

Parser Earley'a

1 Algorytm Earley'a

Parser Earley'a operuje na konstrukcji sekwencji zbiorów, czasami nazywanych zbiorami Earley'a. Jeżeli na wejściu znajduje się słowo $x_1x_2x_3 \dots x_n$ o długości n parser zbuduje $n + 1$ zbiorów Earley'a: jeden zbiór początkowy I_0 oraz po jednym dla każdego z parsowanych terminali. Każdy element tych zbiorów składa się z rozważanej sytuacji gramatyki oraz wskaźnika do odpowiedniego zbioru Earley'a. Sytuacja ma postać

$$(A \rightarrow \alpha \cdot \beta, j)$$

gdzie kropka stanowi odgraniczenie pomiędzy tym co już przeczytaliśmy (α) a tym co jeszcze zostało do sparsowania (β). Identyfikator j jest wskaźnikiem do zbioru I_j .

Zbiór I_i jest liczony na podstawie zbioru początkowego a następny zbiór I_{i+1} powstaje przez zastosowanie następujących trzech kroków wykonywanych tak długo, aż nic nowego nie będzie mogło powstać w kolejnej iteracji.

1. SCANNER

jeśli $(A \rightarrow \dots \cdot a \dots, j) \in I_i$ oraz $a = x_{i+1}$ to dodaj $(A \rightarrow \dots a \dots, j)$ do I_{i+1} .

2. PREDICATOR

jeśli $(A \rightarrow \dots \cdot B \dots, j) \in I_i$ to dodaj $(B \rightarrow \cdot \alpha, i)$ do I_i dla każdej produkcji $B \rightarrow \alpha$.

3. COMPLETER

jeśli $(A \rightarrow \dots \cdot, j) \in I_i$ to dodaj $(B \rightarrow \dots A \dots, k)$ do I_i dla każdego zbioru $(B \rightarrow \dots \cdot A \dots, k) \in I_j$.

Zbiór początkowy zawiera produkcję początkową $[S' \rightarrow S, 0]$, gdzie S' to nowy symbol początkowy gramatyki rozszerzonej. Słowo zostanie zaakceptowane jeżeli parser dojdzie do produkcji akceptującej $[S' \rightarrow S, 0]$.

1.

$$E \rightarrow id$$

$$E \rightarrow (E)$$

$$E \rightarrow E \text{ op } E$$

Parsing słowa $a + b * c$

Gramatyka rozszerzona

$$S \rightarrow E$$

$$E \rightarrow id$$

$$E \rightarrow (E)$$

$$E \rightarrow E \text{ op } E$$

Stan początkowy będzie miał postać:

$$I_0 = \begin{array}{ll} S \rightarrow \cdot E, & 0 \\ E \rightarrow \cdot id, & 0 \\ E \rightarrow \cdot (E), & 0 \\ E \rightarrow \cdot E \text{ op } E, & 0 \end{array}$$

Kiedy widzimy na wejściu widzimy terminal a , wykonujemy taki ruch aby go przeczytać.

$$E \rightarrow id \cdot, 0$$

Widzimy, że powyższy stan nie jest kompletny ponieważ musimy domknąć zbiór stosując trzeci krok (COMPLETER). Szukamy zatem w zbiorze I_0 tych sytuacji które mają po prawej stronie kropki nieterminal E . Otrzymujemy zatem kompletny zbiór I_1 .

$$\begin{aligned}
I_1 = & E \rightarrow id., & 0 \\
& S \rightarrow E., & 0 \\
& E \rightarrow E. \text{ op } E, & 0
\end{aligned}$$

Następnie na wejściu widzimy terminal +, musimy więc przeczytać op.

$$\begin{aligned}
I_2 = & E \rightarrow E \text{ op } .E, & 0 \\
& E \rightarrow .id, & 2 \\
& E \rightarrow .(E), & 2 \\
& E \rightarrow .E \text{ op } E, & 2
\end{aligned}$$

Na wejściu widzimy terminal b, więc wykonujemy kolejny krok.

$$\begin{aligned}
I_3 = & E \rightarrow id., & 2 \\
& E \rightarrow E \text{ op } E., & 0 \\
& E \rightarrow E. \text{ op } E, & 2 \\
& E \rightarrow E. \text{ op } E, & 0 \\
& S \rightarrow E., & 0
\end{aligned}$$

Teraz na wejściu widzimy terminal *.

$$\begin{aligned}
I_4 = & E \rightarrow E \text{ op } .E, & 2 \\
& E \rightarrow E \text{ op } .E, & 0 \\
& E \rightarrow .id, & 4 \\
& E \rightarrow .(E), & 4 \\
& E \rightarrow .E \text{ op } E, & 4
\end{aligned}$$

W końcu na wejściu czytamy znak c.

$$\begin{aligned}
I_5 = & E \rightarrow id., & 4 \\
& E \rightarrow E \text{ op } E., & 2 \\
& E \rightarrow E \text{ op } E., & 0 \\
& E \rightarrow E. \text{ op } E, & 4 \\
& E \rightarrow E. \text{ op } E, & 2 \\
& E \rightarrow E. \text{ op } E, & 0 \\
& S \rightarrow E., & 0
\end{aligned}$$

2.

$$\begin{aligned}
E & \rightarrow E + T \mid T \\
T & \rightarrow T * F \mid F \\
F & \rightarrow (E) \mid a
\end{aligned}$$

Parsing słowa $a + a * a$

Gramatyka rozszerzona

$$\begin{aligned}
S & \rightarrow E \\
E & \rightarrow E + T \mid T \\
T & \rightarrow T * F \mid F \\
F & \rightarrow (E) \mid a
\end{aligned}$$

Stan początkowy będzie miał postać:

$$\begin{aligned}
I_0 = \quad & S \rightarrow \cdot E, & 0 \\
& E \rightarrow \cdot E + T, & 0 \\
& E \rightarrow \cdot T, & 0 \\
& T \rightarrow \cdot T * F, & 0 \\
& T \rightarrow \cdot F, & 0 \\
& F \rightarrow \cdot a, & 0 \\
& F \rightarrow \cdot (E), & 0
\end{aligned}$$

Kiedy widzimy na wejściu widzimy terminal a , wykonujemy taki ruch aby go przeczytać.

$$F \rightarrow a\cdot, 0$$

Widzimy, że powyższy stan nie jest kompletny ponieważ musimy domknąć zbiór stosując trzeci krok (COMPLETER). Szukamy zatem w zbiorze I_0 tych sytuacji które mają po prawej stronie kropki nieterminal F . Otrzymujemy zatem kompletny zbiór I_1 .

$$\begin{aligned}
I_1 = \quad & F \rightarrow a\cdot, & 0 \\
& T \rightarrow F\cdot, & 0 \\
& E \rightarrow T\cdot, & 0 \\
& T \rightarrow T \cdot * F, & 0 \\
& S \rightarrow E\cdot, & 0 \\
& E \rightarrow E \cdot + T, & 0
\end{aligned}$$

Następnie na wejściu widzimy terminal $+$.

$$\begin{aligned}
I_2 = \quad & E \rightarrow E + \cdot T, & 0 \\
& T \rightarrow \cdot T * F, & 2 \\
& T \rightarrow \cdot F, & 2 \\
& F \rightarrow \cdot a, & 2 \\
& F \rightarrow \cdot (E), & 2
\end{aligned}$$

Na wejściu widzimy terminal b , więc wykonujemy kolejny krok.

$$\begin{aligned}
I_3 = \quad & F \rightarrow a\cdot, & 2 \\
& T \rightarrow F\cdot, & 2 \\
& E \rightarrow E + T\cdot, & 0 \\
& T \rightarrow T \cdot * F, & 2
\end{aligned}$$

Teraz na wejściu widzimy terminal $*$.

$$\begin{aligned}
I_4 = \quad & T \rightarrow T * \cdot F, & 2 \\
& F \rightarrow \cdot a, & 4 \\
& F \rightarrow \cdot (E), & 4
\end{aligned}$$

W końcu na wejściu czytamy znak c .

$$\begin{aligned}
I_5 = \quad & F \rightarrow a\cdot, & 4 \\
& T \rightarrow T * F\cdot, & 2 \\
& E \rightarrow E + T\cdot, & 0 \\
& T \rightarrow T \cdot * F, & 2 \\
& S \rightarrow E\cdot, & 0 \\
& E \rightarrow E \cdot + T, & 0
\end{aligned}$$

3.

Gramatyka rozszerzona

$$S' \rightarrow S$$

$S \rightarrow L = E$
 $L \rightarrow a$
 $L \rightarrow a(E)$
 $E \rightarrow E + V$
 $E \rightarrow V$
 $V \rightarrow a$
 $V \rightarrow a(E)$
 $V \rightarrow (E)$

Parsing słowa $a(a + a) = a$

Stan początkowy będzie miał postać:

$I_0 = \begin{array}{ll} S' \rightarrow \bullet S, & 0 \\ S \rightarrow \bullet L = E, & 0 \\ L \rightarrow \bullet a, & 0 \\ L \rightarrow \bullet a(E), & 0 \end{array}$

Kiedy widzimy na wejściu widzimy terminal a , wykonujemy taki ruch aby go przeczytać.

$I_1 = \begin{array}{ll} L \rightarrow a \bullet, & 0 \\ L \rightarrow a \bullet (E), & 0 \\ S \rightarrow L \bullet = E, & 0 \end{array}$

Następnie na wejściu widzimy terminal $($.

$I_2 = \begin{array}{ll} L \rightarrow a(\bullet E), & 0 \\ E \rightarrow \bullet E + V, & 2 \\ E \rightarrow \bullet V, & 2 \\ V \rightarrow \bullet a, & 2 \\ V \rightarrow \bullet a(E), & 2 \\ V \rightarrow \bullet (E), & 2 \end{array}$

Na wejściu widzimy terminal a .

$I_3 = \begin{array}{ll} V \rightarrow a \bullet, & 2 \\ V \rightarrow a \bullet (E), & 2 \\ E \rightarrow V \bullet, & 2 \\ L \rightarrow a(E \bullet), & 0 \\ E \rightarrow E \bullet + V, & 2 \end{array}$

Teraz na wejściu widzimy terminal $+$.

$I_4 = \begin{array}{ll} E \rightarrow E + \bullet V, & 2 \\ V \rightarrow \bullet a, & 4 \\ V \rightarrow \bullet a(E), & 4 \\ V \rightarrow \bullet (E), & 4 \end{array}$

Na wejściu czytamy znak a .

$I_5 = \begin{array}{ll} V \rightarrow a \bullet, & 4 \\ V \rightarrow a \bullet (E), & 4 \\ E \rightarrow E + V \bullet, & 2 \\ L \rightarrow a(E \bullet), & 0 \\ E \rightarrow E \bullet + V, & 2 \end{array}$

Na wejściu czytamy znak $)$.

$$I_6 = \begin{array}{ll} L \rightarrow a(E)\cdot, & 0 \\ S \rightarrow L\cdot = E, & 0 \end{array}$$

Na wejściu czytamy znak =.

$$I_7 = \begin{array}{ll} S \rightarrow L = \cdot E, & 0 \\ E \rightarrow \cdot E + V, & 7 \\ E \rightarrow \cdot V, & 7 \\ V \rightarrow \cdot a, & 7 \\ V \rightarrow \cdot a(E), & 7 \\ V \rightarrow \cdot (E), & 7 \end{array}$$

Czytamy znak a .

$$I_8 = \begin{array}{ll} V \rightarrow a\cdot, & 7 \\ V \rightarrow a\cdot (E), & 7 \\ E \rightarrow V\cdot, & 7 \\ S \rightarrow L = E\cdot, & 0 \\ E \rightarrow E\cdot + V, & 7 \\ S' \rightarrow S\cdot, & 0 \end{array}$$

4.

Gramatyka rozszerzona

$$\begin{array}{l} S' \rightarrow S \\ S \rightarrow S - L \\ S \rightarrow L \\ L \rightarrow (S) \\ L \rightarrow a \\ L \rightarrow a(S) \end{array}$$

Parsing słowa $a(a - a)$

Stan początkowy będzie miał postać:

$$I_0 = \begin{array}{ll} S' \rightarrow \cdot S, & 0 \\ S \rightarrow \cdot S - L, & 0 \\ S \rightarrow \cdot L, & 0 \\ L \rightarrow \cdot (S) & 0 \\ L \rightarrow \cdot a & 0 \\ L \rightarrow \cdot a(S) & 0 \end{array}$$

Na wejściu widzimy terminal a .

$$I_1 = \begin{array}{ll} L \rightarrow a\cdot, & 0 \\ L \rightarrow a\cdot (S), & 0 \\ S \rightarrow L\cdot, & 0 \\ S' \rightarrow S\cdot, & 0 \\ S \rightarrow S\cdot - L, & 0 \end{array}$$

Następnie na wejściu widzimy terminal $($.

$$I_2 = \begin{array}{ll} L \rightarrow a(\cdot S), & 0 \\ S \rightarrow \cdot S - L, & 2 \\ S \rightarrow \cdot L, & 2 \\ L \rightarrow \cdot (S) & 2 \\ L \rightarrow \cdot a & 2 \\ L \rightarrow \cdot a(S) & 2 \end{array}$$

Na wejściu widzimy terminal a .

$$I_3 = \begin{array}{ll} L \rightarrow a\bullet, & 2 \\ L \rightarrow a\bullet(S), & 2 \\ S \rightarrow L\bullet, & 2 \\ L \rightarrow a(S\bullet), & 0 \\ S \rightarrow S\bullet-L, & 2 \end{array}$$

Teraz na wejściu widzimy terminal $-$.

$$I_4 = \begin{array}{ll} S \rightarrow S-L\bullet, & 2 \\ L \rightarrow \bullet(S), & 4 \\ L \rightarrow \bullet a, & 4 \\ L \rightarrow \bullet a(S), & 4 \end{array}$$

Na wejściu czytamy znak a .

$$I_5 = \begin{array}{ll} L \rightarrow a\bullet, & 4 \\ L \rightarrow a\bullet(S), & 4 \\ S \rightarrow S-L\bullet, & 2 \\ L \rightarrow a(S\bullet), & 0 \\ S \rightarrow S\bullet-L, & 2 \end{array}$$

Na wejściu czytamy znak $)$.

$$I_6 = \begin{array}{ll} L \rightarrow a(S)\bullet, & 0 \\ S \rightarrow L\bullet, & 0 \\ S' \rightarrow S\bullet, & 0 \\ S \rightarrow S\bullet-L, & 0 \end{array}$$

5.

Gramatyka rozszerzona

$$\begin{array}{l} S' \rightarrow S \\ S \rightarrow iSeS \\ S \rightarrow iS \\ S \rightarrow a \end{array}$$

Parsing słowa $iaea$

Stan początkowy będzie miał postać:

$$I_0 = \begin{array}{ll} S' \rightarrow \bullet S, & 0 \\ S \rightarrow \bullet iSeS, & 0 \\ S \rightarrow \bullet iS, & 0 \\ S \rightarrow \bullet a & 0 \end{array}$$

Na wejściu widzimy terminal i .

$$I_1 = \begin{array}{ll} S \rightarrow i\bullet SeS, & 0 \\ S \rightarrow i\bullet S, & 0 \\ S \rightarrow \bullet iSeS, & 1 \\ S \rightarrow \bullet iS, & 1 \\ S \rightarrow \bullet a & 1 \end{array}$$

Następnie na wejściu widzimy terminal kolejny i .

$$\begin{aligned}
I_2 = \quad & S \rightarrow i \cdot SeS, & 1 \\
& S \rightarrow i \cdot S, & 1 \\
& S \rightarrow \cdot iSeS, & 2 \\
& S \rightarrow \cdot iS, & 2 \\
& S \rightarrow \cdot a & 2
\end{aligned}$$

Na wejściu widzimy terminal a .

$$\begin{aligned}
I_3 = \quad & S \rightarrow a \cdot & 2 \\
& S \rightarrow iS \cdot eS, & 1 \\
& S \rightarrow iS \cdot, & 1 \\
& S \rightarrow iS \cdot eS, & 0 \\
& S \rightarrow iS \cdot, & 0 \\
& S' \rightarrow S \cdot, & 0
\end{aligned}$$

Teraz na wejściu widzimy terminal e .

$$\begin{aligned}
I_4 = \quad & S \rightarrow iSe \cdot S, & 1 \\
& S \rightarrow iSe \cdot S, & 0 \\
& S \rightarrow \cdot iSeS, & 4 \\
& S \rightarrow \cdot iS, & 4 \\
& S \rightarrow \cdot a & 4
\end{aligned}$$

Na wejściu czytamy znak a .

$$\begin{aligned}
I_5 = \quad & S \rightarrow a \cdot & 4 \\
& S \rightarrow iSeS \cdot, & 1 \\
& S \rightarrow iSeS \cdot, & 0 \\
& S \rightarrow iS \cdot eS, & 0 \\
& S \rightarrow iS \cdot, & 0 \\
& S' \rightarrow S \cdot, & 0
\end{aligned}$$