

Санкт-Петербургский государственный университет

Математическое обеспечение и администрирование  
информационных систем

Системное Программирование

Азимов Рустам Шухратулович

# Обзор задач синтаксического анализа графов

Курсовая работа

Научный руководитель:  
к. ф.-м. н., ст. преп. Григорьев С. В.

Санкт-Петербург  
2016

SAINT-PETERSBURG STATE UNIVERSITY

Software and Administration of Information Systems

Software Engineering

Rustam Azimov

# Survey of graph parsing problems

Course Work

Scientific supervisor:  
Senior Lecturer Semen Grigorev

Saint-Petersburg  
2016

# Оглавление

|   |   |
|---|---|
| Введение  | 4 |
| 1. Постановка задачи  | 5 |
| 2. Общая таблица для классов задач синтаксического анализа графов | 6 |
| Список литературы   | 7 |

# Введение

Во многих областях возникают задачи поиска путей в графах, удовлетворяющих определенным условиям. Например, на искомые пути могут быть наложены ограничения на длину, или производится поиск лишь простых путей. Но при работе со сложными системами зачастую таких ограничений бывает недостаточно. Поэтому широко распространено использование ограничений на метки ребер/вершин путей помеченного графа. Для заданного алфавита  $\Sigma$  и ориентированного графа  $G$ , ребра которого помечены символами из  $\Sigma$ , естественно выбрать в качестве ограничений на пути — формальный язык  $L$ , где для искомого пути  $p$  графа  $G$  выполняется  $l(p) \in L$ . Здесь  $l(p)$  означает слово  $w \in \Sigma^*$ , полученное последовательной конкатенацией меток пути  $p$ . Задачи поиска путей в графе, которые используют такие ограничения с формальными языками, являются задачами синтаксического анализа графов.

Задачи с различным сочетанием описанных ограничений используются в таких областях, как графовые базы данных [2], биоинформатика [1] и др. Кроме того, в контексте задачи бывает необходимо отвечать на различного рода вопросы, связанные с искомыми в графе путями. Если рассматривать процесс поиска ответа на такие вопросы, как вычисление запроса для графовой структуры данных, то типы вопросов, на которые отвечает задача принято называть семантикой запроса. Например, в задачах, использующих реляционную семантику запроса, для каждой пары вершин  $m, n$  графа  $G$  и нетерминала  $N$  входной грамматики необходимо ответить, существует ли путь  $p$  из вершины  $m$  в вершину  $n$ , такой что  $l(p)$  принадлежит языку, порожденному нетерминалом  $N$ .

В качестве формальных грамматик, описывающих язык в задачах синтаксического анализа графа, широко используются регулярные и контекстно-свободные грамматики. Но существуют грамматики, обладающие большей выразительной мощностью. Например, конъюнктивные грамматики [3], которые уже нашли свое применение в биоинформатике при работе с РНК [4]. Чтобы исследовать применимость ранее не используемых грамматик в задачах синтаксического анализа графов, необходимо структурировать и привести к общей терминологии существующие результаты этой области.

В данной работе будет произведен обзор существующих результатов для различных вариаций задачи синтаксического анализа графов.

# 1. Постановка задачи

В рамках данной работы необходимо выполнить обзор существующих результатов для различных задач синтаксического анализа графов, для чего требуется решить перечисленные ниже задачи.

- Изучить литературу, посвящённую решению задач синтаксического анализа графов.
- Выявить основные классы задач синтаксического анализа графов и их ключевые характеристики.
- Составить общую таблицу, отражающую связь классов задач и характеристик.

## 2. Общая таблица для классов задач синтаксического анализа графов

В данном разделе приведена таблица, содержащая существующие результаты для различных классов задач в области синтаксического анализа графов. Данные результаты были приведены к общей терминологии для упрощения таблицы.

Рассматриваемые классы задач отличаются:

- семантикой запроса (поиск кратчайшего пути, простого пути и т.д.);
- типом входного графа (без циклов, планарный и т.д.);
- классом входной формальной грамматики (регулярная грамматика, контекстно-свободная и т.д.).

| Query semantic | Type of graph $G = (V, E)$ | Regular Grammar R       | CFG $C = (\Sigma, N, P, S)$ | Conjunctive grammar | CSG             |
|----------------|----------------------------|-------------------------|-----------------------------|---------------------|-----------------|
| shortest path  | general graph              | $O( R  G \log( R  G ))$ | FP                          |                     | undecidable     |
| simple path    | directed chain             | $O( V )$                | $O( V ^3)$                  |                     | PSPACE-complete |
|                | treewidth bounded          | FP                      | NP-complete                 |                     | PSPACE-complete |
|                | planar/grid graph          | NP-complete             | NP-complete                 |                     | PSPACE-complete |
| relational     | general graph              |                         | $O( N  E  + ( N  V )^3)$    | undecidable         |                 |
| subgraph       | acyclic graph              |                         | $O( V ^3 m^3)$              |                     |                 |
| single-path    | general graph              |                         | $O(f)$                      |                     |                 |
| all-path       | general graph              |                         | $O( N  E  + ( N  V )^3)$    |                     |                 |

Таблица 1: Общая таблица, отражающая связь классов задач синтаксического анализа графов и их характеристик. *CFG* и *CSG* означают контекстно-свободную и контекстно-зависимую грамматики соответственно. *FP* означает, что искомым путь может быть найден на детерминированной машине Тьюринга за полиномиальное время, даже если спецификация формального языка является частью входных данных. *m* означает максимальную длину пути в графе, а  $f = |N||V|^2((|N||V|^2)\log(|N||V|^2) + |P||V|^3 + \min(|N|, |P|)|E|) + 2^{|N||V|^2-1}$ .

## Список литературы

- [1] Anderson J., Novák Á., Sükösd Z. Quantifying variances in comparative RNA secondary structure prediction // BMC Bioinformatics. — 2013. — P. 14–149.
- [2] Mendelzon A., Wood P. Finding Regular Simple Paths in Graph Databases // SIAM J. Computing. — 1995. — Vol. 24, no. 6. — P. 1235–1258.
- [3] Okhotin A. Conjunctive and Boolean grammars: the true general case of the context-free grammars // Computer Science Review. — 2013. — Vol. 9. — P. 27–59.
- [4] Zier-Vogel R. Predicting RNA secondary structure using a stochastic conjunctive grammar : Master's thesis / R. Zier-Vogel ; University of Manitoba. — 2012.