECT * FROM Table_1"
"SELECT * FROM Table_2"

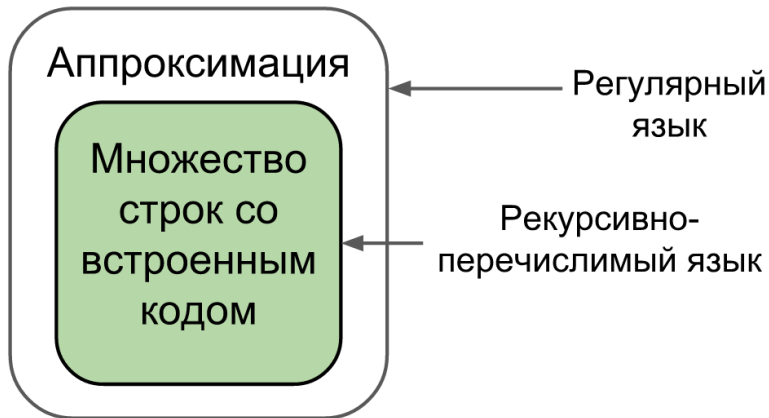
Существующие решения

- Проверка включения языков
 - ▶ Java String Analyzer – регулярная аппроксимация строкового выражения
 - ▶ PHP String Analyzer – контекстно-свободная аппроксимация строкового выражения
- Поддержка встроенных языков в IDE
 - ▶ Varis – плагин к Eclipse IDE для поддержки JS и HTML в PHP: подсветка синтаксиса, навигация
 - ▶ IntelliLang – поддержка встроенных языков в IntelliJ IDEA
 - ▶ PHPStorm – IDE для PHP с поддержкой встроенных языков
 - ▶ Alvor – плагин к Eclipse IDE для проверки встроенного в Java SQL

Архитектура



Аппроксимация



Аппроксимация: поиск строк

```
Entries getEntries() {
```

```
    String usual = "Hello!";
```

```
    String sql = "SELECT * FROM ";
```

```
    if(cond)
```

```
        sql = sql + "Table_1";
```

```
    else
```

```
        sql = sql + "Table_2";
```

```
    return db.exec(sql);
```

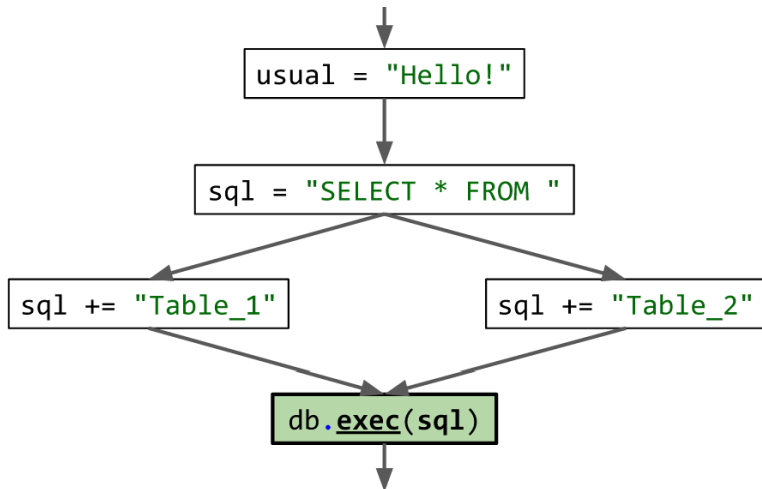
```
}
```

Тут нет кода

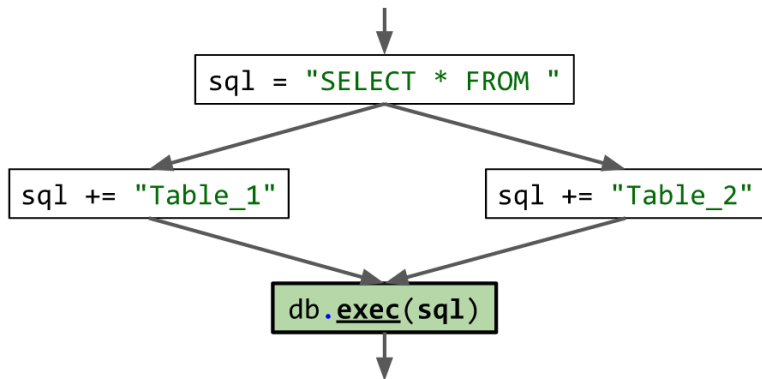
Эта переменная
содержит код

Ищем функции
с таким именем

Аппроксимация: граф потока управления

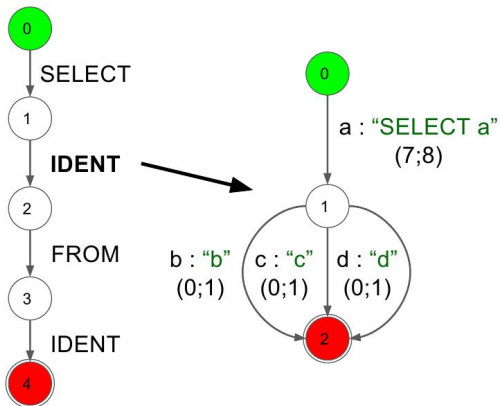
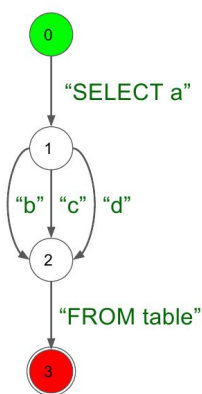


Платформа: построение аппроксимации



Лексический анализ

- Автомат над строками → автомат над токенами
 - Токен – идентификатор + конечный автомат
 - Привязка лексических единиц к исходному коду



Синтаксический анализ

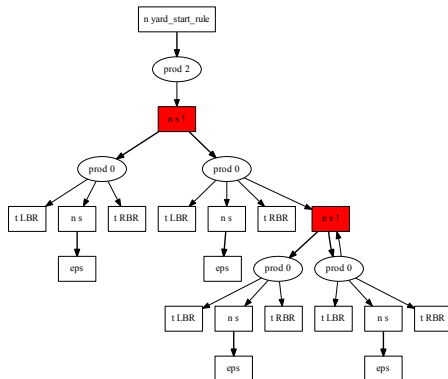
Грамматика:

- (0) $start_rule ::= s$
- (1) $s ::= LBR\ s\ RBR\ s$
- (2) $s ::= \varepsilon$

Вход:



Результат (SPPF):



- YaccConstructor – платформа для исследований в области синтаксического анализа
- Наша платформа – часть YaccConstructor
 - ▶ Генератор абстрактных лексических анализаторов
 - ▶ Генератор абстрактных синтаксических анализаторов
 - ▶ Модульная архитектура для языковых расширений
- Плагин для ReSharper
 - ▶ Расширяемая архитектура, позволяющая легко поддержать любой встроенный язык
 - ▶ Основной язык должен поддерживаться в ReSharper
 - ▶ Анализаторы SQL и Calc, встроенных в C# и JavaScript

Демонстрация

```
8  static int Calculate(bool cond)
9  {
10     var query = "insert into y(u, v)";
11     query += "values(1,2)";
12     Program.ExecuteImmediate(query);
13
14     string expr = "(10";
15     for (int i = 0; i < 10; ++i)
16     {
17         expr += " + 1";
18         if (cond)
19             expr += TrueCaseStr();
20     }
21     return Program.Eval(expr + ") /2");
22 }
23
24 static string TrueCaseStr()
25 {
26     return "*3";
27 }
```



```
17 public static void Execute(bool cond)
18 {
19     string query = "varX = 1;";
20     if (cond)
21         query += "varY = 2;";
22     query += "varZ = varX + varY;";
23     Program.ExtEval(query);
24 }
```

- Контакты:
 - ▶ Хабибуллин Марат: maratx387@gmail.com
 - ▶ Иванов Андрей: ivanovandrew2004@gmail.com
 - ▶ Григорьев Семён: Semen.Grigorev@jetbrains.com
- Исходный код YaccConstructor: <https://github.com/YaccConstructor>
- Google+ сообщество:
<https://plus.google.com/u/0/communities/102842370317111619055>