

Левая рекурсия в PEG

Автор: Екатерина Вербицкая

Лаборатория языковых инструментов JetBrains
Санкт-Петербургский государственный университет
Математико-механический факультет

23 ноября 2015г.

Parser Expression Grammars

- PEG G — четверка (V, T, P, p_S) , где
 - ▶ V — конечное множество нетерминалов
 - ▶ T — алфавит (конечное множество терминалов)
 - ▶ P — функция из V в выражения (parser expression)
 - ▶ p_S — стартовое выражение
- Parser expression
 - ▶ Пустая строка ε
 - ▶ Терминал a
 - ▶ Нетерминал A
 - ▶ Последовательность $p_1 p_2$, где p_1, p_2 — parser expression
 - ▶ Упорядоченный выбор p_1 / p_2 , где p_1, p_2 — parser expression
 - ▶ 0-или-больше p^* , где p — parser expression
 - ▶ Предикат $!p$, где p — parser expression

$$G[p] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')$$

- Выражение p парсит строку xy , съедая x и оставляя y , возвращая x' как результат
- Если справа *fail*, значит, распарсить строку не удалось

Операционная семантика PEG: пустая строка

$$\overline{G[\varepsilon] \quad x \stackrel{\text{PEG}}{\leadsto} (x, \varepsilon)}$$

$$\overline{G[a] \quad ax \stackrel{\text{PEG}}{\leadsto} (x, a)}$$

$$\frac{}{G[a] \quad ax \stackrel{\text{PEG}}{\leadsto} (x, a)}$$

$$\frac{}{G[b] \quad ax \stackrel{\text{PEG}}{\leadsto} \text{fail}}, \quad b \neq a$$

$$\overline{G[a] \quad ax \stackrel{\text{PEG}}{\leadsto} (x, a)}$$

$$\overline{G[b] \quad ax \stackrel{\text{PEG}}{\leadsto} \text{fail}}, \quad b \neq a$$

$$\overline{G[a] \quad \varepsilon \stackrel{\text{PEG}}{\leadsto} \text{fail}}$$

$$\frac{G[P(A)] \quad xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}{G[A] \quad xy \overset{\text{PEG}}{\rightsquigarrow} (y, A[x'])}$$

$$\frac{G[P(A)] \quad xy \stackrel{\text{PEG}}{\rightsquigarrow} (y, x')}{G[A] \quad xy \stackrel{\text{PEG}}{\rightsquigarrow} (y, A[x'])}$$

$$\frac{G[P(A)] \quad x \stackrel{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[A] \quad x \stackrel{\text{PEG}}{\rightsquigarrow} \text{fail}}$$

Операционная семантика PEG: последовательность

$$\frac{G[p_1] \ xyz \overset{\text{PEG}}{\rightsquigarrow} (yz, x') \quad G[p_2] \ yz \overset{\text{PEG}}{\rightsquigarrow} (z, y')}{G[p_1 \ p_2] \ xyz \overset{\text{PEG}}{\rightsquigarrow} (z, x'y')}$$

Операционная семантика PEG: последовательность

$$\frac{G[p_1] \ xyz \xrightarrow{\text{PEG}} (yz, x') \quad G[p_2] \ yz \xrightarrow{\text{PEG}} (z, y')}{G[p_1 p_2] \ xyz \xrightarrow{\text{PEG}} (z, x'y')}$$

$$\frac{G[p_1] \ xy \xrightarrow{\text{PEG}} (y, x') \quad G[p_2] \ y \xrightarrow{\text{PEG}} \text{fail}}{G[p_1 p_2] \ xy \xrightarrow{\text{PEG}} \text{fail}}$$

Операционная семантика PEG: последовательность

$$\frac{G[p_1] \ xyz \xrightarrow{\text{PEG}} (yz, x') \quad G[p_2] \ yz \xrightarrow{\text{PEG}} (z, y')}{G[p_1 p_2] \ xyz \xrightarrow{\text{PEG}} (z, x'y')}$$

$$\frac{G[p_1] \ xy \xrightarrow{\text{PEG}} (y, x') \quad G[p_2] \ y \xrightarrow{\text{PEG}} \text{fail}}{G[p_1 p_2] \ xy \xrightarrow{\text{PEG}} \text{fail}}$$

$$\frac{G[p_1] \ x \xrightarrow{\text{PEG}} \text{fail}}{G[p_1 p_2] \ x \xrightarrow{\text{PEG}} \text{fail}}$$

$$\frac{G[p_1] \quad xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}{G[p_1 / p_2] \quad xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}$$

Операционная семантика PEG: выбор

$$\frac{G[p_1] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}{G[p_1 / p_2] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}$$

$$\frac{G[p_1] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail} \quad G[p_2] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[p_1 / p_2] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}$$

Операционная семантика PEG: выбор

$$\frac{G[p_1] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}{G[p_1 / p_2] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}$$

$$\frac{G[p_1] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail} \quad G[p_2] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[p_1 / p_2] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}$$

$$\frac{G[p_1] \ xy \overset{\text{PEG}}{\rightsquigarrow} \text{fail} \quad G[p_2] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}{G[p_1 / p_2] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}$$

$$\frac{G[p] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[!p] \ x \overset{\text{PEG}}{\rightsquigarrow} (x, \varepsilon)}$$

Операционная семантика PEG: предикат не

$$\frac{G[p] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[!p] \ x \overset{\text{PEG}}{\rightsquigarrow} (x, \varepsilon)}$$

$$\frac{G[p] \ xy \overset{\text{PEG}}{\rightsquigarrow} (y, x')}{G[!p] \ xy \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}$$

Операционная семантика PEG: повторение

$$\frac{G[p] \ x \overset{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[p^*] \ x \overset{\text{PEG}}{\rightsquigarrow} (x, \varepsilon)}$$

Операционная семантика PEG: повторение

$$\frac{G[p] \ x \xrightarrow{\text{PEG}} \text{fail}}{G[p^*] \ x \xrightarrow{\text{PEG}} (x, \varepsilon)}$$

$$\frac{G[p] \ xyz \xrightarrow{\text{PEG}} (yz, x') \quad G[p^*] \ yz \xrightarrow{\text{PEG}} (z, y')}{G[p^*] \ xyz \xrightarrow{\text{PEG}} (z, x'y')}$$

Ограниченная левая рекурсия

- A^n имеет не более n леворекурсивных вызовов A , A^0 всегда завершается ошибкой

$$E^0 ::= \text{fail}$$

$$E^1 ::= E^0 + n/n = \perp + n/n = n$$

$$E^2 ::= E^1 + n/n = n + n/n$$

$$E^3 ::= E^2 + n/n = (n + n/n) + n/n$$

...

$$E^n ::= E^{n-1} + n/n$$

- Ищем значение n для каждого леворекурсивного нетерминала
- Подбирается такая граница, чтобы префикс, обработанный правилом, имел максимальную длину
- Промежуточные значения сохраняются в табличку L
 - ▶ $L[(A, x) \rightarrow X](B, y) = L(B, y)$, если $B \neq A$ или $y \neq x$
 - ▶ $L[(A, x) \rightarrow X](A, x) = X$

Обработка леворекурсивного нетерминала

$$\frac{\begin{array}{l} (A, xyz) \notin \mathcal{L} \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto \mathbf{fail}] \xrightarrow{\text{PEG}} (yz, x') \\ G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto (yz, x')] \xrightarrow{\text{INC}} (z, (xy)') \end{array}}{G[A] \quad xyz \quad \mathcal{L} \xrightarrow{\text{PEG}} (z, A[(xy)'])}$$

Обработка леворекурсивного нетерминала

$$\frac{(A, xyz) \notin \mathcal{L} \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto \mathbf{fail}] \stackrel{\text{PEG}}{\rightsquigarrow} (yz, x') \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto (yz, x')] \stackrel{\text{INC}}{\rightsquigarrow} (z, (xy)')}{G[A] \quad xyz \quad \mathcal{L} \stackrel{\text{PEG}}{\rightsquigarrow} (z, A[(xy)'])}$$

$$\frac{(A, x) \notin \mathcal{L} \quad G[P(A)] \quad x \quad \mathcal{L}[(A, x) \mapsto \mathbf{fail}] \stackrel{\text{PEG}}{\rightsquigarrow} \mathbf{fail}}{G[A] \quad x \quad \mathcal{L} \stackrel{\text{PEG}}{\rightsquigarrow} \mathbf{fail}}$$

Обработка леворекурсивного нетерминала

$$\frac{(A, xyz) \notin \mathcal{L} \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto \mathbf{fail}] \xrightarrow{\text{PEG}} (yz, x') \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto (yz, x')] \xrightarrow{\text{INC}} (z, (xy)')}{G[A] \quad xyz \quad \mathcal{L} \xrightarrow{\text{PEG}} (z, A[(xy)'])}$$

$$\frac{(A, x) \notin \mathcal{L} \quad G[P(A)] \quad x \quad \mathcal{L}[(A, x) \mapsto \mathbf{fail}] \xrightarrow{\text{PEG}} \mathbf{fail}}{G[A] \quad x \quad \mathcal{L} \xrightarrow{\text{PEG}} \mathbf{fail}}$$

$$\frac{\mathcal{L}(A, xy) = \mathbf{fail}}{G[A] \quad xy \quad \mathcal{L} \xrightarrow{\text{PEG}} \mathbf{fail}}$$

Обработка леворекурсивного нетерминала

$$\frac{(A, xyz) \notin \mathcal{L} \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto \text{fail}] \xrightarrow{\text{PEG}} (yz, x') \quad G[P(A)] \quad xyz \quad \mathcal{L}[(A, xyz) \mapsto (yz, x')] \xrightarrow{\text{INC}} (z, (xy)')}{G[A] \quad xyz \quad \mathcal{L} \xrightarrow{\text{PEG}} (z, A[(xy)'])}$$

$$\frac{(A, x) \notin \mathcal{L} \quad G[P(A)] \quad x \quad \mathcal{L}[(A, x) \mapsto \text{fail}] \xrightarrow{\text{PEG}} \text{fail}}{G[A] \quad x \quad \mathcal{L} \xrightarrow{\text{PEG}} \text{fail}}$$

$$\frac{\mathcal{L}(A, xy) = \text{fail}}{G[A] \quad xy \quad \mathcal{L} \xrightarrow{\text{PEG}} \text{fail}}$$

$$\frac{\mathcal{L}(A, xy) = (y, x')}{G[A] \quad xy \quad \mathcal{L} \xrightarrow{\text{PEG}} (y, A[x'])}$$

$$\frac{G[P(A)] \ xyzw \ \mathcal{L}[(A, xyzw) \mapsto (yzw, x')] \stackrel{\text{PEG}}{\rightsquigarrow} (zw, (xy)') \quad G[P(A)] \ xyzw \ \mathcal{L}[(A, xyzw) \mapsto (zw, (xy)')] \stackrel{\text{INC}}{\rightsquigarrow} (w, (xyz)')}{G[P(A)] \ xyzw \ \mathcal{L}[(A, xyzw) \mapsto (yzw, x')] \stackrel{\text{INC}}{\rightsquigarrow} (w, (xyz)')}, \text{ where } y \neq \varepsilon$$

$$\frac{G[P(A)] \ x y z w \ \mathcal{L}[(A, x y z w) \mapsto (y z w, x')] \stackrel{\text{PEG}}{\rightsquigarrow} (z w, (x y)') \quad G[P(A)] \ x y z w \ \mathcal{L}[(A, x y z w) \mapsto (z w, (x y)')] \stackrel{\text{INC}}{\rightsquigarrow} (w, (x y z)')}{G[P(A)] \ x y z w \ \mathcal{L}[(A, x y z w) \mapsto (y z w, x')] \stackrel{\text{INC}}{\rightsquigarrow} (w, (x y z)')}, \text{ where } y \neq \varepsilon$$

$$\frac{G[P(A)] \ x \ \mathcal{L} \stackrel{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[P(A)] \ x \ \mathcal{L} \stackrel{\text{INC}}{\rightsquigarrow} \mathcal{L}(A, x)}$$

$$\frac{G[P(A)] \ x y z w \ \mathcal{L}[(A, x y z w) \mapsto (y z w, x')] \stackrel{\text{PEG}}{\rightsquigarrow} (z w, (x y)') \quad G[P(A)] \ x y z w \ \mathcal{L}[(A, x y z w) \mapsto (z w, (x y)')] \stackrel{\text{INC}}{\rightsquigarrow} (w, (x y z)')}{G[P(A)] \ x y z w \ \mathcal{L}[(A, x y z w) \mapsto (y z w, x')] \stackrel{\text{INC}}{\rightsquigarrow} (w, (x y z)')}, \text{ where } y \neq \varepsilon$$

$$\frac{G[P(A)] \ x \ \mathcal{L} \stackrel{\text{PEG}}{\rightsquigarrow} \text{fail}}{G[P(A)] \ x \ \mathcal{L} \stackrel{\text{INC}}{\rightsquigarrow} \mathcal{L}(A, x)}$$

$$\frac{G[P(A)] \ x y z \ \mathcal{L}[(A, x y z) \mapsto (z, (x y)')] \stackrel{\text{PEG}}{\rightsquigarrow} (y z, x')}{G[P(A)] \ x y z \ \mathcal{L}[(A, x y z) \mapsto (z, (x y)')] \stackrel{\text{INC}}{\rightsquigarrow} (z, (x y)')}$$

- Left recursion in Parsing Expression Grammars:
<http://arxiv.org/pdf/1207.0443.pdf>