



Модернизация набора данных CFPQ_DATA

Автор: Вадим Игоревич Абзалов, 18.Б11-мм

Научный руководитель: к. ф.-м. н., доцент кафедры информатики СПбГУ С. В.
Григорьев

Санкт-Петербургский государственный университет
Кафедра системного программирования

12 июня 2021 г.



Задача поиска путей с контекстно-свободными ограничениями

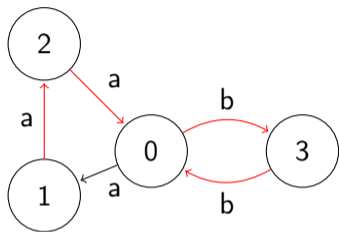


Рис. 1. Граф

$$S \rightarrow a S b$$
$$S \rightarrow a b$$

Рис. 2. Грамматика

CFPQ алгоритмы

- **Граф:** на рёбрах метки
- **Запрос:** контекстно-свободная грамматика
- **Путь в графе:** слово из языка этой грамматики
- **Задача:** искать такие пути

Области применения

- Базы данных
- Биоинформатика
- Статический анализ кода

Экспериментальное исследование CFPQ алгоритмов

- Проводить на реальных данных
- Определять производительность
- Сравнить с уже существующими решениями

⚠ Нет единого набора данных

Проект CFPQ_Data

- (1/2) Множество различных реальных и синтетических графов
- (2/2) Вспомогательные инструменты

⚠ Имеет ряд технических ограничений

Требования к проекту CFPQ_DATA

Проблемы

- ? Набор данных может быть загружен только целиком
- ? Отсутствуют инструменты манипулирования графами
- ? Отсутствует информация о структуре графов, включенных в набор

Требования

- ✓ Предоставить возможность загрузки конкретного графа из набора
- ✓ Предоставить инструменты для манипулирования графами
- ✓ Предоставить информацию о структуре графов, включенных в набор

📌 Цель и задачи

Цель: модернизация существующего набора данных CFPQ_DATA для создания унифицированного средства подготовки проведения экспериментального исследования CFPQ алгоритмов

Задачи

- Модернизация архитектуры набора данных
- Добавление новых возможностей работы с данными
 - Загрузка конкретных графов из набора данных
 - Преобразование графов в другие форматы
 - Получение информации о графе
 - Трансформация графов
- Публикация Python пакета для работы с набором данных и документации к нему

Набор графовых данных	Описание	Тип	Дата	#Вершин	#Рёбер
SNAP: Stanford Network Analysis Project	✓	✓		✓	✓
The webgraph framework I: compression techniques	✓		✓	✓	✓
Layered Label Propagation: A MultiResolution Coordinate-Free Ordering for Compressing Social Networks	✓		✓	✓	✓
Требования к CFPQ_Data	✓			✓	✓

Модернизация архитектуры

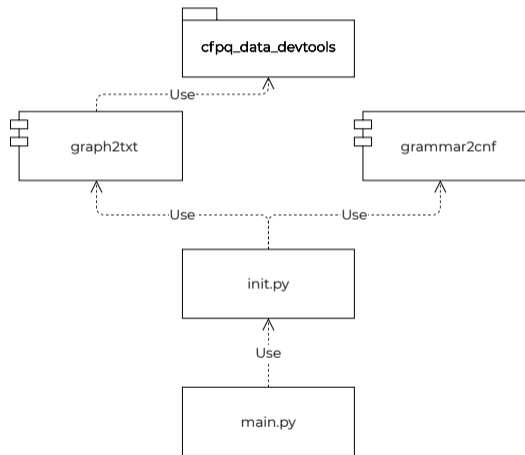


Рис. 3. Архитектура «до»

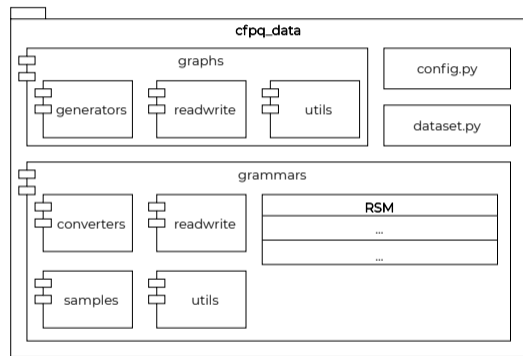


Рис. 4. Архитектура «после»

</> Добавление новых возможностей

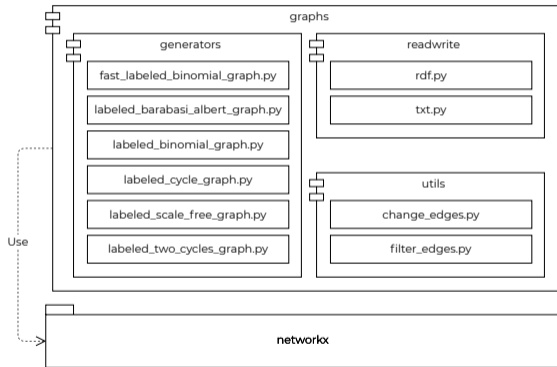


Рис. 5. Модуль работы с графами

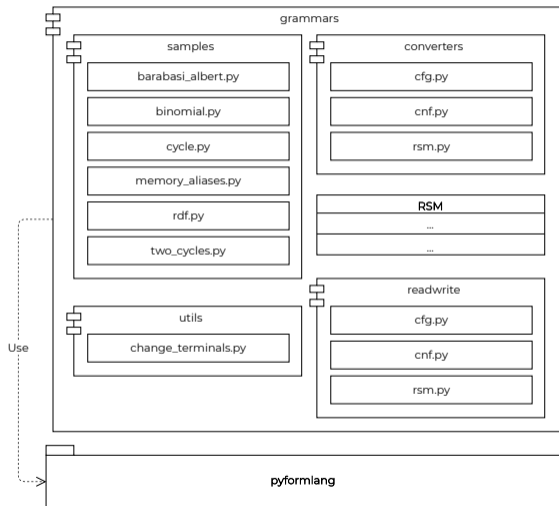


Рис. 6. Модуль работы с грамматиками

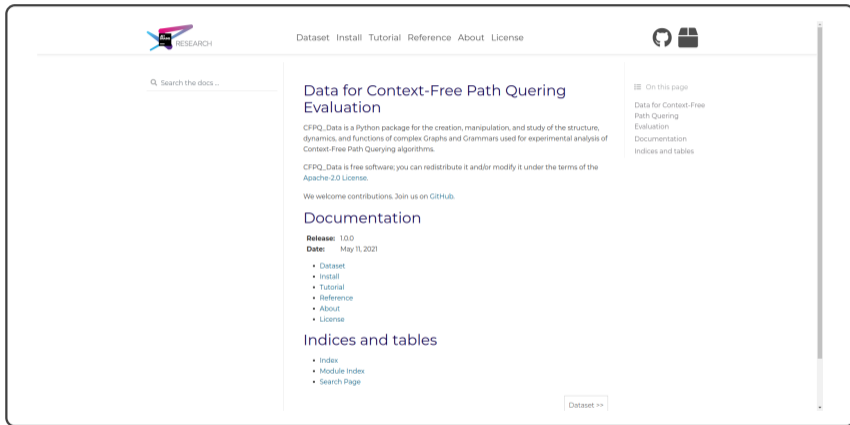


Рис. 8. Веб-сайт проекта

- Информация о наборе данных
- Руководство по установке
- Обучающее руководство
- Документация

☑ Результаты

- ✓ Модернизирована архитектура набора данных
- ✓ Добавлены новые возможности работы с данными
 - ✓ Загрузка конкретных графов из набора данных
 - ✓ Преобразование графов в другие форматы
 - ✓ Получение информации о графе
 - ✓ Трансформация графов
- ✓ Опубликован Python пакет¹ для работы с набором данных и документация к нему²

¹Python пакет: <https://pypi.org/project/cfpq-data>, дата последнего доступа — 04.06.2021

²Веб-сайт с документацией: https://jetbrains-research.github.io/CFPQ_Data/, дата последнего доступа — 04.06.2021