บทที่ 6

บทสรุปและข้อเสนอแนะ

6.1 บทสรุป

ระบบคลังข้อมูลเป็นเครื่องมือสำคัญที่ช่วยในการสนับสนุนการตัดสินใจของผู้บริหาร โดยมีการรวบรวมข้อมูลจำนวนมหาศาลจากหลากหลายแหล่งมาเก็บไว้ในคลังข้อมูลอย่างเป็นระบบ เตรียมไว้สำหรับการสอบถามที่มีความซับซ้อนและเป็นแบบทันทีทันใด กลไกในการเข้าถึงข้อมูลจึงเป็นสิ่งสำคัญ การทำดัชนีบิตแมปเป็นวิธีหนึ่งที่ทำให้การเข้าถึงข้อมูลเป็นไปอย่างรวดเร็ว เนื่องจากเป็นวิธีการทำดัชนีที่อยู่บนพื้นฐานของบิตแมป จึงสนับสนุนการทำงานของฮาร์ดแวร์ ซึ่งที่ผ่านมามีการพัฒนาขั้นตอนการสร้างดัชนีบิตแมปต่อ ๆ กันมาเพื่อให้สอดคล้องกับความต้องการ และให้มีประสิทธิภาพมากยิ่งขึ้น

ในอดีตที่ผ่านมามีเทคนิคการทำดัชนีบิตแมปที่น่าสนใจ คือ ดัชนีบิตแมปแบบพื้นฐาน ดัชนีบิตแมปแบบช่วง ดัชนีบิตแมปแบบกระจาย และดัชนีบิตแมปแบบเข้ารหัส ซึ่งดัชนีบิตแมปแต่ละชนิดมีเทคนิคการสร้างดัชนีและการสอบถามแบบค่าเท่ากันที่ไม่เหมือนกัน จึงทำให้มีประสิทธิภาพในแต่ละด้านแตกต่างกันไป คือ ดัชนีบิตแมปแบบพื้นฐานมีประสิทธิภาพในด้านเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันดีที่สุด แต่ประสิทธิภาพในด้านพื้นที่ที่ใช้ในการจัดเก็บดัชนีแย่ที่สุด ส่วนดัชนีบิตแมปแบบเข้ารหัสมีประสิทธิภาพในด้านพื้นที่ที่ใช้ในการจัดเก็บดัชนีดีที่สุด แต่ประสิทธิภาพในด้านเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันแย่ที่สุด ส่วนดัชนีบิตแมปแบบช่วงและดัชนีบิตแมปแบบกระจายมีประสิทธิภาพในด้านเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันใกล้เคียงกัน แต่ดัชนีบิตแมปแบบกระจายมีประสิทธิภาพในด้านพื้นที่ที่ใช้ในการจัดเก็บดัชนีมากกว่าดัชนีบิตแมปแบบช่วง

ในงานวิทยานิพนธ์นี้ ได้นำเสนอขั้นตอนวิธีในการสร้างดัชนีบิตแมปแบบใหม่สำหรับการสอบถามแบบค่าเท่ากันจากคลังข้อมูล ดัชนีบิตแมปแบบใหม่นี้มีชื่อว่า ดัชนีบิตแมปแบบคู่กัน (Dual Bitmap Index) โดยใช้หลักการที่ว่า แต่ละค่าของแอทริบิวต์ถูกแทนด้วย 2 บิตแมปเวกเตอร์ และทำให้ทุกบิตแมปเวกเตอร์ถูกใช้อย่างคุ้มค่ามากที่สุด ซึ่งสามารถพิจารณาการเปรียบเทียบลักษณะที่สำคัญของดัชนีบิตแมปแบบคู่กันที่คิดค้นขึ้นกับดัชนีบิตแมปที่เคยมีมาได้ดังตาราง 6-1

ตาราง 6-1 แสดงลักษณะที่สำคัญของดัชนีบิตแมปทั้ง 5 ชนิด (C คือ จำนวนค่าที่เป็นไปได้ของแอทริบิวต์ที่เลือกมาทำดัชนี)

ลักษณะที่สำคัญ	ดัชนีบิตแมป				
	แบบพื้นฐาน	แบบช่วง	แบบกระจาย	แบบเข้ารหัส	แบบคู่กัน
จำนวนบิตแมปเวกเตอร์ที่ใช้ในการแทนแต่ละค่าของแอทริบิวต์	1	$\leq\lfloor C/2 \rfloor$	2	$\lceil \log_2 C \rceil$	2
จำนวนค่าของแอทริบิวต์ที่ถูกแทนด้วยแต่ละบิตแมปเวกเตอร์	1	$\lfloor C/2 \rfloor$	$\leq\lceil \sqrt{C} \rceil+1$	C	$\left\lceil \sqrt{2C+\frac{1}{4}}+\frac{1}{2} \right\rceil-1$
จำนวนบิตแมปเวกเตอร์ที่ใช้ในการแทนค่าของแอทริบิวต์ที่มีคาร์ดินอลิตี้ C (พื้นที่)	C	$\lceil C/2 \rceil$	$\lceil 2\sqrt{C} \rceil$	$\lceil \log_2 C \rceil$	$\left\lceil \sqrt{2C+\frac{1}{4}}+\frac{1}{2} \right\rceil$
จำนวนบิตแมปเวกเตอร์ที่อ่าน : จำนวนครั้งในการดำเนินการตรรกะสำหรับการสอบถามแบบค่าเท่ากัน (เวลา)	1:0	2:2 (1AND,1NOT)	2:1 (1AND)	$\lceil \log_2 C \rceil$: ใช้การเทียบค่าจากตารางการเทียบค่า	2:1 (1AND)

จากตาราง 6-1 จะเห็นได้ว่า ดัชนีบิตแมปแต่ละชนิดมีลักษณะบางประการที่แตกต่างกัน ส่งผลให้มีข้อดีข้อจำกัดที่แตกต่างกันไป ซึ่งดัชนีบิตแมปแต่ละชนิดถูกสร้างขึ้นมามีความเหมาะสมกับการสอบถามที่แตกต่างกัน เมื่อพิจารณาการสอบถามแบบค่าเท่ากัน พบว่าดัชนีบิตแมปแต่ละชนิดมีลักษณะที่เป็นจุดด้อย ดังนี้

- ดัชนีบิตแมปแบบพื้นฐานมีการใช้แต่ละบิตแมปเวกเตอร์ไม่คุ้มค่า คือ แต่ละบิตแมปเวกเตอร์ใช้แทนค่าของแอทริบิวต์ได้เพียงค่าเดียว ซึ่งถือว่าน้อยเกินไป

- ดัชนีบิตแมปแบบช่วงในบางกรณีมีการใช้จำนวนบิตแมปเวกเตอร์ในการแทนแต่ละค่าของแอทริบิวต์มากเกินไป (ใช้ $\leq\lfloor C/2 \rfloor$ บิตแมปเวกเตอร์) เพราะในการสอบถามแบบค่าเท่ากันมีการอ่านเพียง 2 บิตแมปเวกเตอร์เท่านั้น จึงควรใช้ 2 บิตแมปเวกเตอร์ในการแทนแต่ละค่าของแอทริบิวต์ก็เพียงพอ

- ดัชนีบิตแมปแบบกระจายมีบางบิตแมปเวกเตอร์ถูกใช้อย่างไม่คุ้มค่า โดยเฉพาะบิตแมปเวกเตอร์ Z^0 ใช้แทนค่าของแอทริบิวต์ได้เพียงค่าเดียวเท่านั้น

- ดัชนีบิตแมปแบบเข้ารหัสมีการใช้จำนวนบิตแมปเวกเตอร์ในการแทนแต่ละค่าของแอทริบิวต์มากเกินไป (ใช้ $\lceil \log_2 C \rceil$ บิตแมปเวกเตอร์)

- ดัชนีบิตแมปแบบคู่กันมีการใช้จำนวนบิตแมปเวกเตอร์ในการแทนแต่ละค่าของแอทริบิวต์มากกว่าดัชนีบิตแมปแบบพื้นฐาน คือ ดัชนีบิตแมปแบบคู่กันใช้ 2 บิตแมปเวกเตอร์ แต่ดัชนีบิตแมปแบบพื้นฐานใช้ 1 บิตแมปเวกเตอร์

ในการคิดค้นเทคนิคการสร้างดัชนีบิตแมป มีวัตถุประสงค์ในการนำไปใช้และความต้องการที่แตกต่างกัน ได้แก่ รูปแบบการสอบถามที่ต้องการ เช่น การสอบถามแบบค่าเท่ากัน การสอบถามแบบความเป็นสมาชิก และการสอบถามแบบเป็นช่วง เป็นต้น หรือความต้องการเกี่ยวประสิทธิภาพในแง่ต่าง ๆ เช่น ประสิทธิภาพในแง่เวลา ประสิทธิภาพในแง่พื้นที่ และประสิทธิภาพในแง่ Space-Time Trade-off เป็นต้น สำหรับในงานวิทยานิพนธ์นี้ เราระบุรูปแบบการสอบถามเป็นการสอบถามแบบค่าเท่ากัน จึงขอสรุปประสิทธิภาพที่ต้องการสำหรับการสอบถามแบบค่าเท่ากัน และดัชนีบิตแมปที่ควรเลือกใช้ ดังตาราง 6-2

ตาราง 6-2 สรุปประสิทธิภาพที่ต้องการสำหรับการสอบถามแบบค่าเท่ากันและดัชนีบิตแมปที่ควรเลือกใช้

ประสิทธิภาพที่ต้องการ	**ดัชนีบิตแมปที่ควรเลือกใช้**
ประสิทธิภาพในแง่ของพื้นที่	ดัชนีบิตแมปแบบเข้ารหัส
ประสิทธิภาพในแง่ของเวลา	ดัชนีบิตแมปแบบพื้นฐาน
ประสิทธิภาพในแง่ Space-Time Trade-off	ดัชนีบิตแมปแบบคู่กัน

จากผลการวิเคราะห์และผลการทดลองเปรียบเทียบดัชนีบิตแมปแบบคู่กันที่คิดค้นขึ้นกับดัชนีบิตแมปที่เคยมีมา พบว่า ดัชนีบิตแมปแบบพื้นฐานใช้เวลาในการสอบถามน้อยที่สุด แต่กลับต้องสูญเสียพื้นที่ในการจัดเก็บดัชนีมากที่สุด จึงเหมาะสำหรับการสร้างดัชนีบนแอทริบิวต์ที่มีคาร์ดินอลิตี้ต่ำ ๆ และต้องการประสิทธิภาพในแง่ของเวลา ส่วนดัชนีบิตแมปแบบเข้ารหัสประหยัดพื้นที่ในการจัดเก็บดัชนีมากที่สุด แต่เวลาในการสอบถามช้าที่สุด จึงเหมาะสำหรับการสร้างดัชนีบนแอทริบิวต์ที่มีคาร์ดินอลิตี้สูง ๆ และต้องการประสิทธิภาพในแง่ของพื้นที่ นอกจากนั้นยังมีอีกทางเลือกหนึ่งในการทำดัชนี คือ ดัชนีบิตแมปแบบคู่กัน ซึ่งสามารถประหยัดพื้นที่ได้มาก แต่ยังคงมีประสิทธิภาพในการสอบถามอยู่ เนื่องจากใช้เพียงการดำเนินการทางตรรกะในการสอบถามเท่านั้น เหมาะสำหรับการสร้างดัชนีบนแอทริบิวต์ที่มีคาร์ดินอลิตี้สูง ๆ และต้องการประสิทธิภาพในแง่ Space-Time Trade-off ส่วนดัชนีบิตแมปแบบช่วงและดัชนีบิตแมปแบบกระจายเหมาะสำหรับการสอบถามแบบอื่นมากกว่า จากที่กล่าวมาทั้งหมด จะเห็นได้

ว่า ดัชนีบิตแมปแต่ละชนิดมีข้อดี และข้อจำกัดที่แตกต่างกันไป การนำไปใช้จึงควรเลือกให้เหมาะสมกับความต้องการ

6.2 ข้อเสนอแนะและงานในอนาคต

เมื่อมีการสอบถามแบบความเป็นสมาชิก เราสามารถดำเนินการบนดัชนีบิตแมปแบบคู่กันได้ด้วยสอบถามแบบค่าเท่ากันเพื่อหาคำตอบของแต่ละค่าที่สอบถาม แล้วคำตอบที่ได้จะมาจากการนำบิตแมปเวกเตอร์ผลลัพธ์ทั้งหมดมาดำเนินการตรรกะ OR ระหว่างบิตแมปเวกเตอร์จนครบทุกบิตแมปเวกเตอร์ ตัวอย่างเช่น ต้องการสอบถาม $"A=\{1,6\}"$ ซึ่งเมื่อพิจารณาการสอบถามแบบค่าเท่ากันของค่า 1 และ 6 จะได้ว่า $"A=1"=D^2 \wedge D^0$ และ $"A=6"=D^4 \wedge D^0$ ดังนั้นบิตแมปเวกเตอร์ที่ต้องอ่าน คือ D^0, D^2 และ D^4 จากนั้นดำเนินการตรรกะระหว่างบิตแมปเวกเตอร์ดังนี้

1) ดำเนินการตรรกะ AND ระหว่างบิตแมปเวกเตอร์ D^2 และ D^0

2) ดำเนินการตรรกะ AND ระหว่างบิตแมปเวกเตอร์ D^4 และ D^0

3) ดำเนินการตรรกะ OR ระหว่างบิตแมปเวกเตอร์ผลลัพธ์ในข้อ 1 และ 2 จะได้ว่า บิตแมปเวกเตอร์ในข้อ 3 คือ บิตแมปเวกเตอร์ผลลัพธ์ของการสอบถามที่ต้องการ

นอกจากนั้น กรณีที่มีการสอบถามแบบความเป็นสมาชิกที่ประกอบด้วยการถามค่าของแอทริบิวต์ 2 ค่า ที่มีบิตแมปเวกเตอร์ที่ต้องอ่านซ้ำกัน แล้วเราสามารถลดขั้นตอนได้ด้วยการดำเนินการตรรกะ OR ระหว่างบิตแมปเวกเตอร์ที่ไม่ซ้ำกันก่อน แล้วนำผลลัพธ์ที่ได้มาดำเนินการตรรกะ AND กับบิตแมปเวกเตอร์ที่ไม่ซ้ำ จากตัวอย่างเดิม ต้องการสอบถาม $"A=\{1,6\}"$ เมื่ออ่านบิตแมปเวกเตอร์ D^0, D^2 และ D^4 แล้วดำเนินการตรรกะระหว่างบิตแมปเวกเตอร์ดังนี้

1) ดำเนินการตรรกะ OR ระหว่างบิตแมปเวกเตอร์ D^2 และ D^4

2) ดำเนินการตรรกะ AND ระหว่างบิตแมปเวกเตอร์ผลลัพธ์ในข้อ 1 กับ D^0

จะได้ว่า บิตแมปเวกเตอร์ในข้อ 2) คือ บิตแมปเวกเตอร์ผลลัพธ์ของการสอบถามที่ต้องการ ซึ่งเห็นได้ว่าวิธีหลังดำเนินการตรรกะระหว่างบิตแมปเวกเตอร์น้อยกว่าวิธีแรก เหตุผลที่เราสามารถใช้วิธีที่สองในการหาคำตอบได้ เนื่องจากรูปแบบการดำเนินการตรรกะเพื่อหาคำตอบในวิธีแรกสามารถลดรูปได้ กล่าวคือ $(a \wedge c) \vee (b \wedge c) = (a \vee b) \wedge c$ นั่นเอง

อย่างไรก็ตามเมื่อมีการสอบถามแบบความเป็นสมาชิกที่ประกอบด้วยการถามค่าของแอทริบิวต์มากกว่า 2 ค่า ยังไม่สามารถหาวิธีในการลดขั้นตอนดำเนินการตรรกะระหว่างบิตแมปเวกเตอร์ได้ ซึ่งเป็นงานที่วางแผนจะทำต่อในอนาคต