

บทที่ 6

การแก้ไขการผิดไวยากรณ์โดยวิธีการแจกส่วนแบบ จีแอลอาร์-ไอจีแอลอาร์

วิธีการแจกส่วนแบบ จีแอลอาร์-ไอจีแอลอาร์ หรือ จีแอลอาร์-อินเวิร์ทจีแอลอาร์ (GLR-IGLR or GLR-Inverted GLR parsing) เป็นทางออกทางหนึ่งที่จะช่วยครอบคลุมข้อผิดพลาดที่อาจถูกมองข้ามไปจากการแก้ไขข้อผิดพลาด ณ จุดที่ตรวจพบเท่านั้น โดยที่การแจกส่วนแบบนี้ใช้การทำงานกับตารางแอลอาร์ 2 ตารางด้วยกัน โดยที่ตารางแรกเป็นตารางแอลอาร์ปกติ ที่ได้จากการแปลงกฎการผลิตให้อยู่ในรูปตาราง ตัวอย่างเช่นไวยากรณ์ไม่พึ่งบริบทข้างล่างนี้สามารถแปลงให้อยู่ในรูปตารางแอลอาร์ดังแสดงในรูป

- (1) S -> NP VP
- (2) S -> S PP
- (3) NP -> n
- (4) NP -> NP PP
- (5) PP -> p NP
- (6) VP -> v NP
- (7) VP -> v PP

รูปที่ 6.1 แสดงกฎการผลิตของไวยากรณ์ไม่พึ่งบริบทอย่างง่ายในภาษาไทย

State	Action Field				Goto Field			
	n	v	p	\$	NP	PP	VP	S
0	sh3				2			1
1			sh5	acc		4		
2		sh8	sh5			7	6	
3		re3	re3	re3				
4			re2	re2				
5	sh3				9			
6			re1	re1				
7		re4	re4	re4				
8	sh3		sh5		10	11		
9		re5	re5/sh5	re5		7		
10			re6/sh5	re6		7		
11			re7	re7				

ตารางที่ 6.1 ตารางแอลอาร์ตารางที่ 1 ที่ได้จากการแปลงกฎการผลิตในรูปที่ 6.1

ส่วนตารางแอลอาร์ตารางที่ 2 ได้มาจากการกลับลำดับสัญลักษณ์ทางขวามือของกฎการผลิต ของ ตารางที่ 1 เช่น กฎการผลิตข้อ 1 คือ $S \rightarrow NP VP$ จะแปลงสัญลักษณ์ทางขวามือของกฎเป็น $S \rightarrow VP NP$ แล้ว ทำการแปลงกฎการผลิตที่สลับตำแหน่งของสัญลักษณ์ทางขวามือแล้วเหล่านี้ เป็นตารางแอลอาร์ตารางที่ 2 ดัง แสดงดังรูป

- (1) $S \rightarrow VP NP$ (5) $PP \rightarrow NP p$
 (2) $S \rightarrow PP S$ (6) $VP \rightarrow NP v$
 (3) $NP \rightarrow n$ (7) $VP \rightarrow PP v$
 (4) $NP \rightarrow PP NP$

รูปที่ 6.2 แสดงกฎการผลิตที่ได้จากการกลับไวยากรณ์ไม่พียงบริบทในรูปที่ 6.1

State	Action Field				Goto Field			
	n	v	p	\$	NP	PP	VP	S
0	sh5				3	4	2	1
1				acc				
2	sh5				6	7		
3		sh9	sh8					
4	sh5	sh12			11	4	2	10
5		re3	re3	re3				
6			sh8	re1				
7	sh5				13	7		
8	re5	re5						
9	re6							
10				re2				
11		re4/sh9	re4/sh8	re4				
12	re7							
13		re4	re4/sh8	re4				

ตารางที่ 6.2 ตารางแอลอาร์ตารางที่ 2 ที่ได้จากการแปลงกฎการผลิตในรูปที่ 6.2

การทำงานของตัวแจงส่วนแบบจีแอลอาร์-ไอจีแอลอาร์ นี้จะทำการแจงส่วนประโยคสองครั้ง โดยครั้งแรกแจงส่วนตามปกติจากทางซ้ายมือไปทางขวามือ หรือแจงส่วนประโยคจากต้นประโยคไปทางท้ายประโยคด้วย ตารางแอลอาร์ตารางแรก เมื่อพบข้อผิดพลาดในประโยคก็ยังคงทำการแก้ไขประโยค ณ จุดที่ตรวจพบแล้วทำงานต่อไปตามปกติ

การแจงส่วนประโยคในครั้งที่ 2 จะทำการแจงส่วนประโยคโดยใช้ตารางแอลอาร์ตารางที่ 2 และแจงส่วนประโยคจากทางขวามือมาทางซ้ายมือ หรือแจงส่วนประโยคจากท้ายประโยคมาต้นประโยค การแจงส่วน

ประโยคในครั้งที่สองนี้ จะช่วยแก้ไขข้อผิดพลาดที่อาจถูกมองข้ามไปในการแจงส่วนครั้งแรก จากตัวอย่างที่แสดงในครั้งก่อน ประโยคผิดไวยากรณ์ 'เขา กัด เธอ ดี ฉัน' เมื่อส่งเข้าแจงส่วนโดยวิธีนี้จะมีการทำงานดังรูปคือ



รูปที่ 6.3 ก) ขั้นตอนแรก การแจงส่วนแบบจีแอลอาร์ ตรวจสอบข้อผิดพลาดและทำการแก้ไขที่ 'ดี'
 ข) ขั้นตอนที่ 2 การแจงส่วนแบบไอจีแอลอาร์ ตรวจสอบข้อผิดพลาดและทำการแก้ไขที่ 'กัด'

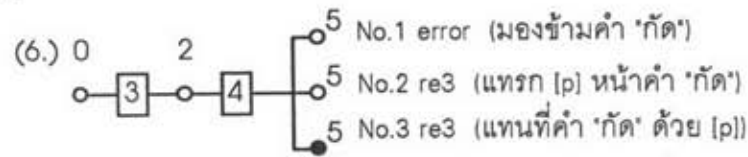
ขั้นตอนในการประมวลผลครั้งแรกของตัวแจงส่วนแบบจีแอลอาร์ปกติ จะมีการทำงานดังรูปที่แสดงในบทที่ 5 รูปที่ 5.2 ถึงรูปที่ 5.4 ข้อผิดพลาดทางไวยากรณ์จะถูกตรวจพบที่คำกริยา "ดี" และประโยคจะถูกแก้ไขเป็น "เขา กัด เธอ [p] ฉัน" ส่วนการทำงานในขั้นตอนที่สอง การแจงส่วนแบบไอจีแอลอาร์ มีการทำงานดังแสดงดังรูปที่ 6.4 ถึงรูปที่ 6.6 โดยใช้ตารางแอลอาร์ตารางที่ 6.2 ในการแจงส่วนประโยค

ประโยคที่ใช้ในการแจงส่วนครั้งแรก : เขา กัด เธอ ดี ฉัน
 กลับประโยคเพื่อใช้ในการแจงส่วนครั้งที่ 2 เป็น : ฉัน ดี เธอ กัด เขา
 ชนิดของคำที่ใช้ : n v n v n



รูปที่ 6.4 แสดงขั้นตอนการแจงส่วนประโยคโดยวิธีไอจีแอลอาร์ ข้อผิดพลาดถูกตรวจพบที่คำ "กัด"

ในการแจงส่วนครั้งที่ 2 นี้ข้อผิดพลาดจะถูกตรวจพบที่คำกริยา "กัด" โครงสร้างข้อมูลในชั้นตอนที่ (6) จะถูกแยกออกดังรูป



path No.1 next word = "เขา"

path No.2 next word = "unknown [p]"

path No.3 next word = "unknown [p]"

รูปที่ 6.5 แสดงการแยกสายโครงสร้างข้อมูลในชั้นตอนที่ (6) หลังพบข้อผิดพลาดที่ "กัด"

ด้วยการจำกัดขอบเขตการแก้ไขประโยค โดยให้มีข้อผิดพลาดได้เพียงจุดเดียวเท่านั้นในแต่ละประโยค จึงมีแต่โครงสร้างสายที่ 3 เท่านั้นที่สามารถแจงส่วนประโยคได้จนสิ้นสุด



รูปที่ 6.6 แสดงการแทนที่คำบุพบท [p] ที่คำ "กัด" และการทำงานต่อจนจบประโยค

เมื่อได้ผลการแก้ไขประโยคและนำมากลับลำดับให้ถูกต้องแล้ว ผลของการแก้ไขประโยคจากการแจงส่วนครั้งที่ 2 คือ "เขา [p] เธอ ดี ฉัน" ซึ่งสามารถแก้ไขข้อผิดพลาดที่ "กัด" ได้

จากตัวอย่างที่กำหนดให้จะเห็นว่าการแจงส่วนแบบ จีแอลอาร์-ไอจีแอลอาร์ สามารถแก้ไขข้อผิดพลาดที่คำ "wish" ซึ่งถูกมองข้ามจากการแก้ไขข้อผิดพลาด ณ จุดที่ตรวจพบเพียงเดียวได้

การแจงส่วนแบบนี้มีข้อดี คือครอบคลุมข้อผิดพลาดได้มากขึ้น เขียนโปรแกรมในการทำงานงานง่าย เพราะใช้การทำงานเหมือนกับการแก้ไข ณ จุดที่ตรวจพบทุกอย่าง เพียงแต่ต้องแปลงกฎการผลิตเพื่อสร้างตาราง

แอลอาร์ตารางที่ 2 และกลับประโยคที่จะแฉงส่วนให้สลับตำแหน่งจากหลังมาหน้า เพื่อการแฉงส่วนในชั้นตอนที่ 2 เท่านั้น

ข้อเสียของการแฉงส่วนแบบนี้ก็คือในกรณีที่มีข้อผิดพลาดมากกว่า 1 จุดในประโยคอาจมีข้อผิดพลาดบางแห่งที่ยังคงถูกมองข้ามไป เช่น

ประโยคที่ถูกไวยากรณ์	:	เขา กับ เธอ ดี ฉัน กับ น้อง
ประโยคที่ผิดไวยากรณ์	:	เขา กัด เธอ ดี ฉัน กัด น้อง
ประโยคที่แก้ไขแล้ว		
ชั้นตอนแรก	:	เขา กัด เธอ [p] ฉัน [p] น้อง
ชั้นตอนที่ 2	:	เขา [p] เธอ [p] ฉัน กัด น้อง

ในตัวอย่างนี้การผิดไวยากรณ์เกิดจากการพิมพ์ผิดที่ 'กับ' เป็น 'กัด' ทั้ง 2 คำ ทำให้ในประโยคนี้มีคำกริยาทั้งหมด 3 คำ การทำงานของตัวแฉงส่วนแบบจีแอลอาร์-ไอจีแอลอาร์ ชั้นตอนแรกทำงานจากซ้ายไปขวาดังนั้น คำกริยาคำที่ 2 และ 3 หรือ คำ 'ดี' และ 'กัด' คำหลัง จะถูกจับเป็นข้อผิดพลาดและแก้ไขประโยคเป็น 'เขา กัด เธอ [p] ฉัน [p] น้อง'

การทำงานในชั้นตอนที่ 2 ของตัวแฉงส่วน จะแฉงส่วนประโยคจากขวาไปซ้าย ดังนั้นคำที่ถูกตรวจพบเป็นข้อผิดพลาดคือคำ 'กัด' คำแรก และคำ 'ดี' จากนั้นจะแก้ไขประโยคเป็น 'เขา [p] เธอ [p] ฉัน กัด น้อง' จะเห็นได้ว่าคำตอบที่ได้ทั้งสองประโยคไม่สามารถแก้ไขความผิดพลาดที่เกิดขึ้นได้หมด มีข้อผิดพลาดบางแห่งถูกมองข้ามไปในการแก้ไขแต่ละครั้ง ที่เป็นเช่นนี้เพราะการแก้ไขข้อผิดพลาดของตัวแฉงส่วนแบบ จีแอลอาร์-ไอจีแอลอาร์ นั้นไม่มีการย้อนกลับมาพิจารณาจุดอื่นที่อาจทำให้เกิดข้อผิดพลาดได้หลังจากมีการตรวจพบข้อผิดพลาดแล้ว ในบทต่อไปจะกล่าวถึงการแฉงส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟ ซึ่งสามารถจัดการกับข้อผิดพลาดประเภทนี้ได้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย