

Синтаксический анализ графа с помощью парсер-комбинаторов

Автор: Смолина Софья Константиновна
Соавтор: Вербицкая Екатерина Андреевна

Санкт-Петербургский государственный электротехнический университет

4 апреля 2017г.

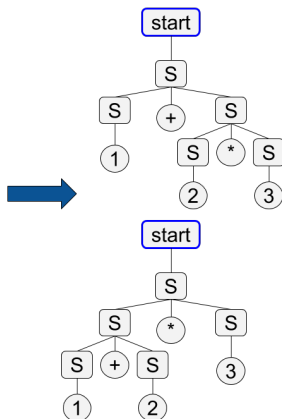
Синтаксический анализ

Синтаксический анализ — это процесс разбора входной последовательности в соответствии с некоторой формальной грамматикой

Грамматика:

$start ::= s$
 $s ::= s + s$
 $s ::= s * s$
 $s ::= num$

Вход: 1+2*3



Синтаксический анализ графов

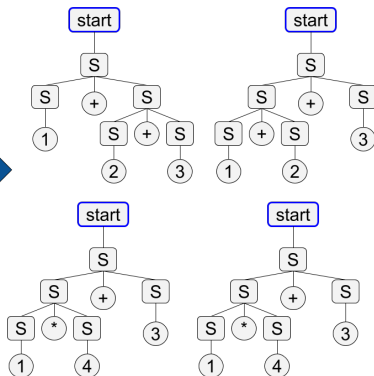
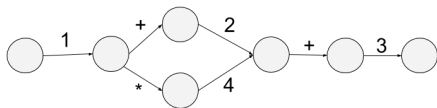
Грамматика:

$start ::= s$
 $s ::= s + s$
 $s ::= s * s$
 $s ::= num$

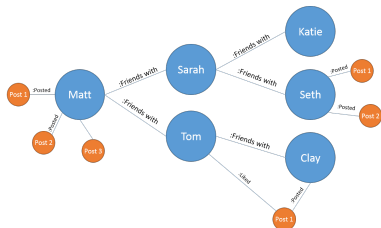
Результат (лес разбора)



Вход:



Синтаксический анализ графов. Применение



Существующие решения. Проблема

- **Решение на основе GLL**

Context-Free Path Querying with Structural Representation of Result, Semyon Grigorev, Anastasiya Ragozina, 2017

- **Решение на основе RNGLR**

Relaxed Parsing of Regular Approximations of String-Embedded Languages, Ekaterina Verbitskaia, Semyon Grigorev, Dmitry Avdyukhin, 2016

- **Решение на основе алгоритма Эрли**

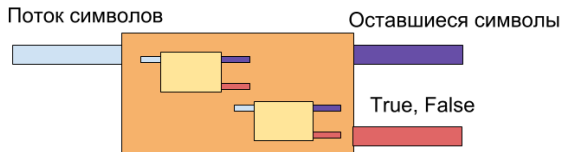
Subgraph Queries by Context-free Grammars, Petteri Sevon and Lauri Eronen, 2008

Проблема: невозможно писать запросы непосредственно в коде программы.

Парсер-комбинаторы

Парсер-комбинатор — это функция высшего порядка, которая на вход получает множество парсеров и возвращает новый парсер.

Парсер



Библиотека парсер-комбинаторов

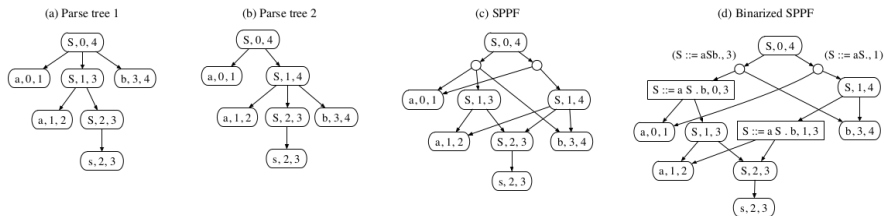
- Анализатор терминального символа
- Анализаатор пустой строки
- Комбинатор последовательности
- Комбинатор правила

Пример парсера

val E: **Nonterminal**

```
= syn( E ~~ "+"~~ E
      | E ~~ "*"~~ E
      | Num )
```

Результат работы (SPPF)



Постановка цели и задачи

Цель работы: Реализовать механизм для синтаксического анализа графов на основе техники парсер-комбинаторов

Задачи:

- Модифицировать библиотеку Meerkat для реализации синтаксического анализа графов
- Протестировать решение

Требование:

- Анализатор конструирует лес разбора
- Анализатор формируется в коде программы

Ключевые изменения:

- Изменен входной тип данных на граф
- Внесены изменения в базовый анализатор для терминала
- Интеграция с графовой базой данных Neo4j

Позиция в строке \Rightarrow Вершина графа

Следующий символ \Rightarrow Множество символов на исходящих рёбрах вершины

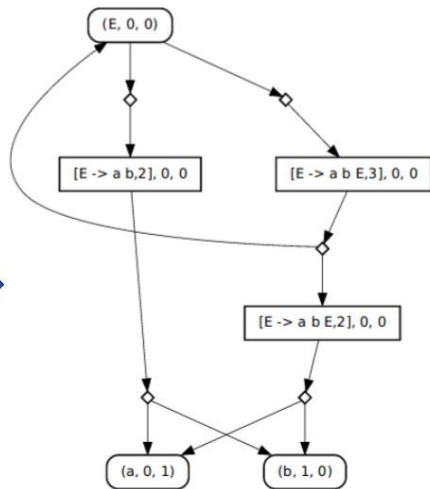
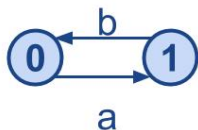
Синтаксический разбор продолжается по тому пути, который начинается с ребра с соответствующим символом

Пример работы

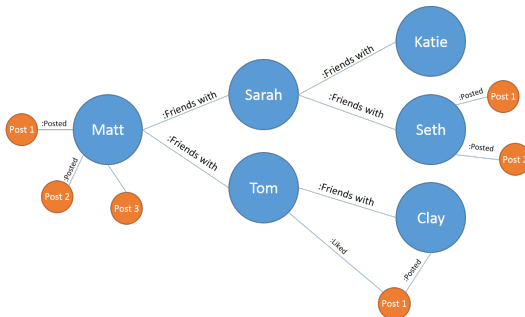
Грамматика:

$E \rightarrow a b E \mid a b$

Граф:



Интеграция с графовой базой данных Neo4j



Что было сделано?

- Внесены изменения в библиотеку
- Библиотрека интегрирована с графовой базой данных
- Протестированно решение

Результат:

- Получен синтаксический анализатор графов

Планируется:

- Реализовать семантические действия