

Ограничения на входную грамматику

ANTLR — генератор нисходящих анализаторов для формальных языков. ANTLR преобразует контекстно-свободную грамматику в программу на языках программирования (Python, C++ и др.).

Правила:

- Структура входного потока определяется набором взаимно ссылающихся правил. Каждое правило имеет имя, необязательный набор аргументов, необязательный пункт «throws», необязательный init-action, необязательный возвращаемое значение и альтернативу или альтернативы. Каждая альтернатива содержит ряд элементов, определяющих, что и где следует сопоставлять.
- Идентификаторы, начинающиеся с заглавной буквы, являются ссылками на токены. Последующие символы могут быть любой буквой, цифрой или символом подчеркивания.
- Идентификаторы, начинающиеся со строчной буквы, являются ссылками на правила парсера ANTLR. Последующие символы могут быть любой буквой, цифрой или символом подчеркивания. Лексические правила не могут ссылаться на правила парсера.
- Имя файла, содержащего грамматику, должно совпадать с именем грамматики и иметь расширение .g4.
- Обязательно наличие заголовка *grammar* <X> в начале файла X.g4.
- Символ конца файла EOF добавляется автоматически в правилах парсера.

ANTLR:

- использует LL(k) парсер, нисходящий анализ;
- поддерживает грамматики с прямой левой рекурсией и записанные в расширенной форме Бэкуса — Наура.