



# Создание набора данных для экспериментального анализа алгоритмов вычисления контекстно-свободных запросов

**Автор:** Азат Ришатович Ахметьянов, 18.Б11-мм  
**Научный руководитель:** к.ф.-м.н., доц. С.В. Григорьев

Санкт-Петербургский государственный университет  
Кафедра системного программирования

06 июня 2020г.

- Алгоритмы поиска путей в графах с контекстно-свободными ограничениями
  - ▶ Путь в графе — слово в формальном языке, контекстно-свободная грамматика — спецификация языка
  - ▶ Необходимо найти такие пути в графе, что соответствующее им слово лежит в заданном языке
- Области применения
  - ▶ Биоинформатика
  - ▶ Статический анализ кода
  - ▶ Семантическая паутина

- Важность существования набора данных для экспериментального анализа алгоритмов
  - ▶ Сравнение производительности алгоритмов
  - ▶ Анализ алгоритмов на данных, моделирующих реальные сценарии их применения
- В области КС запросов такого набора данных нет

# Постановка задачи

**Целью** работы является создание репозитория, содержащего данные (КС запросы и графы из различных областей) для экспериментального анализа алгоритмов поиска путей с КС ограничениями

**Задачи:**

- Унифицировать представление данных
- Добавить возможность получать из запросов грамматики в Нормальной Форме Хомского (НФХ)
- Создать сайт проекта с описанием набора данных, индексируемый Google Dataset Search
- Интегрировать набор данных с графовыми БД

- Представление графов в формате RDF
  - ▶ Связи "объект, предикат, субъект"
  - ▶ Поддержка сторонними инструментами
- Улучшенный формат запросов
  - ▶ Синтаксический сахар для разделения частей продукций
  - ▶ Поддержка регулярных выражений в правых частях продукций
  - ▶ Грамматика в выбранном формате:

---

$$\begin{array}{l} S \\ a \ b \\ S \rightarrow (a \ S \ b)^* \end{array}$$

---

- Обработка расширенного формата запросов для представления в НФХ
  - ▶ Построение ДКА из регулярных выражений, вычленение новых продукций грамматики из его переходов
  - ▶ Присоединение полученных продукций к КС грамматике в построении
  - ▶ Преобразование грамматики в НФХ при помощи библиотеки `pyformlang`

- Сайт<sup>1</sup> проекта создан с помощью Github Pages и статического конструктора Jekyll
  - ▶ Метаданные, необходимые для его индексации Google Dataset Search
  - ▶ Описание набора данных

---

<sup>1</sup>[https://jetbrains-research.github.io/CFPQ\\_Data/](https://jetbrains-research.github.io/CFPQ_Data/)

- Добавлен инструмент для загрузки данных в RedisGraph
- Загрузка данных в Neo4j
  - ▶ Предлагается использование плагина Neosemantics



- Унифицировано представление данных
- Добавлен инструмент для получения из представленных запросов КС грамматик в НФХ
- Создан сайт с описанием набора данных, индексируемый Google Dataset Search
- Предоставлена возможность загрузки набора данных в Neo4j, RedisGraph