

Непосредственная поддержка грамматик в расширенной форме Бэкуса-Наура в генераторах синтаксических анализаторов

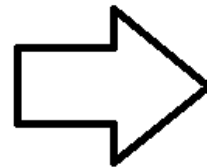
Выполнил: Алефиров А.А., 461 гр.
Научный руководитель: Григорьев С.В.
Рецензент: Беляков А. М.

Санкт-Петербург
2014

Расширенная форма грамматик Бэкуса-Наура (EBNF)

- Регулярные выражения в правых частях правил.
- Более выразительное, понятное и удобное описание грамматик.
- Задача создания парсеров для языков, грамматики которых описаны в EBNF форме.

$S \rightarrow (A|B)(C|D)$



$S \rightarrow AC$

$S \rightarrow AD$

$S \rightarrow BC$

$S \rightarrow BD$

EBNF в генераторах синтаксических анализаторов

- Поддерживается, однако проходит удаление конструкций регулярных выражений.
 - Порождаемые парсеры создают вывод не в терминах грамматики, заданной пользователем.
 - Увеличиваются таблицы состояний и переходов автоматов анализаторов.

Постановка задачи

Цель

Создание генератора синтаксических анализаторов, сохраняющих расширенную форму Бэкуса-Наура.

Задачи

- Разработать алгоритмическое решение.
- Реализовать решение в рамках проекта YaccConstructor.
- Провести апробацию.

Алгоритмы и решения

- LR-анализ.
 - Paul Walton Purdom Jr., Cynthia A. Brown. Parsing extended LR(k) grammars.
 - Модификация состояний и использования стека LR-анализатора.
- GLR, RNLGR, возможность работы с неоднозначными грамматиками.

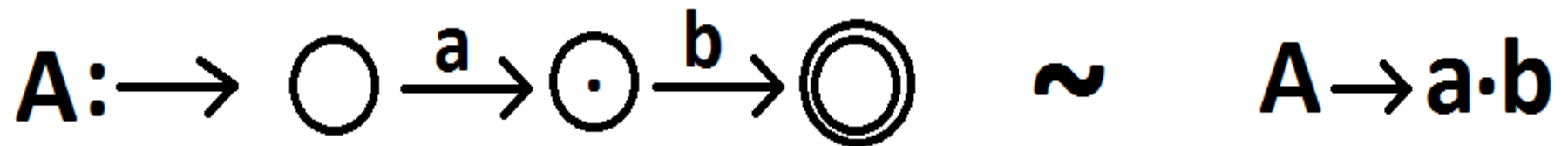
Технологии

- LR
 - Yacc и потомки: Bison, Ocaml yacc и др.
 - Существующая реализация RINGLR в YaccConstructor.
- LL
 - ANTLR.

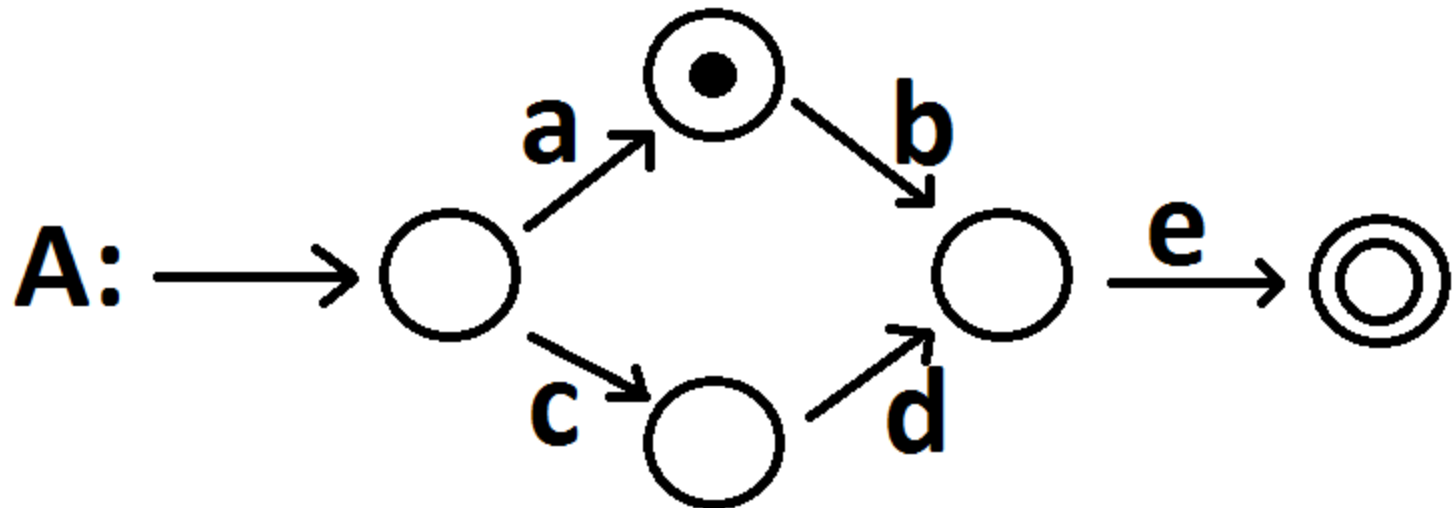
В основном не поддерживают EBNF или трансформируют ее.

Работа с правой регулярной частью продукций

- Создание эквивалентных НКА.
- Состояние автомата в правой части продукции является позицией в нотации Эрли:



$A \rightarrow (a \cdot b \mid cd)e$



Работа со стеком

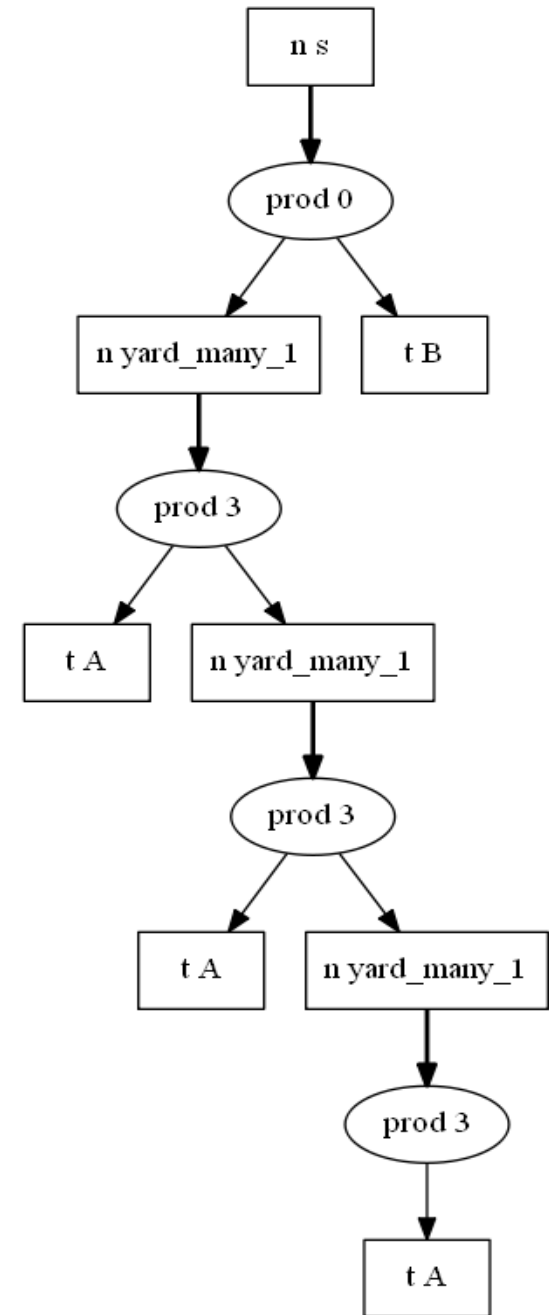
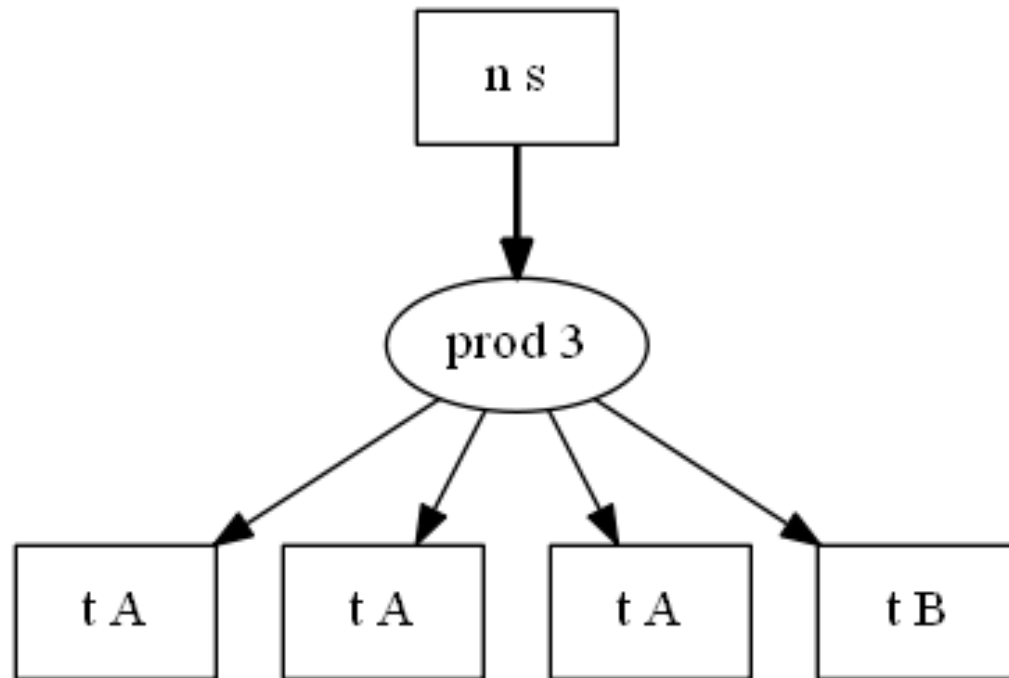
- Правые части правил не имеют фиксированной длины.
- Решение — метки в стеке, управляющие сверткой.
- Возникают конфликты при создании меток
 - оригинальный алгоритм предлагает трансформировать грамматику;
 - разрешаются средствами GLR.

$s : A^*B$

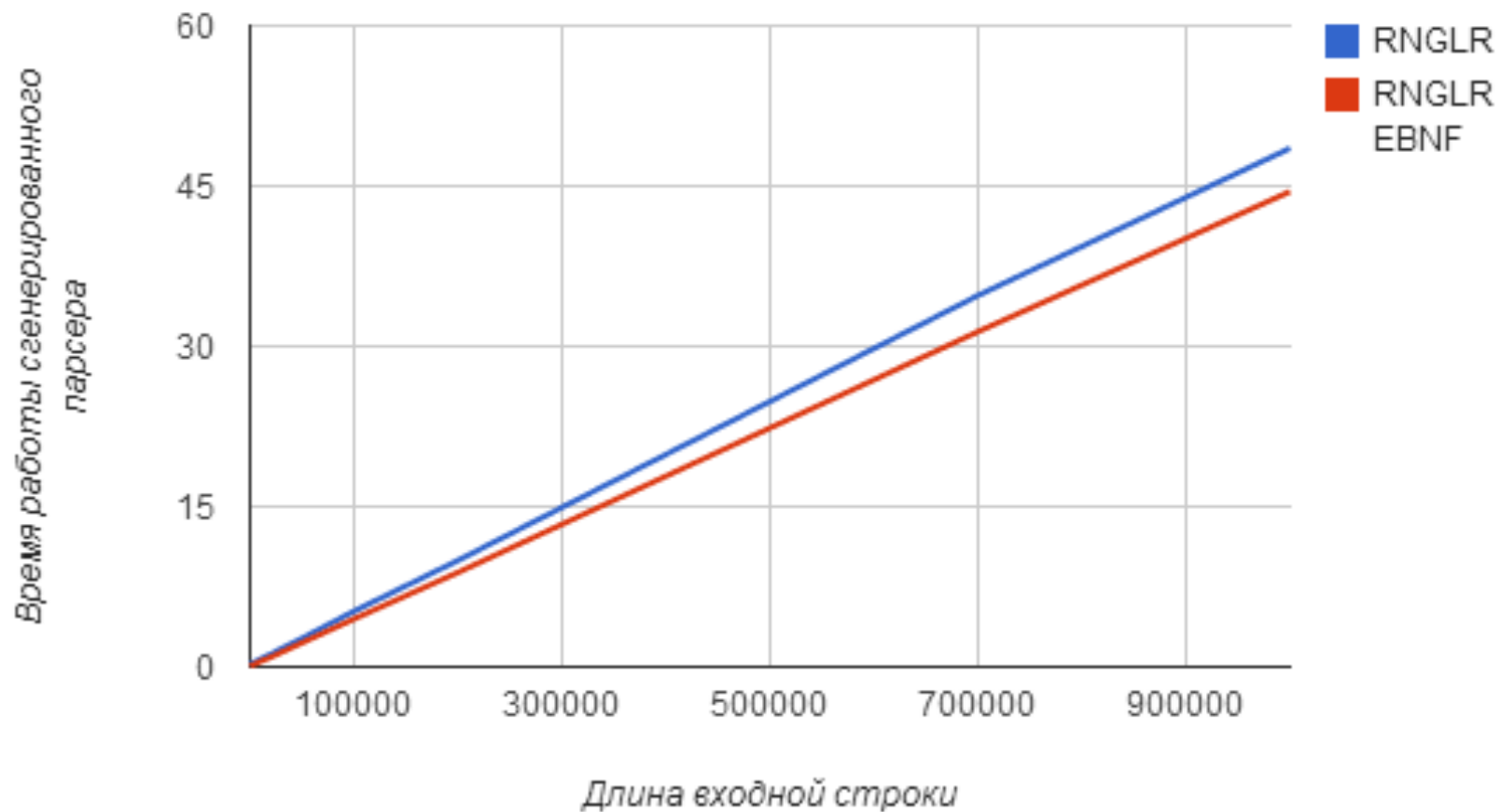
$s : A a C$

$a : A^*$

Вход: A A A B



Апробация RNLGR EBNF на грамматике арифметических выражений



Результаты

- Разработано алгоритмическое решение
 - RNLGR разрешает конфликты стека.
- Реализовано решение в рамках проекта YaccConstructor.
- Проведена апробация.
 - Сохранение EBNF ускоряет синтаксический анализ.