RSM-based Generalized LL Parsing Algorithm All You Need is Generalized GLL

Semyon Grigorev

October 17, 2023

Definition (Recursive State Machine)

Recursive state machine (RSM) is a tuple $\mathcal{R} = \langle \mathcal{N}, \Sigma, B, B_S, Q, Q_S \rangle$

- N is a set of nonterminal symbols
- \bullet Σ is a set of terminal symbols
- Q is a set of states of the RSM
- Q_S is a set of the start states of all boxes
- $B = \{B_{N_i} \mid N_i \in \mathcal{N}\}$ is a set of **boxes**, where each

 $B_{N_i} = \langle Q_{N_i}, q_S, Q_F^{N_i}, \delta \rangle$ is a deterministic finite automaton without ε -transitions

- $\triangleright Q_{N_i} \subseteq Q$ is a set of states of the box
- $q_S \in Q_{N_i} \cap Q_S$ is the start state of the box
- $Q_F^{N_i} \subseteq Q_{N_i}$ is a set of final states of the box
- $\delta \subseteq Q_{N_i} \times (\Sigma \cup Q_S) \times Q_{N_i}$ is a transition function of the box
- $B_S \in B$ is the start box of the RSM

Configuration of RSM

Definition (Configuration of RSM)

A Configuration C_R of the computation of the RSM $R = \langle \mathcal{N}, \Sigma, B, B_S, Q, Q_S \rangle$ over the graph $D = \langle V, E, L \rangle$ is a tuple (q, v, S)

- $q \in Q$ is a current state of RSM
- ullet S is the current stack, which frames have one of two types
 - Return addresses frame (elements of Q) to specify states to continue computation after the call is finished
 - Parsing tree node to store fragments of a parsing tree
- $v \in V$ is the current vertex (current position in the input)

Definition (Step of RSM Computation)

A transition step of the RSM specifies how to get new configurations of RSM, given the current configuration. $C_R \vdash W$ denotes that \mathcal{R} can go to each configuration in W from the configuration C_R

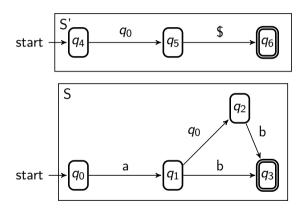
$$(q, v, w_0 :: s :: S) \vdash \{ (q', v', t :: w_0 :: s :: S) \mid (q, t, q') \in \delta, (v, t, v') \in E \}$$

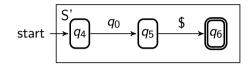
$$\cup \{ (s', v, q' :: w_0 :: s :: S) \mid (q, s', q') \in \delta, s' \in Q_S \}$$

$$\cup \{ (s, v, Node(N_i, w_0) :: S) \mid q \in Q_F^{N_i} \}$$

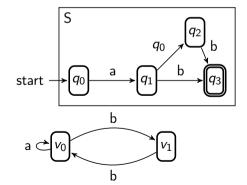
, where w_0 is a possibly empty sequence of terminals and nonterminal nodes.

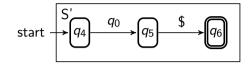
Extended RSM for grammar $S \rightarrow a \ b \mid a \ S \ b$





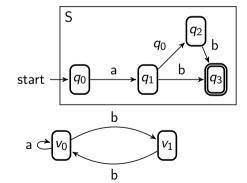
$$(q_4, v_0, []) \vdash \{(q_0, v_0, [q_5])\}$$
 (1)

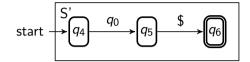


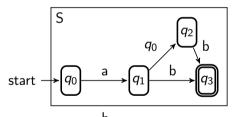


$$(q_4, v_0, []) \vdash \{(q_0, v_0, [q_5])\}$$
 (1)

$$(q_0, v_0, [q_5]) \vdash \{(q_1, v_0, [a, q_5])\}$$
 (2)





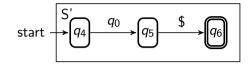


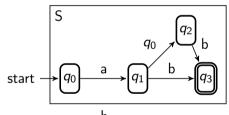
$$a \stackrel{\mathsf{D}}{\smile} v_1$$

$$(q_4, v_0, []) \vdash \{(q_0, v_0, [q_5])\}$$
 (1)

$$(q_0, v_0, [q_5]) \vdash \{(q_1, v_0, [a, q_5])\}$$
 (2)

$$(q_1, v_0, [a, q_5]) \vdash \{(q_0, v_0, [q_2, a, q_5]) , (q_3, v_1, [b, a, q_5])\}$$
(3)





$$a \stackrel{D}{\smile} v_1$$

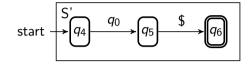
$$(q_4, v_0, []) \vdash \{(q_0, v_0, [q_5])\}$$
 (1)

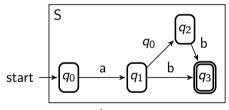
$$(q_0, v_0, [q_5]) \vdash \{(q_1, v_0, [a, q_5])\}$$
 (2)

$$(q_1, v_0, [a, q_5]) \vdash \{(q_0, v_0, [q_2, a, q_5])$$

$$,(q_3,v_1,[b,a,q_5])\}$$
 (3)

$$(q_0, v_0, [q_2, a, q_5]) \vdash \{(q_1, v_0, [a, q_2, a, q_5])\}$$
 (4)





$$a \stackrel{b}{\smile} v_1$$

$$(q_4, v_0, []) \vdash \{(q_0, v_0, [q_5])\}$$
 (1)

$$(q_0, v_0, [q_5]) \vdash \{(q_1, v_0, [a, q_5])\}$$
 (2)

$$(q_1, v_0, [a, q_5]) \vdash \{(q_0, v_0, [q_2, a, q_5])$$

$$,(q_3,v_1,[b,a,q_5])\}$$
 (3)

$$(q_0, v_0, [q_2, a, q_5]) \vdash \{(q_1, v_0, [a, q_2, a, q_5])\}$$
 (4)

$$(q_3, v_1, [b, a, q_5]) \vdash \{(q_5, v_1, [S(a, b)])\}$$
 (5)

start
$$q_4$$
 q_5 q_6 q_6

$$(q_4, v_0, []) \vdash \{(q_0, v_0, [q_5])\}$$
 (1)

$$(q_0, v_0, [q_5]) \vdash \{(q_1, v_0, [a, q_5])\}$$
 (2)

$$(q_1, v_0, [a, q_5]) \vdash \{(q_0, v_0, [q_2, a, q_5])$$

$$,(q_3,v_1,[b,a,q_5])\}$$
 (3)

$$(q_0, v_0, [q_2, a, q_5]) \vdash \{(q_1, v_0, [a, q_2, a, q_5])\}$$
 (4)

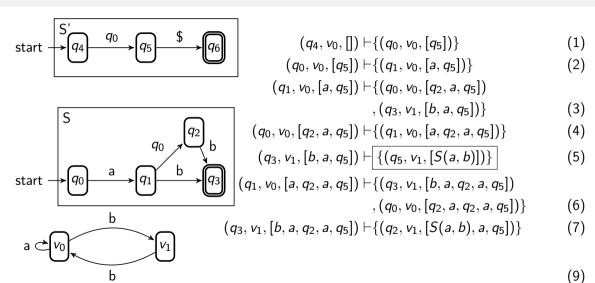
$$(q_3, v_1, [b, a, q_5]) \vdash \{(q_5, v_1, [S(a, b)])\}$$
 (5)

$$(q_1, v_0, [a, q_2, a, q_5]) \vdash \{(q_3, v_1, [b, a, q_2, a, q_5])$$

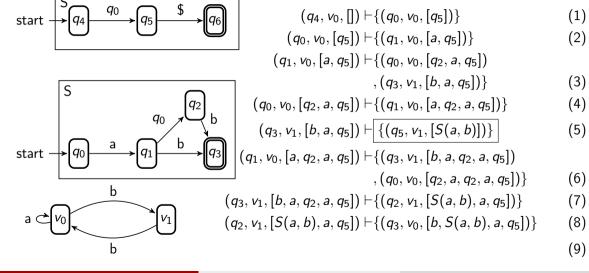
$$,(q_0,v_0,[q_2,a,q_2,a,q_5])\}$$
 (6)

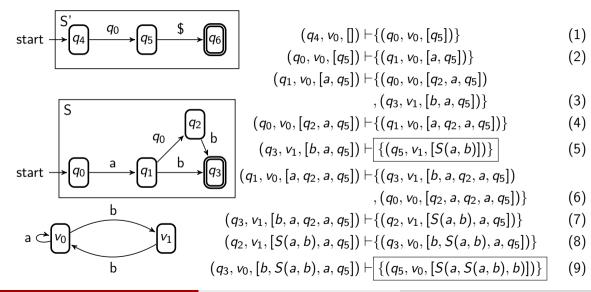
start
$$q_0$$
 q_2 q_2 q_3 q_4 q_5 q_6 q_7 q_8 q_8

$$a \stackrel{D}{\smile} v_1$$



Semyon Grigorev GLL+RSM October 17, 2023 6/10





Paths Representation

Definition (Matched Range)

Range or a matched range $R_{q_j,v_j}^{q_i,v_i}$ corresponds to the fact that there is a chain of transitions from the configuration $(q_i, v_i, \mathcal{S}_i)$ to the configuration $(q_j, v_j, \mathcal{S}_j)$, or $(q_i, v_i, \mathcal{S}_i) \vdash^* (q_j, v_j, \mathcal{S}_j)$. The symbol ϵ denotes the empty range: an empty sequence of configurations.

Paths Representation

Definition (Matched Range)

Range or a matched range $R_{q_j,v_j}^{q_i,v_i}$ corresponds to the fact that there is a chain of transitions from the configuration $(q_i, v_i, \mathcal{S}_i)$ to the configuration $(q_j, v_j, \mathcal{S}_j)$, or $(q_i, v_i, \mathcal{S}_i) \vdash^* (q_j, v_j, \mathcal{S}_j)$. The symbol ϵ denotes the empty range: an empty sequence of configurations.

Definition (Path Index)

Path index for the graph $D = \langle V, E, L \rangle$ and the query $\mathcal{R} = \langle \mathcal{N}, \Sigma, B, B_S, Q, Q_S \rangle$ is a $\mathcal{K} \times \mathcal{K}$ square matrix \mathcal{I} , $\mathcal{K} = |Q| \cdot |V|$. Columns and rows are indexed by tuples $(q_i, v_j), q_i \in Q, v_j \in V$. $\mathcal{I}[(q_i, v_j), (q_l, v_k)]$ is a set which represents information about a $R_{q_l, v_k}^{q_i, v_j}$:

- $t \in \Sigma$ means that $(q_i, v_i, S') \vdash (q_l, v_k, t :: S')$;
- $N_m \in \mathcal{N}$ means that $(q_i, v_j, \mathcal{S}') \vdash^* (q_l, v_k, N_m :: \mathcal{S}')$;
- I_{q_a,v_b} is an intermediate point: $R_{q_l,v_k}^{q_i,v_j}$ was combined from $R_{q_a,v_b}^{q_i,v_j}$ and $R_{q_l,v_k}^{q_a,v_b}$

$$(q_i, v_j, \mathcal{S}') \vdash^* (q_a, v_b, \mathcal{S}'') \vdash^* (q_l, v_k, \mathcal{S}''')$$

Semyon Grigorev GLL+RSM October 17, 2023 7/10

-			
	Configuration	Path	Stack
			q_4, v_0
1	$\vdash (q_4, v_0, (q_4, v_0), \epsilon)$		(44, 70)

	Configuration	Path	Stack
1	$\vdash (q_4, v_0, (q_4, v_0), \epsilon)$		q_4, v_0
2	$(q_4, v_0, (q_4, v_0), \epsilon) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon)\}$		q_4, v_0 q_5, ϵ q_0, v_0

	Configuration	Path	Stack
1	$\vdash (q_4, v_0, (q_4, v_0), \epsilon)$		q_4, v_0
2 3	$(q_4,v_0,(q_4,v_0),\epsilon) \vdash \{(q_0,v_0,(q_0,v_0),\epsilon)\} \ (q_0,v_0,(q_0,v_0),\epsilon) \vdash \{(q_1,v_0,(q_0,v_0),R_{q_0,v_0}^{q_0,v_0})\}$	$[(q_0,v_0),(q_1,v_0)] \leftarrow a$	q_4, v_0 q_5, ϵ q_0, v_0

Configuration	Path	Stack
$\vdash (q_4, v_0, (q_4, v_0), \epsilon)$		q_4, v_0

2
$$(q_4, v_0, (q_4, v_0), \epsilon) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon)\}\ (q_0, v_0, (q_0, v_0), \epsilon) \vdash \{(q_1, v_0, (q_0, v_0), R_{q_1, v_0}^{q_0, v_0})\}\ [(q_0, v_0), (q_1, v_0)] \leftarrow a$$

$$(q_{1},v_{0},(q_{0},v_{0}),R_{q_{1},v_{0}}^{q_{0},v_{0}})\vdash\{\overbrace{(q_{0},v_{0}),\epsilon}^{q_{0},v_{0}},(q_{3},v_{1},(q_{0},v_{0}),R_{q_{3},v_{1}}^{q_{0},v_{0}})\}\\ [(q_{1},v_{0}),(q_{3},v_{1})]\leftarrow b\\ [(q_{0},v_{0}),(q_{3},v_{1})]\leftarrow (q_{1},v_{0})\\ q_{2},R_{q_{1},v_{0}}^{q_{0},v_{0}},(q_{3},v_{1})]\leftarrow (q_{1},v_{0})$$

1
$$\vdash (q_4, v_0, (q_4, v_0), \epsilon)$$
 q_4, v_0

4
$$(q_1, v_0, (q_0, v_0), R_{q_1, v_0}^{q_0, v_0}) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon), (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0})\}$$

$$5 \quad (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0}) \vdash \{(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}), \boxed{(q_5, v_1, (q_4, v_0), R_{q_5, v_1}^{q_4, v_0})}\}$$

$$\begin{aligned} & [(q_1, v_0), (q_3, v_1)] \leftarrow b \\ & [(q_0, v_0), (q_3, v_1)] \leftarrow (q_1, v_0) \end{aligned} \qquad q_2, R_{q_1, v_0}^{q_0, v_0} \\ & [(q_4, v_0), (q_5, v_1)] \leftarrow S \\ & [(q_1, v_0), (q_2, v_1)] \leftarrow S \\ & [(q_0, v_0), (q_2, v_1)] \leftarrow (q_1, v_0) \end{aligned}$$

Path

Stack

1 $\vdash (q_4, v_0, (q_4, v_0), \epsilon)$

 q_4, v_0

2
$$(q_4, v_0, (q_4, v_0), \epsilon) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon)\}$$

3 $(q_0, v_0, (q_0, v_0), \epsilon) \vdash \{(q_1, v_0, (q_0, v_0), R_{q_0, v_0}^{q_0, v_0})\}$

 $[(q_0, v_0), (q_1, v_0)] \leftarrow a$

 q_4, v_0 $\leftarrow q_5, \epsilon \qquad q_0, v_0$

4
$$(q_1, v_0, (q_0, v_0), R_{q_1, v_0}^{q_0, v_0}) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon), (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0})\}$$

$$5 \quad (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0}) \vdash \{(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}), \boxed{(q_5, v_1, (q_4, v_0), R_{q_5, v_1}^{q_4, v_0})}\}$$

$$(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}) \vdash \{(q_3, v_0, (q_0, v_0), R_{q_3, v_0}^{q_0, v_0})\}$$

$$\begin{aligned} & [(q_1, v_0), (q_3, v_1)] \leftarrow b \\ & [(q_0, v_0), (q_3, v_1)] \leftarrow (q_1, v_0) \end{aligned} \qquad q_2, R_{q_1, v_0}^{q_0, v_0} \\ & [(q_4, v_0), (q_5, v_1)] \leftarrow S \\ & [(q_1, v_0), (q_2, v_1)] \leftarrow S \\ & [(q_0, v_0), (q_2, v_1)] \leftarrow (q_1, v_0) \end{aligned}$$

$$\begin{aligned} & q_2, R_{q_1, v_0}^{q_0, v_0} \\ & q_0, v_0 \end{aligned}$$

 q_4, v_0

Path

Stack

 $\vdash (q_4, v_0, (q_4, v_0), \epsilon)$

 q_4, v_0

2
$$(q_4, v_0, (q_4, v_0), \epsilon) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon)\}$$

3 $(q_0, v_0, (q_0, v_0), \epsilon) \vdash \{(q_1, v_0, (q_0, v_0), R_{q_0, v_0}^{q_0, v_0})\}$

$$[(q_0, v_0), (q_1, v_0)] \leftarrow a$$

$$q_5, \epsilon$$
 q_5, ϵ q_0, v_0

 q_4, v_0

4
$$(q_1, v_0, (q_0, v_0), R_{q_1, v_0}^{q_0, v_0}) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon), (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0})\}$$

$$5 \quad (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0}) \vdash \{(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}), \boxed{(q_5, v_1, (q_4, v_0), R_{q_5, v_1}^{q_4, v_0})}\}$$

6
$$(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}) \vdash \{(q_3, v_0, (q_0, v_0), R_{q_3, v_0}^{q_0, v_0})\}$$

$$7 \quad (q_3, v_0, (q_0, v_0), R_{q_3, v_0}^{q_0, v_0}) \vdash \{(q_2, v_0, (q_0, v_0), R_{q_2, v_0}^{q_0, v_0}), \overline{(q_5, v_0, (q_4, v_0), R_{q_5, v_0}^{q_4, v_0})}\}$$

$$[(q_4, v_0), (q_5, v_0)] \leftarrow S$$

$$[(q_1, v_0), (q_2, v_0)] \leftarrow S$$

$$[(q_0, v_0), (q_2, v_0)] \leftarrow (q_1, v_0)$$

$$\vdash (q_4, v_0, (q_4, v_0), \epsilon)$$

 q_4, v_0

2
$$(q_4, v_0, (q_4, v_0), \epsilon) \vdash \{(q_0, v_0, (q_0, v_0), \epsilon)\}$$
3
$$(q_0, v_0, (q_0, v_0), \epsilon) \vdash \{(q_1, v_0, (q_0, v_0), R_{q_0, v_0}^{q_0, v_0}), R_{q_0, v_0}^{q_0, v_0}\}$$

$$[(q_0, v_0), (q_1, v_0)] \leftarrow a$$

 $[(q_0, v_0), (q_3, v_1)] \leftarrow (q_2, v_0)$

 $[(a_2, v_0), (a_3, v_1)] \leftarrow b$

$$q_5, \epsilon$$
 q_0, v_0

8 / 10

$$4 \qquad (q_1, v_0, (q_0, v_0), R_{q_1, v_0}^{q_0, v_0}) \vdash \{ (q_0, v_0, (q_0, v_0), \epsilon), (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0}) \}$$

$$5 \quad (q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0}) \vdash \{(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}), \boxed{(q_5, v_1, (q_4, v_0), R_{q_5, v_1}^{q_4, v_0})}\}$$

$$(q_2, v_1, (q_0, v_0), R_{q_2, v_1}^{q_0, v_0}) \vdash \{(q_3, v_0, (q_0, v_0), R_{q_3, v_0}^{q_0, v_0})\}$$

$$7 \quad (q_3, v_0, (q_0, v_0), R_{q_3, v_0}^{q_0, v_0}) \vdash \{(q_2, v_0, (q_0, v_0), R_{q_2, v_0}^{q_0, v_0}), \boxed{(q_5, v_0, (q_4, v_0), R_{q_5, v_0}^{q_4, v_0})}\}$$

8
$$(q_2, v_0, (q_0, v_0), R_{q_2, v_0}^{q_0, v_0}) \vdash \{ \overline{(q_3, v_1, (q_0, v_0), R_{q_3, v_1}^{q_0, v_0})} \}$$

$$\begin{aligned} & [(q_1, v_0), (q_3, v_1)] \leftarrow b \\ & [(q_0, v_0), (q_3, v_1)] \leftarrow (q_1, v_0) \end{aligned} \qquad q_2, R_{q_1, v_0}^{q_0, v_0} \\ & [(q_4, v_0), (q_5, v_1)] \leftarrow S \\ & [(q_1, v_0), (q_2, v_1)] \leftarrow S \\ & [(q_0, v_0), (q_2, v_1)] \leftarrow (q_1, v_0) \end{aligned}$$

$$\begin{aligned} & [(q_2, v_1), (q_3, v_0)] \leftarrow b \\ & [(q_0, v_0), (q_3, v_0)] \leftarrow (q_2, v_1) \end{aligned}$$

$$\begin{aligned} & [(q_4, v_0), (q_5, v_0)] \leftarrow S \\ & [(q_1, v_0), (q_2, v_0)] \leftarrow S \\ & [(q_0, v_0), (q_2, v_0)] \leftarrow (q_1, v_0) \end{aligned}$$

Semyon Grigorev GLL+RSM October 17, 2023

	q_0, v_0	q_1, v_0	q_2, v_0	q_3, v_0	q_4, v_0	q_5, v_0	q_6, v_0	q_0, v_1	q_1, v_1	q_2, v_1	q_3, v_1	q_4, v_1	q_5, v_1	q_6, v_1	
q_0, v_0	($\{a\}$	$\{I_{q_1,v_0}\}$	$\{I_{q_2,v_1}\}$						$\{I_{q_1,v_0}\}$	$\{I_{q_1,v_0},I_{q_2,v_0}\}$				q_0, v_0
q_1, v_0			{ <i>S</i> }							{ S }	{ <i>b</i> }				q_1, v_0
q_2, v_0											{ <i>b</i> }				q_2, v_0
q_3, v_0															q_3, v_0
q_4, v_0						{ S }							{ S }		q_4, v_0
q_5, v_0															q_5, v_0
q_6, v_0															q_6, v_0
q_0, v_1															q_0, v_1
q_1, v_1															q_1, v_1
q_2, v_1				$\{b\}$											q_2, v_1
q_3, v_1															q_3, v_1
q_4, v_1															q_4, v_1
q_5, v_1															q_5, v_1
q_6, v_1	(q_6, v_1
	q_0, v_0	q_1, v_0	q_2, v_0	q_3, v_0	q_4, v_0	q_5, v_0	q_6, v_0	q_0, v_1	q_1, v_1	q_2, v_1	q_3, v_1	q_4, v_1	q_5, v_1	q_6, v_1	

