บทที่ 4

ดัชนีบิตแมปแบบกระจาย (Scatter Bitmap Indexing)

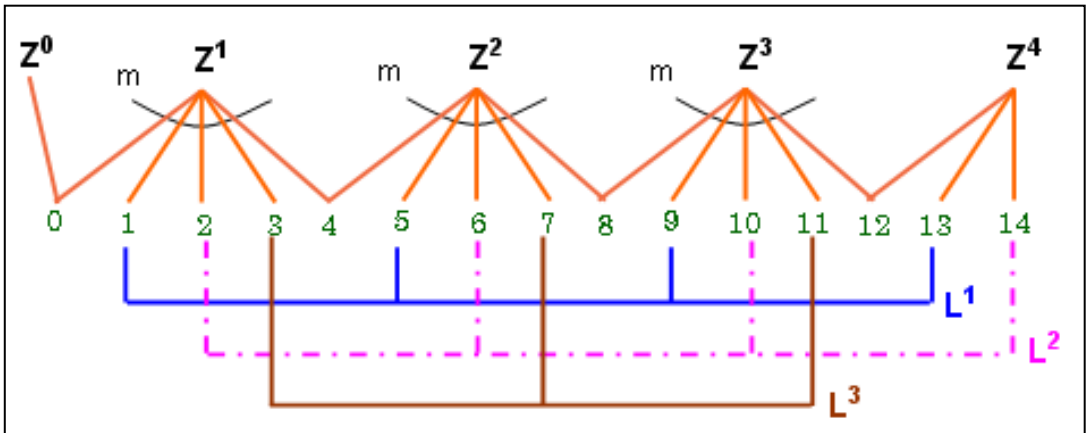
สำหรับบทนี้ จะเปรียบเทียบดัชนีบิตแมปแบบต่าง ๆ ที่มีอยู่แล้ว และนำเสนอดัชนีบิตแมปแบบใหม่ ซึ่งเรียกว่า ดัชนีบิตแมปแบบกระจาย โดยจะกล่าวถึงขั้นตอนวิธีการสร้าง การหาค่าที่เหมาะสมที่สุดของจำนวนสมาชิกภายในกลุ่ม Z (ค่า m) ขั้นตอนวิธีการค้นหาข้อมูลแบบค่าเท่ากัน ขั้นตอนวิธีการค้นหาข้อมูลแบบความเป็นสมาชิก ข้อดีของดัชนีบิตแมปแบบกระจาย และข้อจำกัดของดัชนีบิตแมปแบบกระจาย

จากบทที่ 3 จะเห็นว่า เทคนิคดัชนีบิตแมปแบบต่าง ๆ ที่กล่าวมา จะมีการแทนค่าแต่ละค่าของแอทริบิวต์ในแต่ละเรคอร์ดซ้ำซ้อนกันเกินความจำเป็น ทำให้สิ้นเปลืองพื้นที่ในการจัดเก็บ เช่น ดัชนีบิตแมปแบบพื้นฐาน จะมีการสร้างบิตแมปเวกเตอร์เท่ากับจำนวนค่าของแอทริบิวต์ โดยแต่ละบิตแมปเวกเตอร์แทนค่า 1 ค่าของแอทริบิวต์ ในดัชนีบิตแมปแบบช่วง ค่าใด ๆ ของแอทริบิวต์ จะมีการใช้ $\leq \left\lfloor \frac{C}{2} \right\rfloor$ บิตแมปเวกเตอร์ในการแทนค่า และแต่ละบิตแมปเวกเตอร์แทน $\left\lfloor \frac{C}{2} \right\rfloor$ ค่า ทำให้ดัชนีบิตแมปแบบช่วง ใช้พื้นที่น้อยลงเมื่อเทียบกับดัชนีบิตแมปแบบพื้นฐาน ดังแสดงในตาราง 4-1

ตาราง 4-1 สรุปคุณสมบัติที่สำคัญของดัชนีบิตแมปแบบพื้นฐาน แบบ Range แบบช่วง และแบบเข้ารหัส

ชนิดของดัชนีบิตแมป	จำนวนบิตแมปเวกเตอร์ที่ต้องใช้ในการแทนค่า 1 ค่าของแอทริบิวต์	จำนวนบิตแมปเวกเตอร์ที่ต้องใช้ในการแทนค่าของแอทริบิวต์ที่มีคาร์ดินอลิตี้เท่ากับ C	จำนวนค่าที่แทนโดยใช้ 1 บิตแมปเวกเตอร์
แบบพื้นฐาน	1	C	1
แบบ Range	$\leq$ C-1	C-1	$\leq$ C-1
แบบช่วง	$\leq \left\lfloor \frac{C}{2} \right\rfloor$	$\left\lceil \frac{C}{2} \right\rceil$	$\left\lfloor \frac{C}{2} \right\rfloor$
แบบเข้ารหัส	$\lceil \log_2 C \rceil$	$\lceil \log_2 C \rceil$	C

ในบทนี้ เป็นการนำเสนอเทคนิคแบบใหม่ เรียกว่า ดัชนีบิตแมปแบบกระจาย (Scatter Bitmap Indexing) ซึ่งเป็นการลดพื้นที่ในการจัดเก็บลงมาจากดัชนีบิตแมปแบบช่วง แต่ยังคงประสิทธิภาพการค้นหาอย่างเดิม กล่าวคือ ค่าใดๆ ของแอทริบิวต์ จะมีการใช้ 2 บิตแมปเวกเตอร์ในการแทนค่า โดยแต่ละบิตแมปเวกเตอร์จะแทนค่าหลายค่า หลักการคือ จัดแบ่งกลุ่มค่าของแอทริบิวต์ A เป็นกลุ่มๆ แต่ละกลุ่มมีสมาชิก m ตัว เพื่อสร้างบิตแมปเวกเตอร์ Z^j (ยกเว้น Z^0 ที่แทนค่าแรก) และบิตแมปเวกเตอร์ L^k ดังภาพประกอบ 4-1 การกำหนดค่าไปยังกลุ่มต่างๆ อธิบายในขั้นตอนวิธีการสร้าง ดังต่อไปนี้



ภาพประกอบ 4-1 รูปแบบทั่วไปของดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A (มีค่า C = 15) โดยมีจำนวนสมาชิกภายในกลุ่ม Z เท่ากับ 5 (m = 5)

4.1 ขั้นตอนวิธีการสร้าง (Algorithm)

กำหนดให้ C คือ คาร์ดินอลิตี้ของแอทริบิวต์ที่นำมาสร้างดัชนี และ m คือ จำนวนสมาชิกภายในแต่ละกลุ่ม Z โดยที่ $m \geq 2$

1. กำหนดลำดับที่ให้กับข้อมูลแต่ละค่าของแอทริบิวต์ ที่จะนำมาสร้างดัชนี โดยเริ่มจาก 0, 1, 2, ..., C-1

2. แยกค่าเรียงลำดับ (ที่ได้จากข้อ 1.) ไปยัง 2 กลุ่ม คือ กลุ่ม Z และกลุ่ม L ดังนี้ (ดังภาพประกอบ 4-2)

- การสร้างกลุ่ม Z

1. ให้ค่าลำดับที่แรกสุด (ลำดับที่ 0) อยู่ในกลุ่ม Z^0 (กำหนดให้บิตแมปเวกเตอร์ Z^0 ON)

2. กำหนดสมาชิกให้แก่แต่ละกลุ่ม Z ซึ่งใช้สัญลักษณ์ Z^j โดยใช้สูตร

$$j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1$$

โดย v คือ ค่าลำดับที่ของข้อมูล (v = 0, 1, 2, 3, ..., C-1)

และให้ค่าลำดับที่หารด้วย m-1 ลงตัว อยู่ในกลุ่ม Z ทั้ง 2 กลุ่ม คือ กลุ่ม Z^{j-1} และ Z^{j} (กำหนดให้ บิตแมปเวกเตอร์ Z^{j-1} และ Z^{j} ON)

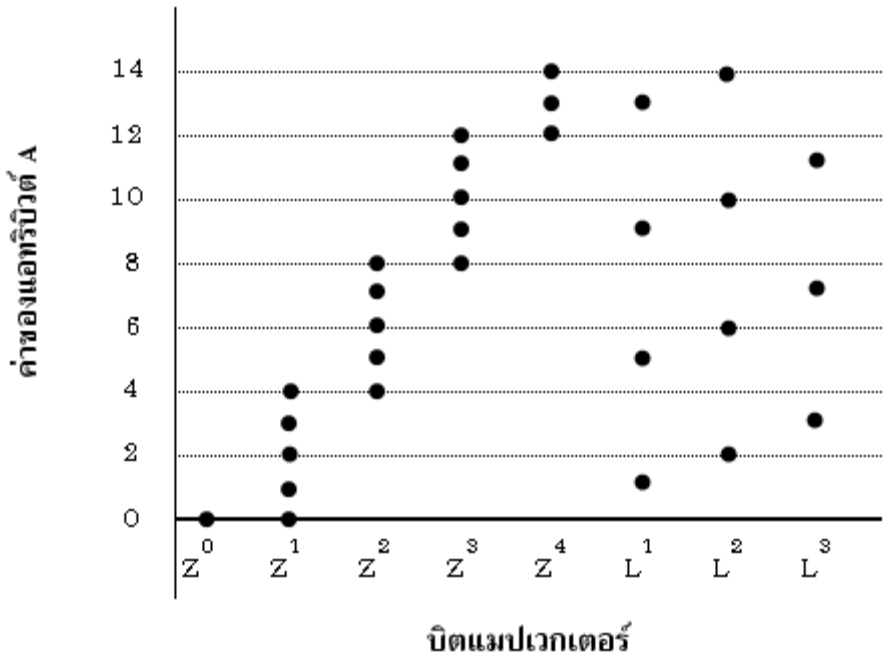
- การสร้างกลุ่ม L

เป็นการกำหนดสมาชิกให้แก่แต่ละกลุ่ม L ซึ่งใช้สัญลักษณ์ L^{k} (กำหนดให้ บิตแมปเวกเตอร์ L^{k} ON) ยกเว้นค่าลำดับที่ k = 0 โดยใช้สูตร

$$k = v \bmod (m-1)$$

โดย v คือ ค่าลำดับที่ของข้อมูล (v = 0, 1, 2, 3, ..., C-1)

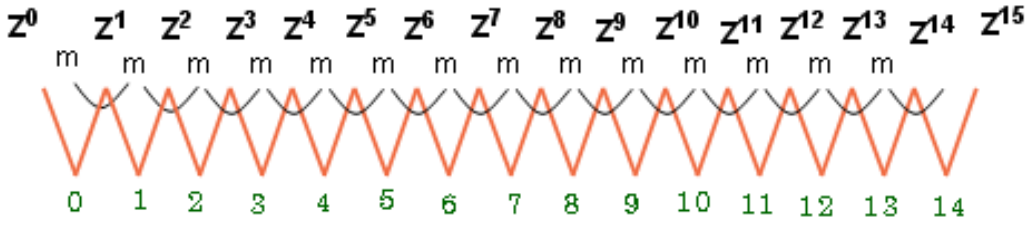
จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้างเท่ากับ $\left\lceil \frac{C}{m-1} \right\rceil + 1$ และจำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้างเท่ากับ $m-2$ (โดยที่ $(m-2) \geq 0$) ดังนั้นจำนวนบิตแมปเวกเตอร์รวมทั้งหมด เท่ากับ $\left\lceil \frac{C}{m-1} \right\rceil + m - 1$ จะเห็นว่าจำนวนบิตแมปเวกเตอร์ที่ต้องสร้าง จะขึ้นอยู่กับค่า m (จำนวนสมาชิกภายในกลุ่ม Z) ถ้าเราเลือกค่า m = 2 จำนวนบิตแมปเวกเตอร์ก็จะมากกว่าดัชนีบิตแมปแบบพื้นฐานอยู่ 1 บิตแมปเวกเตอร์ แต่ถ้าเราเลือกค่า m มาก จำนวนบิตแมปเวกเตอร์ก็จะมาก ซึ่งจะสิ้นเปลืองพื้นที่ในการจัดเก็บดัชนี ดังตัวอย่าง 4-1 จะเป็นกรณีศึกษาสำหรับการเลือกค่า m = 2, 3, 4 และ 7 ตามลำดับ เมื่อ C = 15



ภาพประกอบ 4-2 ตัวอย่างแผนภาพการทำดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ A (มีค่า C = 15, m = 5) ใช้ 8 บิตแมปเวกเตอร์

ตัวอย่าง 4-1 กำหนดให้ C = 15

กรณี m = 2 (ดูภาพประกอบ 4-3)



ภาพประกอบ 4-3 รูปแบบของดัชนีบิตแมปแบบกระจาย เมื่อ C = 15, m = 2

- จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้าง = $\left\lceil \frac{C}{m-1} \right\rceil + 1 = \left\lceil \frac{15}{2-1} \right\rceil + 1 = \left\lceil \frac{15}{1} \right\rceil + 1 = 16$ ซึ่งก็คือ บิตแมปเวกเตอร์ Z^0 ... Z^{15} โดยกำหนดให้

  - บิตแมปเวกเตอร์ Z^0 แทนค่า 0
  - บิตแมปเวกเตอร์ Z^1 แทนค่า 0, 1

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 1$ เมื่อ v = 0

    และ 1 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^1 เพราะ 1 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 1 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{2-1}$ ซึ่งก็คือ Z^1)
  - บิตแมปเวกเตอร์ Z^2 แทนค่า 1, 2

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 2$ เมื่อ v = 1

    และ 2 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^2 เพราะ 2 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 2 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{3-1}$ ซึ่งก็คือ Z^2)
  - บิตแมปเวกเตอร์ Z^3 แทนค่า 2, 3

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 3$ เมื่อ v = 2

และ 3 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^3 เพราะ 3 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 3 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{4-1}$ ซึ่งก็คือ Z^3)

- บิตแมปเวกเตอร์ Z^4 แทนค่า 3, 4

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 4$ เมื่อ v = 3

และ 4 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^4 เพราะ 4 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 4 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{5-1}$ ซึ่งก็คือ Z^4)

- บิตแมปเวกเตอร์ Z^5 แทนค่า 4, 5

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 5$ เมื่อ v = 4

และ 5 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^5 เพราะ 5 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 5 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{6-1}$ ซึ่งก็คือ Z^5)

- บิตแมปเวกเตอร์ Z^6 แทนค่า 5, 6

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 6$ เมื่อ v = 5

และ 6 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^6 เพราะ 6 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 6 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{7-1}$ ซึ่งก็คือ Z^6)

- บิตแมปเวกเตอร์ Z^7 แทนค่า 6, 7

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 7$ เมื่อ v = 6

และ 7 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^7 เพราะ 7 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 7 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{8-1}$ ซึ่งก็คือ Z^7)

- บิตแมปเวกเตอร์ Z^8 แทนค่า 7, 8

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 8$ เมื่อ v = 7

และ 8 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^8 เพราะ 8 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 8 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{9-1}$ ซึ่งก็คือ Z^8)

- บิตแมปเวกเตอร์ Z^9 แทนค่า 8, 9

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 9$ เมื่อ v = 8

และ 9 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^9 เพราะ 9 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 9 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{10-1}$ ซึ่งก็คือ Z^9)

- บิตแมปเวกเตอร์ Z^{10} แทนค่า 9, 10

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 10$ เมื่อ v = 9

และ 10 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{10} เพราะ 10 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 10 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{11-1}$ ซึ่งก็คือ Z^{10})

- บิตแมปเวกเตอร์ Z^{11} แทนค่า 10, 11

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 11$ เมื่อ v = 10

และ 11 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{11} เพราะ 11 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 11 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{12-1}$ ซึ่งก็คือ Z^{11})

- บิตแมปเวกเตอร์ Z^{12} แทนค่า 11, 12

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 12$ เมื่อ v = 11

และ 12 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{12} เพราะ 12 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 12 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{13-1}$ ซึ่งก็คือ Z^{12})

- บิตแมปเวกเตอร์ Z^{13} แทนค่า 12, 13

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 13$ เมื่อ v = 12

และ 13 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{13} เพราะ 13 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 13 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{14-1}$ ซึ่งก็คือ Z^{13})

- บิตแมปเวกเตอร์ Z^{14} แทนค่า 13, 14

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 14$ เมื่อ v = 13

และ 14 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{14} เพราะ 14 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 1 ดังนั้น 14 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{15-1}$ ซึ่งก็คือ Z^{14})

- บิตแมปเวกเตอร์ Z^{15} แทนค่า 14

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 15$ เมื่อ v = 14

- จำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้าง = $m-2$ = 2-2 = 0 ซึ่งก็คือไม่มีบิตแมปเวกเตอร์ L
- จำนวนบิตแมปเวกเตอร์ทั้งหมดที่ต้องสร้าง = จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z + จำนวนบิตแมปเวกเตอร์ในกลุ่ม L = 16 + 0 = 16

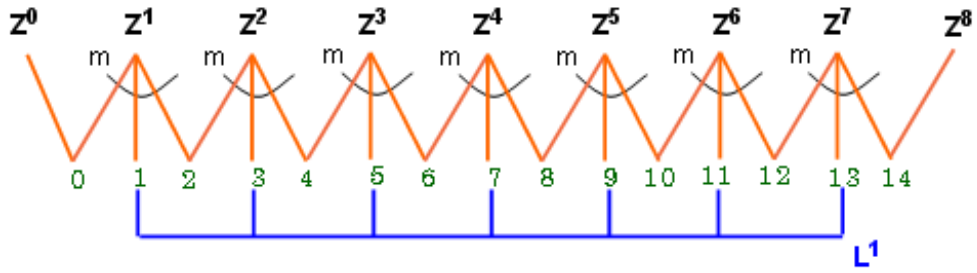
ดังนั้นการลงรหัสเรคอร์ดที่ 5 ซึ่งมีค่าเป็น 3 ในตำแหน่งบิตที่ 5 ของ Z^3 และ Z^4 จะถูกกำหนดค่าบิตให้เป็น 1 ดังภาพประกอบ 4-4

RID	$\P_A(R)$	Z^{15}	Z^{14}	Z^{13}	Z^{12}	Z^{11}	Z^{10}	Z^9	Z^8	Z^7	Z^6	Z^5	Z^4	Z^3	Z^2	Z^1	Z^0
1	14	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	3	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
3	4	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
5	3	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
7	13	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
9	6	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
10	5	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0

(a) เลือกเฉพาะแอทริบิวต์ A (b) ดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A

ภาพประกอบ 4-4 ดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ A
เมื่อ C = 15 และ m = 2 ใช้ 16 บิตแมปเวกเตอร์

กรณี m = 3 (ดูภาพประกอบ 4-5)



ภาพประกอบ 4-5 รูปแบบของดัชนีบิตแมปแบบกระจาย เมื่อ C = 15, m = 3

- จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้าง = $\left\lceil \frac{C}{m-1} \right\rceil + 1 = \left\lceil \frac{15}{3-1} \right\rceil + 1 = \left\lceil \frac{15}{2} \right\rceil + 1 = 9$ ซึ่งก็คือ บิตแมปเวกเตอร์ Z^0 ... Z^8 โดยกำหนดให้

  - บิตแมปเวกเตอร์ Z^0 แทนค่า 0
  - บิตแมปเวกเตอร์ Z^1 แทนค่า 0, 1, 2

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 1$ เมื่อ v = 0, 1

    และ 2 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^1 เพราะ 2 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 2 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{2-1}$ ซึ่งก็คือ Z^1)
  - บิตแมปเวกเตอร์ Z^2 แทนค่า 2, 3, 4

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 2$ เมื่อ v = 2, 3

    และ 4 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^2 เพราะ 4 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 4 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{3-1}$ ซึ่งก็คือ Z^2)
  - บิตแมปเวกเตอร์ Z^3 แทนค่า 4, 5, 6

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 3$ เมื่อ v = 4, 5

    และ 6 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^3 เพราะ 6 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 6 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{4-1}$ ซึ่งก็คือ Z^3)

- บิตแมปเวกเตอร์ Z^4 แทนค่า 6, 7, 8

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 4$ เมื่อ v = 6, 7

และ 8 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^4 เพราะ 8 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 8 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{5-1}$ ซึ่งก็คือ Z^4)

- บิตแมปเวกเตอร์ Z^5 แทนค่า 8, 9, 10

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 5$ เมื่อ v = 8, 9

และ 10 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^5 เพราะ 10 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 10 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{6-1}$ ซึ่งก็คือ Z^5)

- บิตแมปเวกเตอร์ Z^6 แทนค่า 10, 11, 12

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 6$ เมื่อ v = 10, 11

และ 12 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^6 เพราะ 12 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 12 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{7-1}$ ซึ่งก็คือ Z^6)

- บิตแมปเวกเตอร์ Z^7 แทนค่า 12, 13, 14

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 7$ เมื่อ v = 12, 13

และ 14 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^7 เพราะ 14 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 2 ดังนั้น 14 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{8-1}$ ซึ่งก็คือ Z^7)

- บิตแมปเวกเตอร์ Z^8 แทนค่า 14

เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 8$ เมื่อ v = 14

- จำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้าง = $m-2$ = 3-2 = 1 ซึ่งก็คือ บิตแมปเวกเตอร์ L^1 โดยกำหนดให้บิตแมปเวกเตอร์ L^1 แทนค่า 1, 3, 5, 7, 9, 11, 13 เมื่อ v = 1, 3, 5, 7, 9, 11, 13

- จำนวนบิตแมปเวกเตอร์ทั้งหมดที่ต้องสร้าง = จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z + จำนวนบิตแมปเวกเตอร์ในกลุ่ม L = 9 + 1 = 10

ดังนั้นการลงรหัสเรคอร์ดที่ 5 ซึ่งมีค่าเป็น 3 ในตำแหน่งบิตที่ 5 ของ Z^2 และ L^1 จะถูกกำหนดค่าบิตให้เป็น 1 ดังภาพประกอบ 4-6

RID	$\Pi_A(R)$
1	14
2	3
3	4
4	2
5	3
6	1
7	13
8	0
9	6
10	5

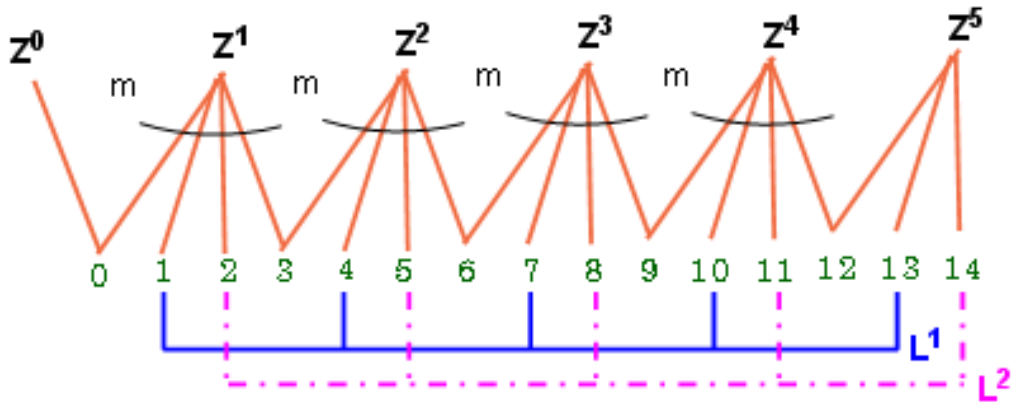
(a) เลือกเฉพาะแอทริบิวต์ A

L^1	Z^8	Z^7	Z^6	Z^5	Z^4	Z^3	Z^2	Z^1	Z^0
0	1	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	0	1	1	0
1	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	1	0
1	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	1	1	0	0	0
1	0	0	0	0	0	1	0	0	0

(b) ดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A

ภาพประกอบ 4-6 ดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ A เมื่อ C = 15 และ m = 3 ใช้ 10 บิตแมปเวกเตอร์

กรณี m = 4 (ดูภาพประกอบ 4-7)



ภาพประกอบ 4-7 รูปแบบของดัชนีบิตแมปแบบกระจาย เมื่อ C = 15, m = 4

- จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้าง = $\left\lceil \frac{C}{m-1} \right\rceil + 1 = \left\lceil \frac{15}{4-1} \right\rceil + 1 = \left\lceil \frac{15}{3} \right\rceil + 1 = 6$ ซึ่งก็คือ บิตแมปเวกเตอร์ Z^0 ... Z^5 เช่น

  - บิตแมปเวกเตอร์ Z^0 แทนค่า 0
  - บิตแมปเวกเตอร์ Z^1 แทนค่า 0, 1, 2, 3

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 1$ เมื่อ v = 0, 1, 2

    และ 3 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^1 เพราะ 3 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 3 ดังนั้น 3 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{2-1}$ ซึ่งก็คือ Z^1)
  - บิตแมปเวกเตอร์ Z^2 แทนค่า 3, 4, 5, 6

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 2$ เมื่อ v = 3, 4, 5

    และ 6 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^2 เพราะ 6 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 3 ดังนั้น 6 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{3-1}$ ซึ่งก็คือ Z^2)
  - บิตแมปเวกเตอร์ Z^3 แทนค่า 6, 7, 8, 9

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 3$ เมื่อ v = 6, 7, 8

    และ 9 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^3 เพราะ 9 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 3 ดังนั้น 9 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{4-1}$ ซึ่งก็คือ Z^3)
  - บิตแมปเวกเตอร์ Z^4 แทนค่า 9, 10, 11, 12

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 4$ เมื่อ v = 9, 10, 11

    และ 12 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^4 เพราะ 12 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 3 ดังนั้น 12 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{5-1}$ ซึ่งก็คือ Z^4)
  - บิตแมปเวกเตอร์ Z^5 แทนค่า 12, 13, 14

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 5$ เมื่อ v = 12, 13, 14

- จำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้าง = $m-2$ = 4-2 = 2 ซึ่งก็คือ บิตแมปเวกเตอร์ $L^1 \ldots L^2$ โดยกำหนดให้
  - บิตแมปเวกเตอร์ L^1 แทนค่า 1, 4, 7, 10, 13 เมื่อ v = 1, 4, 7, 10, 13
  - บิตแมปเวกเตอร์ L^2 แทนค่า 2, 5, 8, 11, 14 เมื่อ v = 2, 5, 8, 11, 14
- จำนวนบิตแมปเวกเตอร์ทั้งหมดที่ต้องสร้าง = จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z + จำนวนบิตแมปเวกเตอร์ในกลุ่ม L = 6 + 2 = 8

ดังนั้นการลงรหัสเรคอร์ดที่ 5 ซึ่งมีค่าเป็น 3 ในตำแหน่งบิตที่ 5 ของ Z^1 และ Z^2 จะถูกกำหนดค่าบิตให้เป็น 1 ดังภาพประกอบ 4-8

RID	$\Pi_A(R)$
1	14
2	3
3	4
4	2
5	3
6	1
7	13
8	0
9	6
10	5

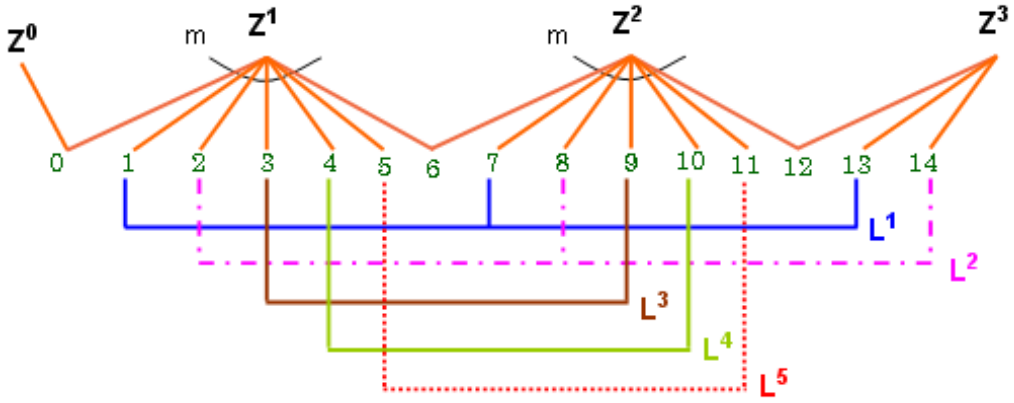
(a) เลือกเฉพาะแอทริบิวต์ A

L^2	L^1	Z^5	Z^4	Z^3	Z^2	Z^1	Z^0
1	0	1	0	0	0	0	0
0	0	0	0	0	1	1	0
0	1	0	0	0	1	0	0
1	0	0	0	0	0	1	0
0	0	0	0	0	1	1	0
0	1	0	0	0	0	1	0
0	1	1	0	0	0	0	0
0	0	0	0	0	0	1	1
0	0	0	0	1	1	0	0
1	0	0	0	0	1	0	0

(b) ดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A

ภาพประกอบ 4-8 ดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ A
เมื่อ C = 15 และ m = 4 ใช้ 8 บิตแมปเวกเตอร์

กรณี m = 7 (ดูภาพประกอบ 4-9)



ภาพประกอบ 4-9 รูปแบบของดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A
เมื่อ C = 15, m = 7

- จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้าง = $\left\lceil \frac{C}{m-1} \right\rceil + 1 = \left\lceil \frac{15}{7-1} \right\rceil + 1 = \left\lceil \frac{15}{6} \right\rceil + 1 = 4$ ซึ่งก็คือ บิตแมปเวกเตอร์ Z^0 ... Z^3 เช่น

  - บิตแมปเวกเตอร์ Z^0 แทนค่า 0
  - บิตแมปเวกเตอร์ Z^1 แทนค่า 0, 1, 2, 3, 4, 5, 6

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 1$ เมื่อ v = 0, 1, 2, 3, 4, 5

    และ 6 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^1 เพราะ 6 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 6 ดังนั้น 6 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{2-1}$ ซึ่งก็คือ Z^1)
  - บิตแมปเวกเตอร์ Z^2 แทนค่า 6, 7, 8, 9, 10, 11, 12

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 2$ เมื่อ v = 6, 7, 8, 9, 10, 11

    และ 12 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^2 เพราะ 12 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 6 ดังนั้น 12 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{3-1}$ ซึ่งก็คือ Z^2)
  - บิตแมปเวกเตอร์ Z^3 แทนค่า 12, 13, 14

    เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 3$ เมื่อ v = 12, 13, 14

- จำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้าง = $m-2$ = 7-2 = 5 ซึ่งก็คือ บิตแมปเวกเตอร์ $L^1 \ldots L^5$ เช่น
  - บิตแมปเวกเตอร์ L^1 แทนค่า 1, 7, 13 เมื่อ v = 1, 7, 13
  - บิตแมปเวกเตอร์ L^2 แทนค่า 2, 8, 14 เมื่อ v = 2, 8, 14
  - บิตแมปเวกเตอร์ L^3 แทนค่า 3, 9 เมื่อ v = 3, 9
  - บิตแมปเวกเตอร์ L^4 แทนค่า 4, 10 เมื่อ v = 4, 10
  - บิตแมปเวกเตอร์ L^5 แทนค่า 5, 11 เมื่อ v = 5, 11
- จำนวนบิตแมปเวกเตอร์ทั้งหมดที่ต้องสร้าง = จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z + จำนวนบิตแมปเวกเตอร์ในกลุ่ม L = 4 + 5 = 9

ดังนั้นการลงรหัสเรคอร์ดที่ 5 ซึ่งมีค่าเป็น 3 ในตำแหน่งบิตที่ 5 ของ Z^1 และ L^3 จะถูกกำหนดค่าบิตให้เป็น 1 ดังภาพประกอบ 4-10

RID	$\Pi_A(R)$
1	14
2	3
3	4
4	2
5	3
6	1
7	13
8	0
9	6
10	5

(a) เลือกเฉพาะแอทริบิวต์ A

L^5	L^4	L^3	L^2	L^1	Z^3	Z^2	Z^1	Z^0
0	0	0	1	0	1	0	0	0
0	0	1	0	0	0	0	1	0
0	1	0	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0
0	0	1	0	0	0	0	1	0
0	0	0	0	1	0	0	1	0
0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	1	1	0
1	0	0	0	0	0	0	1	0

(b) ดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A

ภาพประกอบ 4-10 ดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ A เมื่อ C = 15 และ m = 7 ใช้ 9 บิตแมปเวกเตอร์

จากตัวอย่างข้างบน จะเห็นได้ว่า จำนวนบิตแมปเวกเตอร์ที่ต้องสร้าง จะขึ้นอยู่กับค่า m ถ้าเลือกค่า m น้อยหรือมากเกินไป ก็จะทำให้สิ้นเปลืองพื้นที่ในการจัดเก็บดัชนี ดังนั้นเราต้องเลือกค่า m ที่เหมาะสมสำหรับการสร้างดัชนีบิตแมปแบบกระจาย เพื่อที่จะสร้างดัชนีบิตแมปแบบกระจายที่มีประสิทธิภาพมากที่สุด (ใช้พื้นที่น้อยที่สุด)

4.2 การหาค่าที่เหมาะสมที่สุดของจำนวนสมาชิกภายในกลุ่ม Z (ค่า *m*)

จากที่กล่าวมาในหัวข้อ 4.1 จะเห็นได้ว่า จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้างเท่ากับ $\left\lceil \frac{C}{m-1} \right\rceil + 1$ และจำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้างเท่ากับ $m-2$ (โดยที่ $(m-2) \geq 0$) ดังนั้นจำนวนบิตแมปเวกเตอร์รวมทั้งหมด เท่ากับ $\left\lceil \frac{C}{m-1} \right\rceil + m - 1$

เราสามารถเขียนจำนวนบิตแมปเวกเตอร์รวมทั้งหมดในรูปของฟังก์ชัน m ได้ดังนี้

$$f(m) = \left\lceil \frac{C}{m-1} \right\rceil + m - 1$$

หาอนุพันธ์ของฟังก์ชัน f(m) ได้ดังนี้

$$f'(m) = \frac{d}{dm}\left(\left(\frac{C}{m-1}\right) + m - 1\right)$$

$$= \frac{(m-1)\frac{dC}{dm} - C\frac{d(m-1)}{dm}}{(m-1)^2} + \frac{d(m-1)}{dm}$$

$$= \frac{0-C}{(m-1)^2} + 1$$

$$f'(m) = \frac{-C}{(m-1)^2} + 1$$

เพื่อหาค่าที่ดีที่สุด สามารถทำได้โดยกำหนดให้อนุพันธ์ของฟังก์ชัน เท่ากับ 0 นั่นคือ $f'(m) = 0$ จะได้ว่า

$$\frac{-C}{(m-1)^2} + 1 = 0$$

$$\frac{C}{(m-1)^2} = 1$$

$$C = (m-1)^2$$

$$\sqrt{C} = m-1$$

$$m = \sqrt{C}+1$$

จะได้ว่า $m=\sqrt{C}+1$ ดังนั้นจำนวนสมาชิกในกลุ่มที่มีค่าเท่ากับ $\sqrt{C}+1$ เป็นค่าที่ทำให้มีการสร้างจำนวนบิตแมปเวกเตอร์น้อยที่สุด กล่าวคือ ใช้พื้นที่น้อยที่สุดในการจัดเก็บดัชนี ทำให้ดัชนีบิตแมปแบบกระจายมีประสิทธิภาพมากที่สุด เมื่อ $m=\lceil\sqrt{C}+1\rceil$

เราจะได้ค่าจำนวนบิตแมปเวกเตอร์ที่ต้องสร้างเท่ากับค่าที่น้อยที่สุดของฟังก์ชัน $f(m)$ โดยการแทนค่า $m=\sqrt{C}+1$ ลงในฟังก์ชัน

$$\min\ f(m) = \min\left(\left\lceil\frac{C}{m-1}\right\rceil + m - 1\right)$$

จะได้ว่า

$$\min\ f(m) = \min\left(\left\lceil\frac{C}{(\sqrt{C}+1)-1}\right\rceil + (\sqrt{C}+1) - 1\right)$$

$$= \min\left(\left\lceil\frac{C}{\sqrt{C}}\right\rceil + \sqrt{C}\right)$$

$$= \frac{C+(\sqrt{C})(\sqrt{C})}{\sqrt{C}}$$

$$= \frac{2C}{\sqrt{C}}$$

$$\min\ f(m) = 2\sqrt{C}$$

ดังนั้นจำนวนบิตแมปเวกเตอร์ที่น้อยที่สุดคือ $\lceil 2\sqrt{C}\rceil$

จากที่กล่าวมาข้างต้น จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z เท่ากับ $\left\lceil\frac{C}{m-1}\right\rceil+1$

แทนค่า $m=\sqrt{C}+1$ ลงใน $\left\lceil\frac{C}{m-1}\right\rceil+1$ จะได้ว่า

$$\left\lceil \frac{C}{m-1} \right\rceil + 1 = \frac{C}{\left(\sqrt{C}+1\right)-1} + 1$$

$$= \frac{C}{\sqrt{C}} + 1$$

$$= \sqrt{C} + 1$$

ซึ่งก็คือ จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z เท่ากับ $\sqrt{C}+1$

และแทนค่า $m = \sqrt{C}+1$ ลงใน $m-2$ จะได้ว่า

$$m-2 = \left(\sqrt{C}+1\right)-2$$

$$= \sqrt{C}-1$$

ซึ่งก็คือ จำนวนบิตแมปเวกเตอร์ในกลุ่ม L เท่ากับ $\left\lceil \sqrt{C}-1 \right\rceil$

ตัวอย่างเช่น ถ้าแอทริบิวต์ A มีค่า C เท่ากับ 100 จำนวนบิตแมปเวกเตอร์ที่เราต้องสร้างจะเท่ากับ 20 $(=\left\lceil 2\sqrt{100} \right\rceil)$ บิตแมปเวกเตอร์ ประกอบด้วย 11 บิตแมปเวกเตอร์ในกลุ่ม Z และ 9 บิตแมปเวกเตอร์ในกลุ่ม L

จากตัวอย่าง 4.1 แอทริบิวต์ที่จะนำมาสร้างดัชนีบิตแมปแบบกระจาย ที่มีค่า C = 15 สามารถหาค่าที่เหมาะสมที่สุดของจำนวนสมาชิกภายในกลุ่ม Z (ค่า m) ได้ดังนี้

$$m = \left\lceil \sqrt{C}+1 \right\rceil$$

$$= \left\lceil \sqrt{15}+1 \right\rceil$$

$$= 5$$

ซึ่งจะเห็นได้ว่า ในกรณีที่แอทริบิวต์ที่จะนำมาสร้างดัชนีบิตแมปแบบกระจาย มีค่า C = 15 ควรจะเลือกค่า m = 5 ซึ่งเป็นค่าที่เหมาะสมที่สุด โดย (ดูภาพประกอบ 4-1)

- จำนวนบิตแมปเวกเตอร์ Z ที่ต้องสร้าง = $\left\lceil \frac{C}{m-1} \right\rceil + 1 = \left\lceil \frac{15}{5-1} \right\rceil + 1 = \left\lceil \frac{15}{4} \right\rceil + 1 = 5$ ซึ่งก็คือ บิตแมปเวกเตอร์ Z^0 ... Z^4 โดยกำหนดให้
  - บิตแมปเวกเตอร์ Z^0 แทนค่า 0

- บิตแมปเวกเตอร์ Z^1 แทนค่า 0, 1, 2, 3, 4

  เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 1$ เมื่อ v = 0, 1, 2, 3

  และ 4 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^1 เพราะ 4 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 4 ดังนั้น 4 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{2-1}$ ซึ่งก็คือ Z^1)

- บิตแมปเวกเตอร์ Z^2 แทนค่า 4, 5, 6, 7, 8

  เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 2$ เมื่อ v = 4, 5, 6, 7

  และ 8 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^2 เพราะ 8 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 4 ดังนั้น 8 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{3-1}$ ซึ่งก็คือ Z^2)

- บิตแมปเวกเตอร์ Z^3 แทนค่า 8, 9, 10, 11, 12

  เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 3$ เมื่อ v = 8, 9, 10, 11

  และ 12 ถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^3 เพราะ 12 หารด้วย m-1 ลงตัว ซึ่งในที่นี้ m-1 มีค่าเท่ากับ 4 ดังนั้น 12 จึงถูกแทนค่าด้วยบิตแมปเวกเตอร์ Z^{j-1} (ในที่นี้ $Z^{j-1} = Z^{4-1}$ ซึ่งก็คือ Z^3)

- บิตแมปเวกเตอร์ Z^4 แทนค่า 12, 13, 14

  เพราะ $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = 4$ เมื่อ v = 12, 13, 14

- จำนวนบิตแมปเวกเตอร์ L ที่ต้องสร้าง = $m-2$ = 5-2 = 3 ซึ่งก็คือ บิตแมปเวกเตอร์ $L^1 \ldots L^3$ เช่น
  - บิตแมปเวกเตอร์ L^1 แทนค่า 1, 5, 9, 13 เมื่อ v = 1, 5, 9, 13
  - บิตแมปเวกเตอร์ L^2 แทนค่า 2, 6, 10, 14 เมื่อ v = 2, 6, 10, 14
  - บิตแมปเวกเตอร์ L^3 แทนค่า 3, 7, 11 เมื่อ v = 3, 7, 11
- จำนวนบิตแมปเวกเตอร์ทั้งหมดที่ต้องสร้าง = จำนวนบิตแมปเวกเตอร์ในกลุ่ม Z + จำนวนบิตแมปเวกเตอร์ในกลุ่ม L = 5 + 3 = 8

ดังนั้นการลงรหัสเรคอร์ดที่ 5 ซึ่งมีค่าเป็น 3 ในตำแหน่งบิตที่ 5 ของ Z^1 และ L^3 จะถูกกำหนดค่าบิตให้เป็น 1 ดังภาพประกอบ 4-11

RID	$\Pi_A(R)$
1	14
2	3
3	4
4	2
5	3
6	1
7	13
8	0
9	6
10	5

(a) เลือกเฉพาะแอทริบิวต์ A

L^3	L^2	L^1	Z^4	Z^3	Z^2	Z^1	Z^0
0	1	0	1	0	0	0	0
1	0	0	0	0	0	1	0
0	0	0	0	0	1	1	0
0	1	0	0	0	0	1	0
1	0	0	0	0	0	1	0
0	0	1	0	0	0	1	0
0	0	1	1	0	0	0	0
0	0	0	0	0	0	1	1
0	1	0	0	0	1	0	0
0	0	1	0	0	1	0	0

(b) ดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ A

ภาพประกอบ 4-11 ดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ A เมื่อ C = 15 และ m = 5 (ค่าที่เหมาะสมที่สุด) ใช้ 8 บิตแมปเวกเตอร์

4.3 ขั้นตอนวิธีการค้นหาข้อมูลแบบค่าเท่ากัน

1. หาค่าลำดับที่ของข้อมูลของแอทริบิวต์ที่จะค้นหา

2. คำนวณหาว่าค่าลำดับที่นั้นอยู่ในกลุ่ม Z และ L ใด และใช้ตัวดำเนินการตรรกะ AND โดยใช้สูตร

$$\text{“A = v”} = \begin{cases} Z^{v/(m-1)} \wedge Z^{v/(m-1)+1} & \text{ถ้า } v \bmod (m-1) = 0 \text{--- (4.1)} \\ Z^{v/(m-1)+1} \wedge L^{v \bmod (m-1)} & \text{อื่น ๆ} \end{cases}$$

ตัวอย่างเช่น จากภาพประกอบ 4-11 ถ้าต้องการค้นหา A = 2

- ตรวจสอบดูว่า A = 2 อยู่ในกลุ่ม Z^j ใด โดย $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = \left\lfloor \frac{2}{5-1} \right\rfloor + 1 = 1$

- ตรวจสอบดูว่า A = 2 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m−1) = 2 mod (5−1) = 2

จะได้ว่า A = 2 อยู่ในกลุ่ม Z^1 และ L^2 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ L^2 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge L^2$ ที่มีค่าบิตเท่ากับ 1 ซึ่งก็คือ เรคอร์ดที่ 4

4.4 ขั้นตอนวิธีการค้นหาข้อมูลแบบความเป็นสมาชิก

สำหรับการค้นหาข้อมูลแบบความเป็นสมาชิก สามารถทำได้โดยการค้นหาทีละค่า แล้วมาดำเนินการตรรกะ OR ถ้าข้อมูลนั้นเป็นสมาชิกอยู่ภายในกลุ่ม Z เดียวกัน ก็สามารถอ่านเฉพาะบิตแมปเวกเตอร์ $Z^{v/(m-1)+1}$ เพียงบิตแมปเวกเตอร์เดียว

ตัวอย่างเช่น จากภาพประกอบ 4-11 ถ้าต้องการค้นหา

1) A = 1, 4 และ 6

<u>วิธีการดึงข้อมูล</u>

1.1) ตรวจสอบดูว่า A = 1 อยู่ในกลุ่ม Z^j ใด โดย $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = \left\lfloor \frac{1}{5-1} \right\rfloor + 1 = 1$

1.2) ตรวจสอบดูว่า A = 1 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m−1) = 1 mod (5−1) = 1

1.3) จะได้ว่า A = 1 อยู่ในกลุ่ม Z^1 และ L^1 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ L^1 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge L^1$ ที่มีค่าบิตเท่ากับ 1

1.4) ตรวจสอบดูว่า A = 4 อยู่ในกลุ่ม Z^j ใด โดย $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = \left\lfloor \frac{4}{5-1} \right\rfloor + 1 = 2$

1.5) ตรวจสอบดูว่า A = 4 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m−1) = 4 mod (5−1) = 0 ดังนั้น A = 4 จะอยู่ในกลุ่ม Z^{j-1} ด้วย ซึ่งก็คือ Z^1

1.6) จะได้ว่า A = 4 อยู่ในกลุ่ม Z^1 และ Z^2 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ Z^2 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge Z^2$ ที่มีค่าบิตเท่ากับ 1

1.7) ตรวจสอบดูว่า A = 6 อยู่ในกลุ่ม Z^j ใด โดย j = $\left\lfloor \frac{v}{m-1} \right\rfloor + 1$ = $\left\lfloor \frac{6}{5-1} \right\rfloor + 1 = 2$

1.8) ตรวจสอบดูว่า A = 6 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 6 mod (5-1) = 2

1.9) จะได้ว่า A = 6 อยู่ในกลุ่ม Z^2 และ L^2 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^2 ดำเนินการตรรกะ AND กับ L^2 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^2 \wedge L^2$ ที่มีค่าบิตเท่ากับ 1

1.10) นำผลลัพธ์จากข้อ 1.3), 1.6) และ 1.9) มาดำเนินการตรรกะ OR ในระดับบิตต่อบิต ซึ่งคำตอบที่ได้คือ ผลลัพธ์ของบิตแมปเวกเตอร์ที่มีค่าบิตเท่ากับ 1 (ON)

2) A = 1, 3, 5 และ 14

<u>วิธีการดึงข้อมูล</u>

2.1) ตรวจสอบดูว่า A = 1 อยู่ในกลุ่ม Z^j ใด โดย j = $\left\lfloor \frac{v}{m-1} \right\rfloor + 1$ = $\left\lfloor \frac{1}{5-1} \right\rfloor + 1 = 1$

2.2) ตรวจสอบดูว่า A = 1 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 1 mod (5-1) = 1

2.3) จะได้ว่า A = 1 อยู่ในกลุ่ม Z^1 และ L^1 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ L^1 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge L^1$ ที่มีค่าบิตเท่ากับ 1

2.4) ตรวจสอบดูว่า A = 3 อยู่ในกลุ่ม Z^j ใด โดย j = $\left\lfloor \frac{v}{m-1} \right\rfloor + 1$ = $\left\lfloor \frac{3}{5-1} \right\rfloor + 1 = 1$

2.5) ตรวจสอบดูว่า A = 3 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 3 mod (5-1) = 3

2.6) จะได้ว่า A = 3 อยู่ในกลุ่ม Z^1 และ L^3 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ L^3 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge L^3$ ที่มีค่าบิตเท่ากับ 1

2.7) ตรวจสอบดูว่า A = 5 อยู่ในกลุ่ม Z^j ใด โดย j = $\left\lfloor \frac{v}{m-1} \right\rfloor + 1$ = $\left\lfloor \frac{5}{5-1} \right\rfloor + 1 = 2$

2.8) ตรวจสอบดูว่า A = 5 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 5 mod (5-1) = 1

2.9) จะได้ว่า A = 5 อยู่ในกลุ่ม Z^2 และ L^1 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^2 ดำเนินการตรรกะ AND กับ L^1 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^2 \wedge L^1$ ที่มีค่าบิตเท่ากับ 1

2.10) ตรวจสอบดูว่า A = 14 อยู่ในกลุ่ม Z^j ใด โดย $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = \left\lfloor \frac{14}{5-1} \right\rfloor + 1 = 4$

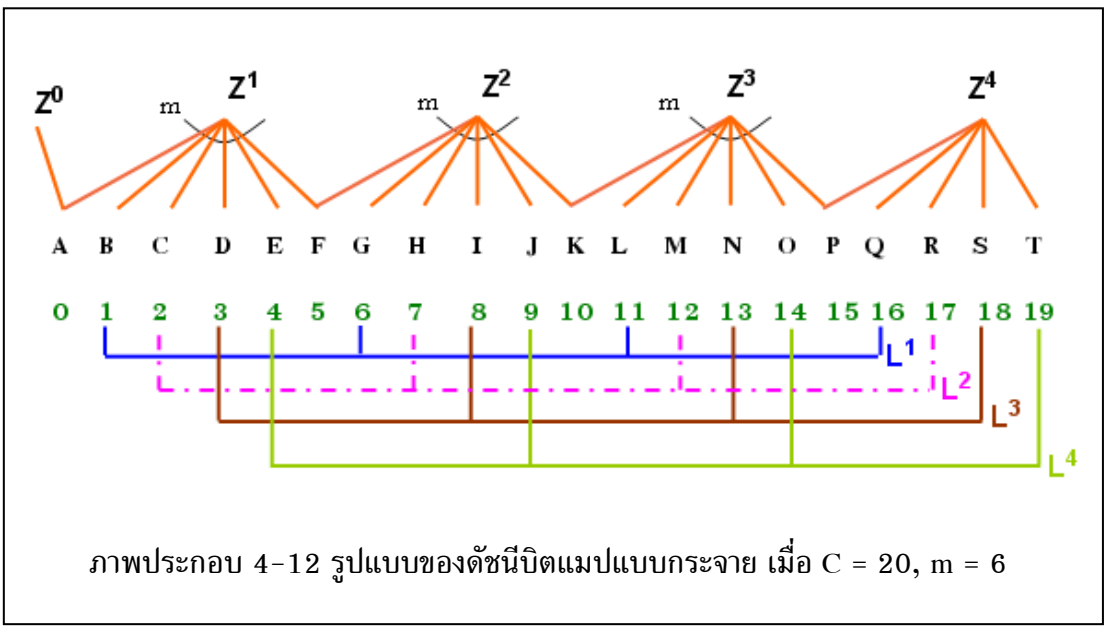
2.11) ตรวจสอบดูว่า A = 14 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 14 mod (5-1) = 2

2.12) จะได้ว่า A = 14 อยู่ในกลุ่ม Z^4 และ L^2 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^4 ดำเนินการตรรกะ AND กับ L^2 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^4 \wedge L^2$ ที่มีค่าบิตเท่ากับ 1

2.13) นำผลลัพธ์จากข้อ 2.3), 2.6), 2.9) และ 2.12) มาดำเนินการตรรกะ OR ในระดับบิตต่อบิต ซึ่งคำตอบที่ได้คือ ผลลัพธ์ของบิตแมปเวกเตอร์ที่มีค่าบิตเท่ากับ 1 (ON)

4.5 การดำเนินการตรรกะระดับบิตระหว่างบิตแมปเวกเตอร์ของต่างแอทริบิวต์

นอกจากนี้ดัชนีบิตแมปแบบกระจายยังมีคุณสมบัติในการดำเนินการตรรกะระดับบิตระหว่างบิตแมปเวกเตอร์ของต่างแอทริบิวต์ได้ด้วย เช่น จากตัวอย่างเดิม 3-2 เมื่อต้องการค้นหาข้อมูล type = 3, 14 และ brand = “B” สามารถทำได้โดยการเลือกแอทริบิวต์ type และ brand มาทำดัชนีบิตแมปแบบกระจาย ได้ดังภาพประกอบ 4-11(b) และ 4-13(b) ตามลำดับ โดยกำหนดให้ brand = “A”, “B”, “C”, …, “S”, “T” มีลำดับที่เท่ากับ 0, 1, 2, ..., 19 ตามลำดับ โดยสามารถแสดงรูปแบบของดัชนีบิตแมปแบบกระจาย เมื่อ C = 20, m = 6 (ค่าที่เหมาะสมมากที่สุด) ได้ดังภาพประกอบ 4-12



ภาพประกอบ 4-12 รูปแบบของดัชนีบิตแมปแบบกระจาย เมื่อ C = 20, m = 6

RID	$\Pi_A(R)$
1	E
2	C
3	B
4	E
5	B
6	A
7	B
8	T
9	F
10	C

(a) เลือกเฉพาะแอทริบิวต์ brand

L^4	L^3	L^2	L^1	Z^4	Z^3	Z^2	Z^1	Z^0
1	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0
1	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0
0	0	0	0	0	0	0	1	1
0	0	0	1	0	0	0	1	0
1	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	1	0
0	0	1	0	0	0	0	1	0

(b) ดัชนีบิตแมปแบบกระจาย บนแอทริบิวต์ brand

ภาพประกอบ 4-13 ดัชนีบิตแมปแบบกระจายบนแอทริบิวต์ brand เมื่อ C=20 และ m = 6 (ค่าที่เหมาะสมที่สุด) ใช้ 9 บิตแมปเวกเตอร์

<u>วิธีการดึงข้อมูล</u>

1) ดึงข้อมูลแอทริบิวต์ type (แอทริบิวต์ A ภาพประกอบ 4-11(b)) โดยการ

1.1) ตรวจสอบดูว่า A = 3 อยู่ในกลุ่ม Z^j ใด โดย $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = \left\lfloor \frac{3}{5-1} \right\rfloor + 1 = 1$

1.2) ตรวจสอบดูว่า A = 3 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 3 mod (5-1) = 3

1.3) จะได้ว่า A = 3 อยู่ในกลุ่ม Z^1 และ L^3 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ L^3 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge L^3$ ที่มีค่าบิตเท่ากับ 1 ดังภาพประกอบ 4-14(a)

1.4) ตรวจสอบดูว่า A = 14 อยู่ในกลุ่ม Z^j ใด โดย $j = \left\lfloor \frac{v}{m-1} \right\rfloor + 1 = \left\lfloor \frac{14}{5-1} \right\rfloor + 1 = 4$

1.5) ตรวจสอบดูว่า A = 14 อยู่ในกลุ่ม L^k ใด โดย k = v mod (m-1) = 14 mod (5-1) = 2

1.6) จะได้ว่า A = 14 อยู่ในกลุ่ม Z^4 และ L^2 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^4 ดำเนินการตรรกะ AND กับ L^2 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^4 \wedge L^2$ ที่มีค่าบิตเท่ากับ 1 ดังภาพประกอบ 4-14(b)

1.7) นำผลลัพธ์จากข้อ 1.3) และ 1.6) มาดำเนินการตรรกะ OR ระดับบิตระหว่างบิตแมปเวกเตอร์ ผลลัพธ์ที่ได้ คือ บิตที่มีค่าเท่ากับ 1 ดังภาพประกอบ 4-14(c)

2) ดึงข้อมูลแอทริบิวต์ brand (แอทริบิวต์ brand ภาพประกอบ 4-13(b)) โดย

2.1) ตรวจสอบดูว่า brand = "B" อยู่ในกลุ่ม Z^j ใด โดย j = $\left\lfloor \frac{v}{m-1} \right\rfloor + 1$ = $\left\lfloor \frac{1}{6-1} \right\rfloor + 1 = 1$

2.2) ตรวจสอบดูว่า brand = "B" อยู่ในกลุ่ม L^k ใด โดย k=v mod (m-1) = 1 mod (6-1) = 1

2.3) จะได้ว่า brand = "B" อยู่ในกลุ่ม Z^1 และ L^1 ดังนั้นจึงอ่านบิตแมปเวกเตอร์ Z^1 ดำเนินการตรรกะ AND กับ L^1 ผลลัพธ์ที่ได้ คือ ผลลัพธ์ของ $Z^1 \wedge L^1$ ที่มีค่าบิตเท่ากับ 1 ดังภาพประกอบ 4-14(d)

3) นำผลลัพธ์จากข้อ 1.7) และ 2.3) มาดำเนินการตรรกะ AND ระดับบิตระหว่างบิตแมปเวกเตอร์ ดังภาพประกอบ 4-14(e)

$Z^1 \wedge L^3$		$Z^4 \wedge L^2$		$(Z^1 \wedge L^3) \vee (Z^4 \wedge L^2)$		$Z^1 \wedge L^1$		ผลลัพธ์
0		1		1		0		0
1		0		1		0		0
0		0		0		1		0
0		0		0		0		0
1	OR	0	⇨	1	AND	1	⇨	1
0		0		0		0		0
0		0		0		1		0
0		0		0		0		0
0		0		0		0		0
0		0		0		0		0
(a)		(b)		(c)		(d)		(e)

โดยที่ (a) ผลลัพธ์จากการ $Z^1 \wedge L^3$ เพื่อค้นหาค่า A=3

(b) ผลลัพธ์จากการ $Z^4 \wedge L^2$ เพื่อค้นหาค่า A=14

(c) ผลลัพธ์จากการค้นหาค่า A=3, 14 มาดำเนินการตรรกะ OR ในระดับบิตต่อบิต

(d) ผลลัพธ์จากการ $Z^1 \wedge L^1$ เพื่อค้นหาค่า brand = "B"

(e) ผลลัพธ์จากการค้นหาค่า A=3, 14 และค่า brand="B" มาดำเนินการตรรกะ AND ในระดับบิตต่อบิต คือ บิตที่ 5 มีค่าบิตเท่ากับ 1

ภาพประกอบ 4-14 การดำเนินการตรรกะระดับบิตของแอทริบิวต์ type และ brand ของดัชนีบิตแมปแบบกระจาย

4.6 ข้อดีของดัชนีบิตแมปแบบกระจาย

ดัชนีบิตแมปแบบกระจาย ประหยัดพื้นที่ได้มากกว่าดัชนีบิตแมปแบบพื้นฐาน แบบ Range และแบบช่วง และเหมาะสำหรับการสอบถามข้อมูลแบบค่าเท่ากันและแบบความเป็นสมาชิก เพราะบิตแมปเวกเตอร์ที่ถูกอ่าน มีจำนวน 2 บิตแมปเวกเตอร์ และในการค้นหาค่าเท่ากันมีการดำเนินการตรรกะ AND เกิดขึ้นเพียงตัวดำเนินการเดียวเท่านั้น

นอกจากนี้ ดัชนีบิตแมปแบบกระจายจะมีประสิทธิภาพยิ่งขึ้น (ใช้เวลาน้อยลง) ในการค้นหาข้อมูลแบบความเป็นสมาชิก กรณีที่จำนวนค่าที่ต้องการค้นหาเท่ากับค่า m และค่าที่ต้องการค้นหานั้นอยู่ในกลุ่ม Z หรือ L เดียวกัน

4.7 ข้อจำกัดของดัชนีบิตแมปแบบกระจาย

ดัชนีบิตแมปแบบกระจาย จะใช้พื้นที่ในการจัดเก็บดัชนีมากกว่าดัชนีบิตแมปแบบเข้ารหัส แต่ดัชนีบิตแมปแบบกระจายจะค้นหาได้เร็วกว่าดัชนีบิตแมปแบบเข้ารหัส และดัชนีบิตแมปแบบกระจายจะใช้พื้นที่มากกว่าดัชนีบิตแมปแบบช่วง ในกรณีที่แอทริบิวต์ที่นำมาสร้างดัชนีมีค่าคาร์ดินอลิตี้น้อยกว่า 15

อย่างไรก็ตาม ในภาพรวม ดัชนีบิตแมปแบบกระจายจะมีประสิทธิภาพและเหมาะกับการสอบถามข้อมูลแบบค่าเท่ากันและแบบความเป็นสมาชิกมากที่สุด เมื่อเปรียบเทียบกับดัชนีบิตแมปแบบพื้นฐาน แบบ Range แบบช่วง และแบบเข้ารหัส ดังตาราง 4-2

ตาราง 4-2 สรุปคุณสมบัติที่สำคัญของดัชนีบิตแมปแบบพื้นฐาน แบบ Range แบบช่วง แบบเข้ารหัส และแบบกระจาย

ชนิดของดัชนีบิตแมป	จำนวนบิตแมปเวกเตอร์ที่ต้องใช้ในการแทนค่า 1 ค่าของแอทริบิวต์	จำนวนบิตแมปเวกเตอร์ที่ต้องใช้ในการแทนค่าของแอทริบิวต์ที่มีคาร์ดินอลิตี้เท่ากับ C	จำนวนค่าที่แทนโดยใช้ 1 บิตแมปเวกเตอร์
แบบพื้นฐาน	1	C	1
แบบ Range	$\leq C-1$	C-1	$\leq C-1$
แบบช่วง	$\leq \left\lfloor \frac{C}{2} \right\rfloor$	$\left\lceil \frac{C}{2} \right\rceil$	$\left\lfloor \frac{C}{2} \right\rfloor$
แบบเข้ารหัส	$\lceil \log_2 C \rceil$	$\lceil \log_2 C \rceil$	C
แบบกระจาย	2	$\lceil 2\sqrt{C} \rceil$	$\leq \lceil \sqrt{C}+1 \rceil$