



Shuffled Languages Parsing

Автор: Горохов Артём В.

Руководитель: Григорьев Семён В.

JetBrains Research, Programming Languages and Tools Lab
Сенкт-Петербургский Государственный Университет

16 декабря 2017

Shuffled Languages

Agent 1

$S := b^*$

Agent 2

$S := b c d$

Agent 3

$S := c^*$

b

b

b

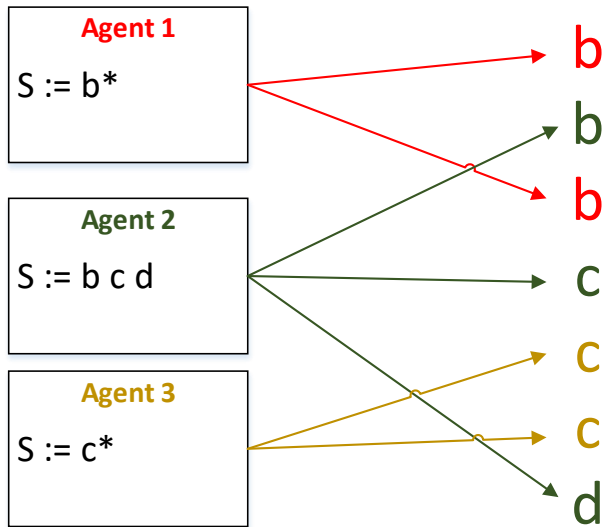
c

c

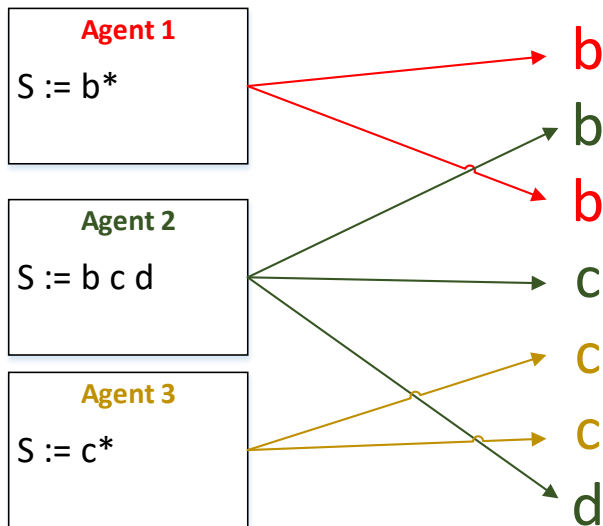
c

d

Shuffled Languages



Shuffled Languages



Проблема: Грамматик может быть очень много, а задача NP-полная

- (Maraist, 2016) Generalized LR parsing and the shuffle operator
- Используется GLR
- Во время разбора отслеживаются позиции сразу во всех грамматиках
- Пространство поиска сужается за счёт поддержания наиболее вероятных веток разбора
- Не строятся деревья разбора

синтаксический анализ графов + метод ветвей и границ

- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе

Input: a b c d

- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе

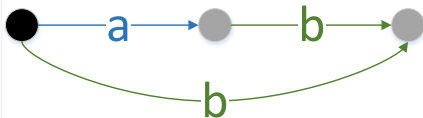
Input: a b c d



Анализ графов

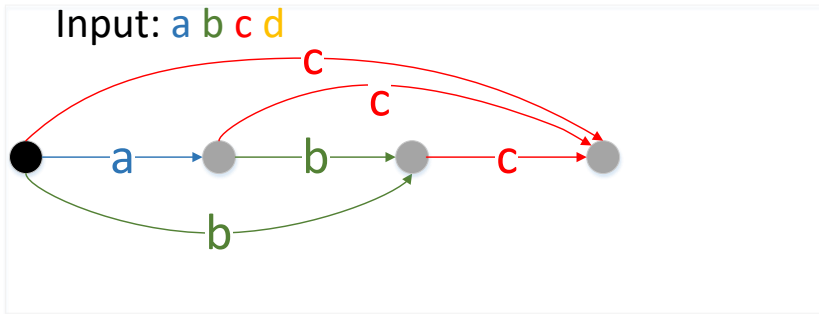
- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе

Input: a b c d



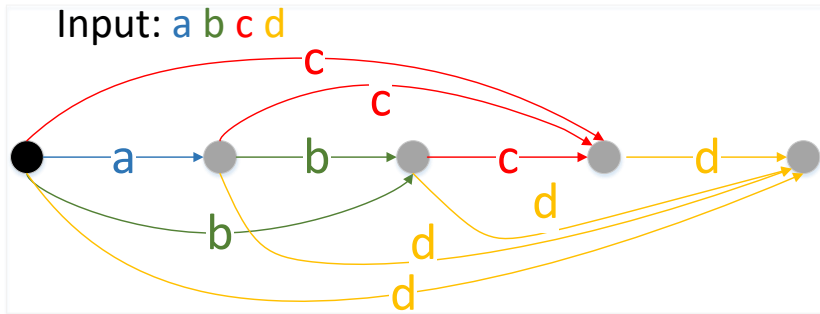
Анализ графов

- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе



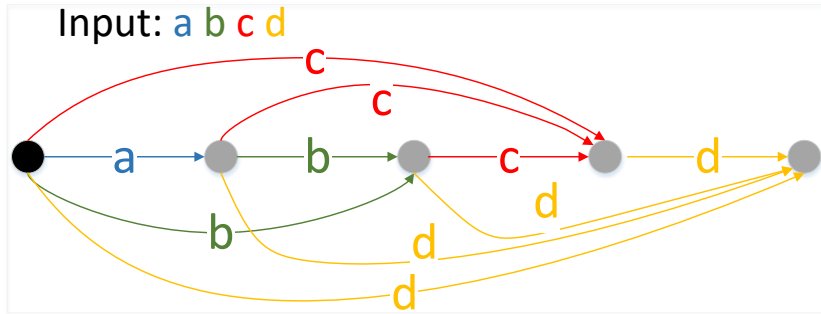
Анализ графов

- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе



Анализ графов

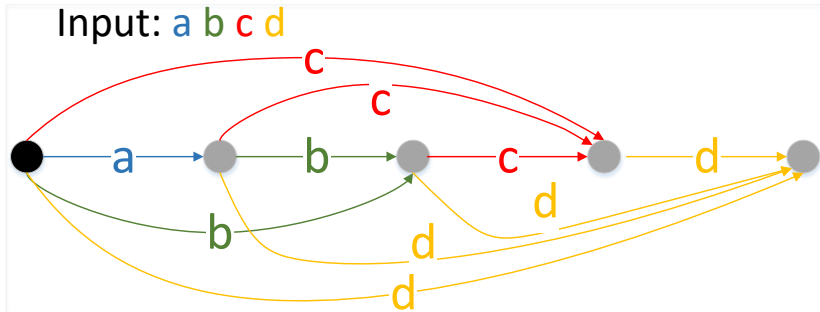
- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе



- Для каждой грамматики анализ производится независимо — можно распараллелить

Анализ графов

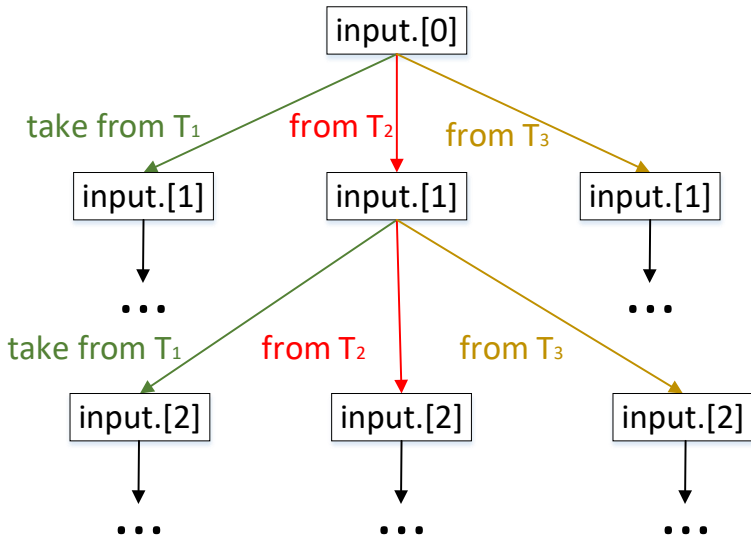
- Выделяем все возможные подпоследовательности из входа и обрабатываем как пути в графе



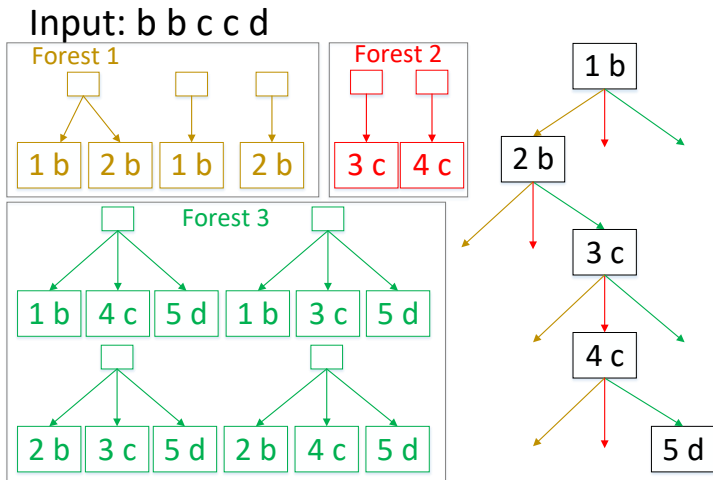
- Для каждой грамматики анализ производится независимо — можно распараллелить
- Результат — лес разбора всех подпоследовательностей для каждой грамматики

Метод ветвей и границ

- Восстанавливаем входную последовательность по лесам разбора



Пример восстановления входа



- Отсечение ветви если невозможно взять токен из данного леса
- Различные ветки обрабатываются независимо — можно

- В поисках данных для адекватного тестирования и сравнения
- Планируется публикация результатов