หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

โครงสร้างข้อมูลของดัชนี

การสร้างดัชนีให้กับแฟ้มข้อมูล เป็นวิธีหนึ่งที่ช่วยในการเรียงลำดับข้อมูลภายในแฟ้มข้อมูล โดยไม่ต้องเกี่ยวข้องกับลำดับการบันทึกและโครงสร้างระเบียนข้อมูล และยังช่วยในการค้นหาระเบียนข้อมูลที่ต้องการ ได้จากการระบุคีย์ โดยทั่วไปดัชนีทุกรูปแบบจะมีส่วนที่คล้ายกัน คือ ประกอบด้วยคีย์ หรือเซตข้อมูลที่ใช้เป็นตัวแทนในการค้นหาระเบียนข้อมูล และตำแหน่งอ้างอิงระเบียนข้อมูล หรือพอยต์เตอร์ โครงสร้างข้อมูลที่ใช้จัดเรียงดัชนีซึ่งง่ายและสะดวกที่สุด ได้แก่ การจัดเรียงลำดับดัชนีภายในโครงสร้างแถวลำดับ(Array Structure) ซึ่งอยู่ในหน่วยความจำ เป็นโครงสร้างที่ให้ความเร็วในการค้นหาข้อมูล และสะดวกในการปรับโครงสร้างเมื่อมีการเปลี่ยนแปลงข้อมูลภายในโครงสร้าง เนื้อที่หน่วยความจำสำหรับโครงสร้างแถวลำดับจะต้องเพียงพอกับข้อมูลของดัชนีเมื่อจำนวนดัชนีเพิ่มขึ้น จึงทำให้โครงสร้างแถวลำดับไม่สามารถนำมาใช้ได้กับระบบงานที่มีจำนวนระเบียนข้อมูลมาก ๆ แต่มีโครงสร้างข้อมูลแบบหนึ่งที่สามารถนำมาใช้จัดเรียงดัชนีที่มีจำนวนมากได้ โครงสร้างข้อมูลแบบนี้จะแบ่งการจัดเรียงดัชนีออกเป็นส่วน ๆ ได้ เพื่อให้สามารถนำการจัดเรียงเฉพาะบางส่วนของโครงสร้าง มาเก็บในหน่วยความจำระหว่างที่ใช้งานได้ ในขณะที่ส่วนที่เหลือยังคงเก็บบันทึกไว้ภายนอก ลักษณะโครงสร้างดัชนีที่จะนำมาใช้คือ โครงสร้างบี-ทรี (B-Tree Structure)

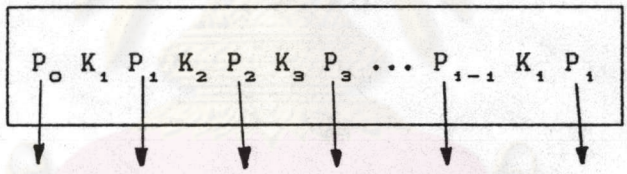
ลักษณะโครงสร้างบี-ทรี

บี-ทรี เป็นโครงสร้างข้อมูลต้นไม้(Tree Structure)รูปแบบหนึ่ง ที่มีการจัดความสมดุลย์ให้กับโนดปลาย(Terminal Node) ภายในโครงสร้างที่เป็นลำดับชั้นของต้นไม้ การทำความสมดุลย์ให้กับโนดปลาย เป็นการกระจายข้อมูลที่ถูกบันทึกภายในโครงสร้างให้แผ่ออกไปไม่ให้มีการเอนเอียงน้ำหนักไปทางปลายข้างใด เพื่อเป็นการเพิ่มประสิทธิภาพในการค้นหาข้อมูลภายในโครงสร้าง โครงสร้างบี-ทรี จึงเป็นโครงสร้างต้นไม้ที่แตกต่างจากโครงสร้างต้นไม้รูปแบบอื่น คือ เริ่มต้นการสร้างโครงสร้างจากโนดปลายและขยายขึ้นไปสู่โนดเบื้องบน แทนที่จะเริ่มจากโนดราก(Root Node) ภายในแต่ละโนด ประกอบด้วย

สมาชิกจำนวนหนึ่ง ซึ่งเป็นกลุ่มข้อมูลที่มีการจัดเรียงลำดับตามค่าของข้อมูลนั้น และมีพอยต์เตอร์ที่ชี้ไปยังโนดต่อไป กลุ่มข้อมูลนั้นคือ คีย์ และ พอยต์เตอร์ของระเบียนข้อมูล ลักษณะที่สำคัญของโครงสร้างบี-ทรีที่มีจำนวนสมาชิกแต่ละโนดได้เท่ากับ m จะสรุปได้ดังนี้

1. ภายในโครงสร้างบี-ทรี ทุกโนดต้องมีสมาชิกได้ไม่เกิน m สมาชิก
2. ทุก ๆ โนด ยกเว้นโนดราก มีสมาชิกได้อย่างน้อยที่สุดเท่ากับ m/2 สมาชิก
3. โนดปลายทั้งหมดภายในบี-ทรี ต้องอยู่ในระดับเดียวกัน และไม่มีการชี้โยงไปยังโนดใด ๆ อีก
4. จำนวนพอยต์เตอร์ที่คู่กับสมาชิก ภายในโนดรากมีได้อย่างน้อยที่สุดเท่ากับ 2 พอยต์เตอร์
5. ในขณะที่ภายในโนดที่มีจำนวนสมาชิก i สมาชิก จะประกอบด้วยจำนวนคีย์เท่ากับ i คีย์ และต้องมีพอยต์เตอร์ที่ชี้ไปที่โนดในระดับที่ต่ำกว่า เท่ากับ i+1 พอยต์เตอร์

ลักษณะภายในโนด จะเป็นดังนี้



K_i แทน คีย์ ที่มีลำดับ $K_1 < K_2 < K_3 < \ldots < K_{i-1} < K_i$

และ $1 < i < m$ (m เท่ากับ จำนวนสมาชิกที่มีได้มากที่สุดในโนด)

P_i แทน พอยต์เตอร์ที่ชี้ไปยังโนดที่อยู่ระดับต่ำกว่า ซึ่งอยู่ในตำแหน่งระหว่าง 2 คีย์ คือ K_i และ K_{i+1}

<u>การค้นหาข้อมูลภายในบี-ทรี</u>

การค้นหาข้อมูลภายในโครงสร้างต้นไม้ เป็นลักษณะของการเดินข้าม(traverse) จากโนดหนึ่งไปยังอีกโนดหนึ่งภายในโครงสร้าง และสำหรับโครงสร้างบี-ทรีซึ่งมีสมาชิกจำนวนหนึ่งภายในโนดด้วย จึงทำให้มีการทำงานถึง 2 ขั้นตอนคือ การค้นหาระหว่างโนดต่อโนด และการค้นหาภายในโนด ซึ่งสรุปวิธีการได้ดังนี้

1. เริ่มต้นจากโนดราก โดยสมมติ x เป็น คีย์ที่ต้องการค้นหา

1.1 ถ้า x มีค่าระหว่างคีย์ที่ K_i และ K_{i+1} การค้นหาจะเริ่มที่โนดที่ถูกชี้โดยพอยต์เตอร์ P_i โดยที่ $1 < i < j$ ซึ่ง j เป็นจำนวนสมาชิกที่มีในโนดนั้น

1.2 ถ้า x มีค่ามากกว่าคีย์ที่ K_j ต้องไปค้นหาที่โนดที่ถูกชี้ด้วยพอยต์เตอร์ P_j

1.3 ถ้า x มีค่าน้อยกว่าคีย์ที่ K_1 ต้องไปค้นหาที่โนดที่ถูกชี้ด้วยพอยต์เตอร์ P_0

2. การค้นหาจะสำเร็จ ต่อเมื่อพบคีย์ที่มีค่าตรงที่ต้องการ แต่ถ้าการค้นหาไปถึงโนดปลาย และไม่พบคีย์ที่ต้องการเพราะไม่มีพอยต์เตอร์ชี้ไปที่โนดใดอีก การค้นหาจะสรุปว่าไม่พบคีย์ที่ต้องการในโครงสร้างบี-ทรี

<u>การเพิ่มเติมข้อมูลในบี-ทรี</u>

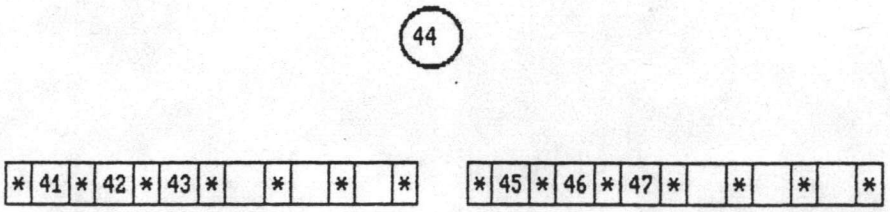
การเพิ่มเติมข้อมูล หรือการขยายโครงสร้างข้อมูลบี-ทรี เป็นการขยายโครงสร้างจากฐานหรือโนดปลายขึ้นไปสู่ยอดเบื้องบน โดยเริ่มจากไม่มีข้อมูลใดเลย การเพิ่มเติมจะต้องมีการกำหนดโนดว่างขึ้นมาโนดหนึ่งสำหรับเติมข้อมูล และระหว่างที่เติมข้อมูลลงไปจะต้องทำการเรียงลำดับข้อมูลเหล่านั้นไปด้วยพร้อมกัน

ข้อมูลเริ่มต้น : 42,43,47,46,44,41 และมี * แทน พอยต์เตอร์

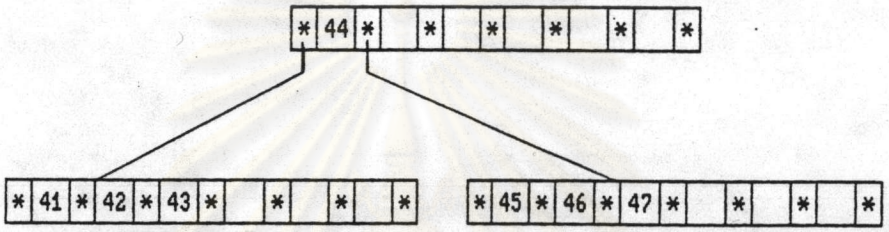
*	41	*	42	*	43	*	44	*	46	*	47	*

เมื่อจำนวนสมาชิกในโนดมีจำนวนเต็มตามที่กำหนด แต่ต้องการเพิ่มข้อมูลลงไปอีก ทำให้เกิดภาวะการแยกตัวและเลื่อนชั้นโดยโนดเดิมนั้นต้องถูกแยกตัวออก(Split) คือ ภายในโนดที่มีจำนวนสมาชิกเท่ากับ j สมาชิก จะถูกแบ่งจำนวนสมาชิกออกไปครึ่งหนึ่ง พร้อมทั้งสร้างโนดใหม่อีกหนึ่งโนด ทำให้แบ่งออกเป็น 2 ฝ่าย โดยกำหนดให้โนดทางซ้ายมือประกอบด้วยสมาชิกตัวที่ 1 ถึงตัวที่ (j/2) - 1 และให้โนดทางขวามือประกอบด้วยสมาชิกตัวที่ (j/2) + 1 ถึงตัวที่ j

เมื่อเพิ่มข้อมูล 45 ทำให้เกิดการแยกตัวของโนดเป็น 2 โนด



สำหรับสมาชิกตัวที่ j/2 ต้องผ่านการเลื่อนระดับ(promote)ขึ้นไปเบื้องบน โดยการสร้างโนดใหม่ที่อยู่ในระดับที่สูงกว่าโนดทั้งสองที่เกิดจากการแยกตัว เพื่อบรรจุสมาชิกตัวที่ j/2 เป็นสมาชิกภายในโนดนี้ และมี 2 พอยต์เตอร์ที่ทำหน้าที่เชื่อมโยงโนดนี้กับโนดทั้งสอง



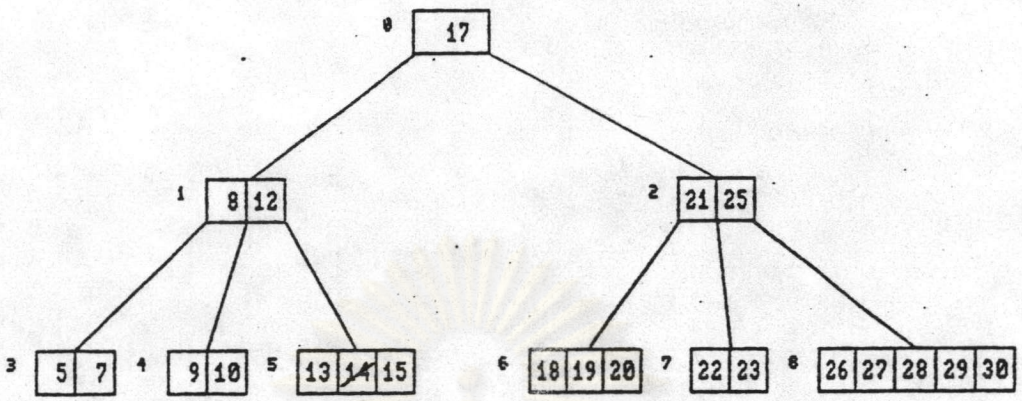
<u>การลบข้อมูลออกจากบี-ทรี</u>

เมื่อการขยายโครงสร้างบี-ทรีต้องมีการรักษาความสมดุลย์ของฐานโครงสร้างต้นไม้ ในการลบข้อมูลออกจากโครงสร้างบี-ทรี จึงต้องคำนึงถึงความสมดุลย์ของฐานโครงสร้างด้วยเช่นกัน การเปลี่ยนแปลงโครงสร้างบี-ทรีที่อาจเกิดขึ้นได้จากการลบข้อมูล แบ่งออกเป็น 6 กรณี คือ

1. เป็นกรณีที่ง่ายที่สุด เมื่อสมาชิกที่ต้องการลบอยู่ที่โนดปลาย การลบสามารถลบออกได้ทันที และทำการเรียงลำดับสมาชิกภายในโนดขึ้นใหม่

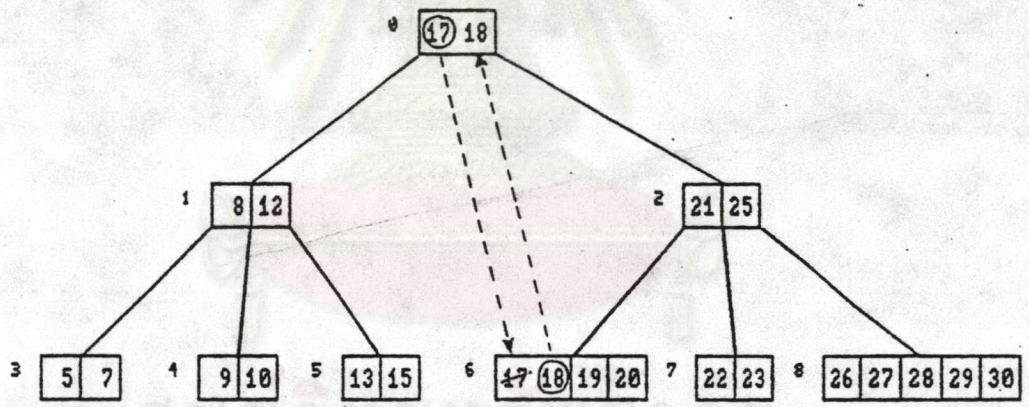
2. เมื่อสมาชิกที่ต้องการลบไม่อยู่ที่โนดปลาย ขั้นตอนการลบจึงต้องเริ่มค้นหาสมาชิกตัวถัดไปที่ค่าใกล้เคียงกับสมาชิกที่จะถูกลบมากที่สุด ซึ่งตำแหน่งของสมาชิกนั้นจะอยู่ที่โนดปลายเสมอและทำการสลับที่ระหว่างสมาชิกทั้งสอง ทำให้สมาชิกที่จะถูกลบอยู่ที่โนดปลายแล้วจึงลบสมาชิกนั้นออกจากโนดปลายได้

เมื่อลบข้อมูล 14 ออกจากโนดที่ 5



ภาพที่ 3.1 ลักษณะการลบข้อมูลออกจากบี-ทรีในกรณีที่ 1

ลบข้อมูล 17 ออกจากโนดที่ 0 ทำให้ต้องสลับที่ข้อมูล 18 ในโนดที่ 6



ภาพที่ 3.2 ลักษณะการลบข้อมูลออกจากบี-ทรีในกรณีที่ 2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ในการลบสมาชิกทั้ง 2 กรณีแรก อาจทำให้จำนวนสมาชิกที่มีในโนดปลายมีจำนวนน้อยกว่าเกณฑ์ที่กำหนด(underflow)คือ มีจำนวนน้อยกว่าครึ่งหนึ่งของจำนวนสมาชิกซึ่งต้องมีภายในโนดจึงเป็นผลให้เกิดการลบข้อมูลที่มีลักษณะต่างออกไปอีก คือ

3. เกิดการกระจายสมาชิก(redistribution) ของโนดข้างเคียงจากข้างใดข้างหนึ่งมาบรรจุภายในโนดซึ่งมีจำนวนน้อยกว่าเกณฑ์ เพื่อให้เกิดความสมดุลย์กับโนดปลาย โดยยืมสมาชิกมาจากโนดที่อยู่ในระดับเดียวกัน และมีโนดพ่อแม่(parent node)เดียวกัน เมื่อถูกกระจายสมาชิกมาแล้ว โนดที่กระจายสมาชิกมานั้นจะต้องมีสมาชิกไม่น้อยกว่าเกณฑ์ด้วยเช่นกัน

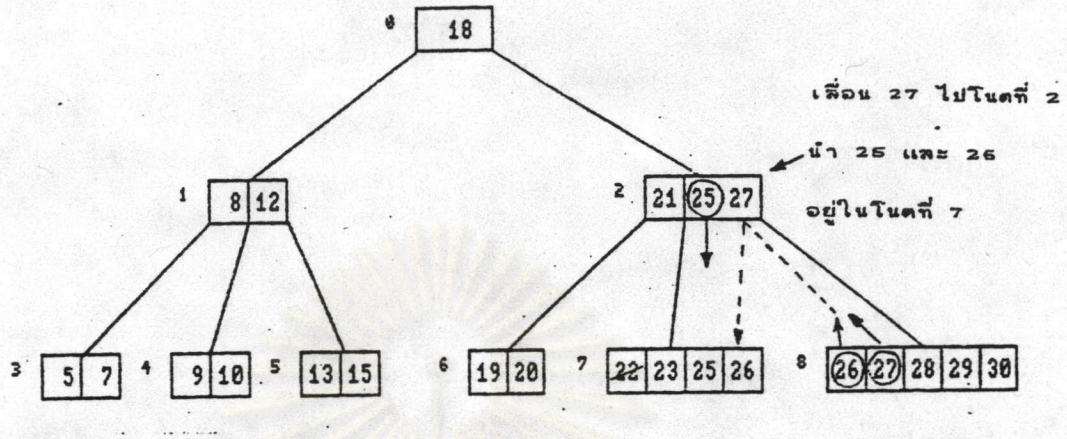
4. ถ้าการกระจายสมาชิกจากโนดข้างเคียงไม่สามารถทำได้ เพราะจำนวนสมาชิกไม่เพียงพอสำหรับการกระจาย จึงต้องทำการรวมสมาชิก(concatenation)ทั้งสองโนดให้กลายเป็นโนดเดียวกัน และขณะเดียวกันจะดึงสมาชิกในโนดพ่อแม่ซึ่งเชื่อมโยงทั้งสองโนดมารวมในโนดนี้ด้วย การรวมสมาชิกจึงเป็นลักษณะการทำงานที่ตรงกันข้ามกับการแยกตัวและการเลื่อนขึ้นของการเพิ่มเติมข้อมูล

5. จากการรวมสมาชิกทั้งสองโนด อาจส่งผลให้จำนวนสมาชิกในโนดระดับที่สูงขึ้นไปเกิดภาวะของจำนวนสมาชิกน้อยกว่าเกณฑ์ได้ ดังนั้นจึงอาจทำให้เกิดการกระจายหรือการรวมสมาชิกอย่างต่อเนื่องตามมาในโนดระดับที่สูงขึ้นไป

6. จากลักษณะการเปลี่ยนแปลงในกรณีที่ 3, 4 และ 5 อาจส่งผลให้เกิดการรวมสมาชิกในโนดราก ซึ่งทำให้ความสูงของต้นไม้บี-ทรีลดลง 1 ชั้น

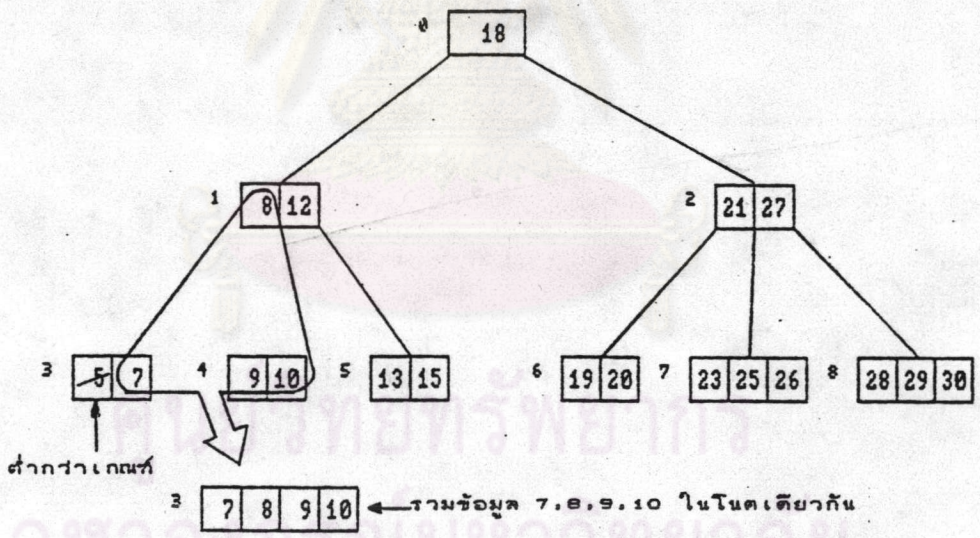
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ลบข้อมูล 22 ออกจากโนดที่ 7 ทำให้เกิดการกระจายสมาชิกมาจากโนดที่ 8



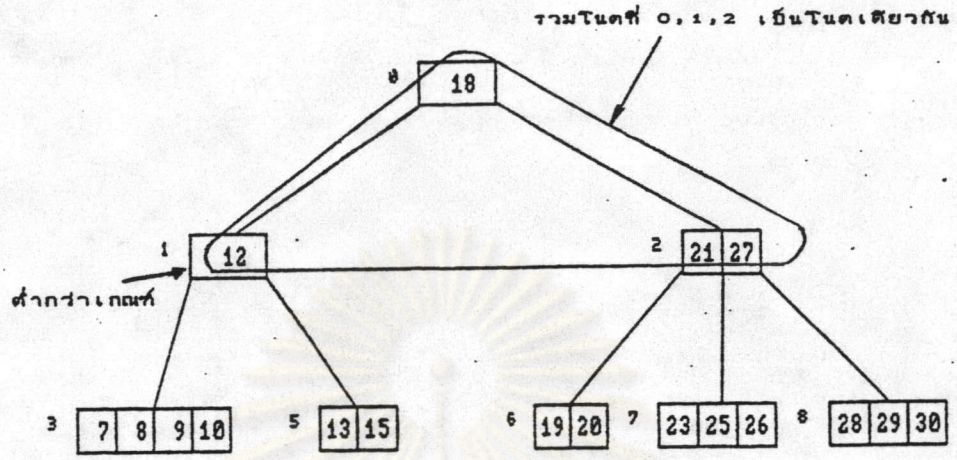
ภาพที่ 3.3 ลักษณะการลบข้อมูลออกจากบี-ทรีในกรณีที่ 3

ลบข้อมูล 5 ทำให้จำนวนสมาชิกภายในโนดที่ 3 ต่ำกว่าเกณฑ์

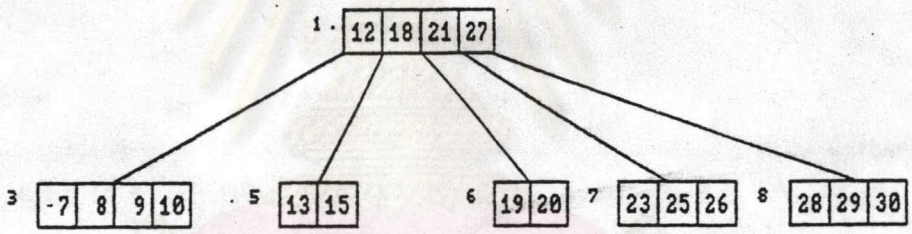


ภาพที่ 3.4 ลักษณะการลบข้อมูลออกจากบี-ทรีในกรณีที่ 4

ทำให้จำนวนสมาชิกในโนดที่ 1 ต่ำกว่าเกณฑ์ และต้องรวมข้อมูลในโนดเดียวกัน



ภาพที่ 3.5 ลักษณะการลบข้อมูลออกจากบี-ทรีในกรณีที่ 5



ภาพที่ 3.6 ลักษณะการลบข้อมูลออกจากบี-ทรีในกรณีที่ 6

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย