

Справка по Pascal Decimal Parser

| | |
|---|----------|
| Справка по Pascal Decimal Parser | 1 |
| Сведения о синтаксическом анализе | 2 |
| Порождающая грамматика | 2 |
| Классификация Хомского | 3 |
| Конечный автомат-распознаватель | 4 |
| Обнаружение и нейтрализация ошибок | 6 |

Сведения о синтаксическом анализе

Язык

Пусть

$$sign = \{+|- \}$$

$num0$ - целое, возможно, начинающееся с нуля

num - целое, не начинающееся с нуля

Тогда

$$L(G[Z]) = [sign] \{ \{0|num\} [e [sign] num] \mid \{num0 [e [sign] num] \mid e [sign] num\} \}$$

Порождающая грамматика

$$V_T = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ., +, -, /, *\}$$

$$V_N = \{ \langle \text{Цифра } 0..9 \rangle, \langle \text{Цифра } 1..9 \rangle, \langle \text{Знак} \rangle, \langle \text{Экспонента} \rangle, \langle \text{Целое допускающее } 0 \rangle, \\ \langle \text{Целое допускающее } E \rangle, \langle \text{Дробная часть} \rangle, \langle E \rangle, \langle \text{Целое со знаком} \rangle, \langle \text{Целое} \rangle, \\ \langle \text{Десятичная константа допускающая } 0 \rangle, \langle \text{Десятичная константа без знака} \rangle, \\ \langle \text{Десятичная константа} \rangle \}$$

$$Z = \langle \text{Десятичная константа} \rangle$$

$$P = \{$$

$$\langle \text{Цифра } 0..9 \rangle \rightarrow 0|1|2|3|4|5|6|7|8|9$$

$$\langle \text{Цифра } 1..9 \rangle \rightarrow 1|2|3|4|5|6|7|8|9$$

$$\langle \text{Знак} \rangle \rightarrow +|-$$

$$\langle \text{Экспонента} \rangle \rightarrow E|e$$

$$\langle \text{Целое допускающее } 0 \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle | \langle \text{Цифра } 0..9 \rangle \langle \text{Целое} \rangle$$

$$\langle \text{Целое допускающее } E \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle | \langle \text{Цифра } 0..9 \rangle \langle \text{Целое} \rangle$$

$$\langle \text{Целое допускающее } E \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle \langle E \rangle$$

$$\langle \text{Дробная часть} \rangle \rightarrow . \langle \text{Целое допускающее } E \rangle | . \langle E \rangle | \Lambda$$

$$\langle E \rangle \rightarrow \langle \text{Экспонента} \rangle \langle \text{Целое со знаком} \rangle | \langle \text{Экспонента} \rangle \langle \text{Целое} \rangle$$

$$\langle \text{Целое со знаком} \rangle \rightarrow \langle \text{Знак} \rangle \langle \text{Целое} \rangle$$

$$\langle \text{Целое} \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle | \langle \text{Цифра } 1..9 \rangle \langle \text{Целое допускающее } 0 \rangle$$

$\langle \text{Десятичная константа допускающая } 0 \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle \langle \text{Дробная часть} \rangle$
 $\langle \text{Десятичная константа допускающая } 0 \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle \langle E \rangle$
 $\langle \text{Десятичная константа допускающая } 0 \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle \langle \text{Десятичная константа} \rangle$
 $\langle \text{Десятичная константа без знака} \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle \langle \text{Дробная часть} \rangle$
 $\langle \text{Десятичная константа без знака} \rangle \rightarrow \langle \text{Цифра } 1..9 \rangle \langle E \rangle$
 $\langle \text{Десятичная константа без знака} \rangle \rightarrow \langle \text{Цифра } 1..9 \rangle \langle \text{Десятичная константа допускающая } 0 \rangle$
 $\langle \text{Десятичная константа} \rangle \rightarrow \langle \text{Знак} \rangle \langle \text{Десятичная константа без знака} \rangle$
 $\langle \text{Десятичная константа} \rangle \rightarrow \langle \text{Десятичная константа без знака} \rangle$
 }

Классификация Хомского

В классификации Хомского эта грамматика соответствует автоматным грамматикам, поскольку все её продукции могут быть приведены к виду:

$$G[A] : A \rightarrow aB \mid a \mid \Lambda, a \in V_T, A, B \in V_N$$

Значения нетерминалов $\langle \text{Цифра } 0..9 \rangle$, $\langle \text{Цифра } 1..9 \rangle$, $\langle \text{Экспонента} \rangle$ и $\langle \text{Знак} \rangle$ однозначно заменяются любым терминалом из соответствующих им продукций. Продукция соответствующая нетерминалу $\langle \text{Десятичная константа} \rangle$ для случая

$$\langle \text{Десятичная константа} \rangle \rightarrow \langle \text{Десятичная константа без знака} \rangle$$

может быть заменён эквивалентными:

$$\langle \text{Десятичная константа} \rangle \rightarrow \langle \text{Цифра } 0..9 \rangle \langle \text{Дробная часть} \rangle$$

$$\langle \text{Десятичная константа} \rangle \rightarrow \langle \text{Цифра } 1..9 \rangle \langle E \rangle$$

$$\langle \text{Десятичная константа} \rangle \rightarrow \langle \text{Цифра } 1..9 \rangle \langle \text{Десятичная константа допускающая } 0 \rangle,$$

которые приводятся к виду продукций автоматной грамматики, в соответствии с общим случаем.

Путём вышеописанных преобразований мы получаем множество продукций грамматики десятичных констант языка PASCAL, соответствующих виду продукций автоматной грамматики, а следовательно и сама грамматика является автоматной.

Конечный автомат-распознаватель

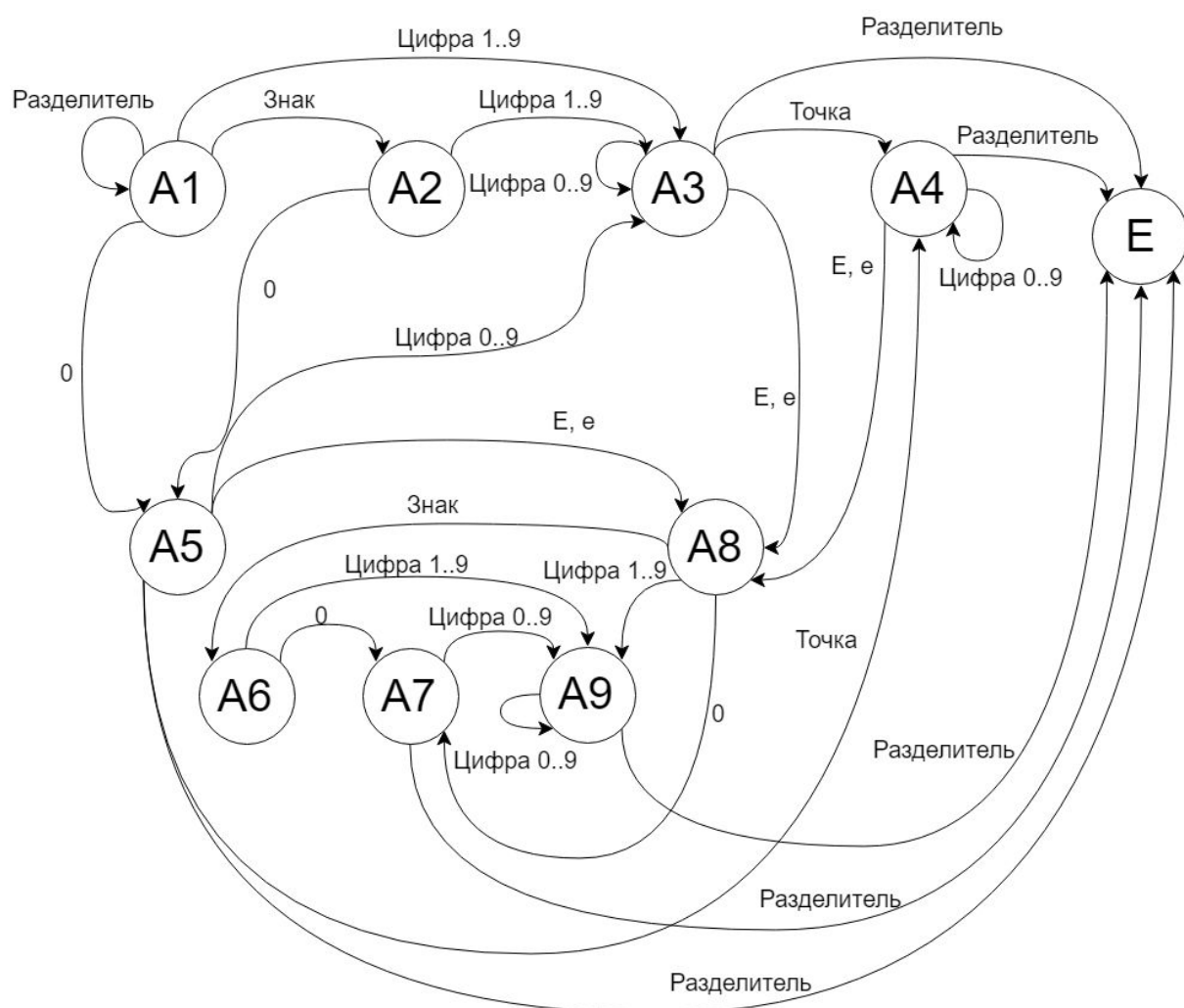


Таблица 1. Состояния конечного автомата

| Обозначение | Описание |
|-------------|--|
| A1 | Чтение десятичной константы (Начальное состояние) |
| A2 | Чтение десятичной константы без знака |
| A3 | Чтение десятичной константы, допускающей ноль в начале |
| A4 | Чтение беззнакового целого, допускающего экспоненциальную форму записи |
| A5 | Чтение десятичной константы, содержащей ноль вначале (нужно для вывода предупреждения) |

| | |
|----|--|
| A6 | Чтение беззнакового целого |
| A7 | Чтение беззнакового целого, содержащего ноль вначале (нужно для вывода предупреждения) |
| A8 | Чтение целого со знаком |
| A9 | Чтение беззнакового целого, допускающего ноль в начале |
| E | Конечное состояние |

A1 - начальное состояние.

Таблица 2. Обозначения на схеме

| Обозначение | Описание |
|--------------------|--|
| Цифра 0..9 | Цифры в диапазоне от 0 до 9 |
| Цифра 1..9 | Цифры в диапазоне от 1 до 9 |
| Точка | Символ-разделитель в записи десятичной дроби: точка |
| Разделитель | Символ, отделяющий число от основного текста: пробел, конец строки, табуляция, переход на следующую строку |
| Знак | Знак десятичной константы: + или - |

Обнаружение и нейтрализация ошибок

Входная цепочка может содержать некорректные данные и требуется каким-то образом выявить и нейтрализовать все имеющиеся синтаксические ошибки. Для решения данной задачи целесообразно использовать метод Айронса.

Далее мы будем использовать следующую терминологию. Ошибкой считается символ во входной цепочке, который в дереве разбора привёл к тупиковой ситуации и нет способа каким-либо образом однозначно его заменить, для устранения проблемы. Предупреждением считаем символ, который также привёл к тупиковой ситуации, но есть возможность корректно заменить его, чтобы корректно продолжить разбор строки.

Для ошибок мы, в соответствии с методом Айронса, будем отсекаать символ, приведший к тупиковой ситуации и продолжать разбор не изменяя состояние автомата. В случае предупреждений тупиковый символ будет однозначно заменён некоторым другим заранее определённым, либо удалён явным образом. После замены конечный автомат также не изменяет своего состояния.

Для данной грамматики во входной цепочке предупреждения возникают в следующих случаях:

- '0' в начале ненулевой десятичной константы. Решение: удаление символа
- '0' в начале целочисленной константы, используемой в порядке экспоненциальной формы записи числа. Решение: удаление символа
- ',' в качестве символа-разделителя при записи десятичной дроби. Решение: замена на символ '.'