



Библиотека YC.QuickGraph: поиск путей с КС-ограничениями в графах

Автор: Свитков Сергей Андреевич, 344 группа
Научный руководитель: ст. пр., к. ф-м н. Григорьев С.В.

Санкт-Петербургский государственный университет
Кафедра системного программирования

27 апреля 2017г.

- Ориентированные графы с метками на рёбрах
 - ▶ Графовые базы данных
 - ▶ Социальные графы
 - ▶ Биоинформатика
 - ▶ ...
- Пути в подобных графах могут представлять интерес
 - ▶ Последовательность из меток на рёбрах
- КС-грамматики
 - ▶ Позволяют задать КС-язык
 - ▶ КС-язык можно использовать как язык запросов
- Поиск интересующих строчек в графе
 - ▶ Последовательность меток на ребрах в графе выводима в заданном языке

- Большинство существующих языков запросов к графам — регулярные
 - ▶ Cypher (Neo4J)
 - ▶ Gremlin (Titan)
- Контекстно-свободные языки запросов
 - ▶ Большое количество теории
 - ▶ Существуют реализации, но с очень ограниченной функциональностью

Conjunctive Context-Free Path Queries

- 2014, Jelle Hellings
- Обобщение существующего регулярного языка CRPQ до КС-языка CCFPQ
- Позволяет использовать КС-языки для запросов
- Использует СУК для синтаксического анализа графов
- Результат — КС-отношение $R = (N, n, m)$, где N — нетерминал, из которого выводим путь из вершины n в m
- Минусы
 - ▶ Нет практической реализации
 - ▶ Представление результата только в одном формате

Subgraph Queries by Context-free Grammars

- 2008, Petteri Sevon and Lauri Eronen
- Применение КС-запросов в биоинформатике
- Для синтаксического анализа используется Earley Parser
- Результат — связный подграф, порожденный множеством путей, строки из меток на ребрах которых выводимы из заданной грамматики
- Минусы
 - ▶ Алгоритм имеет приемлемое время работы только на небольших входных данных
 - ▶ Проблемы при наличии циклов в графе
 - ▶ Предложенный алгоритм представляет результат только в одном формате

Ослабленный синтаксический анализ динамически формируемых выражений на основе алгоритма GLL

- 2016, Рагозина А.К.
- Синтаксический анализ регулярной аппроксимации (конечный автомат, граф), основанный на GLL
- Обработывает большие входные данные
- Работает с графами, рёбра которых помечены токенами — численными значениями
- Результат — лес разбора всех путей в графе, выводимых в заданном языке (SPPF)
- Реализован в рамках YaccConstructor (.NET)

- Существующие библиотеки для работы с графами для .NET
 - ▶ QuickGraph — заброшен в 2011
 - ▶ YC.QuickGraph — основывается на QuickGraph, поддерживается сейчас

- Существенные минусы у существующих практических реализаций (один формат результата, слишком наивные алгоритмы, ...)
- Хочется представить реализацию, не имеющую перечисленных недостатков
- Представление результата запроса должно быть возможно в нескольких форматах:
 - ▶ Кратчайший путь
 - ▶ Подграф
 - ▶ Множество путей
 - ▶ КС-отношение
- Можно использовать результаты работы Рагозиной А. К. для синтаксического анализа графов, если преобразовать пользовательские объекты на рёбрах графа в токены

Постановка задачи

Целью работы является создание библиотеки, позволяющей выполнять КС-запросы к графам и предоставлять пользователю результат запроса в различных форматах

Задачи:

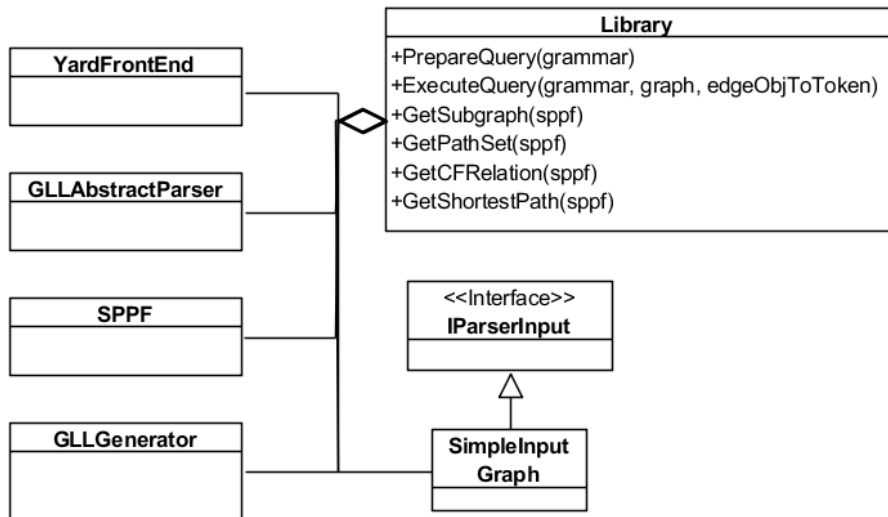
- Разработать архитектуру решения
- Реализовать библиотеку
- Провести тестирование библиотеки
- Представить конечный результат в виде NuGet-пакета для использования в языках платформы .NET

На данный момент достигнуты следующие результаты

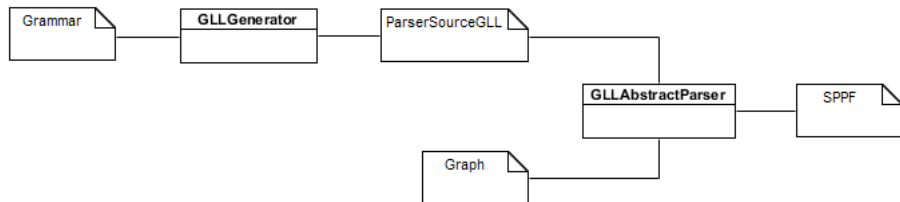
- Изучена предметная область
- Спроектирована архитектура решения
- Реализовано
 - ▶ Поддержка задания грамматики строкой
 - ▶ Возможность задания графа с пользовательскими объектами на рёбрах
 - ▶ Поддержка преобразования пользовательского объекта к токену
 - ▶ Начальный прототип: позволяет задать запрос, граф и получить результат — SPPF

Архитектура решения

Упрощенная архитектура решения:



Процесс исполнения запроса:



Дальнейшие планы

- Реализовать преобразования SPPF к различным форматам
- Протестировать полученное решение
- Написать документацию, собрать NuGet-пакет