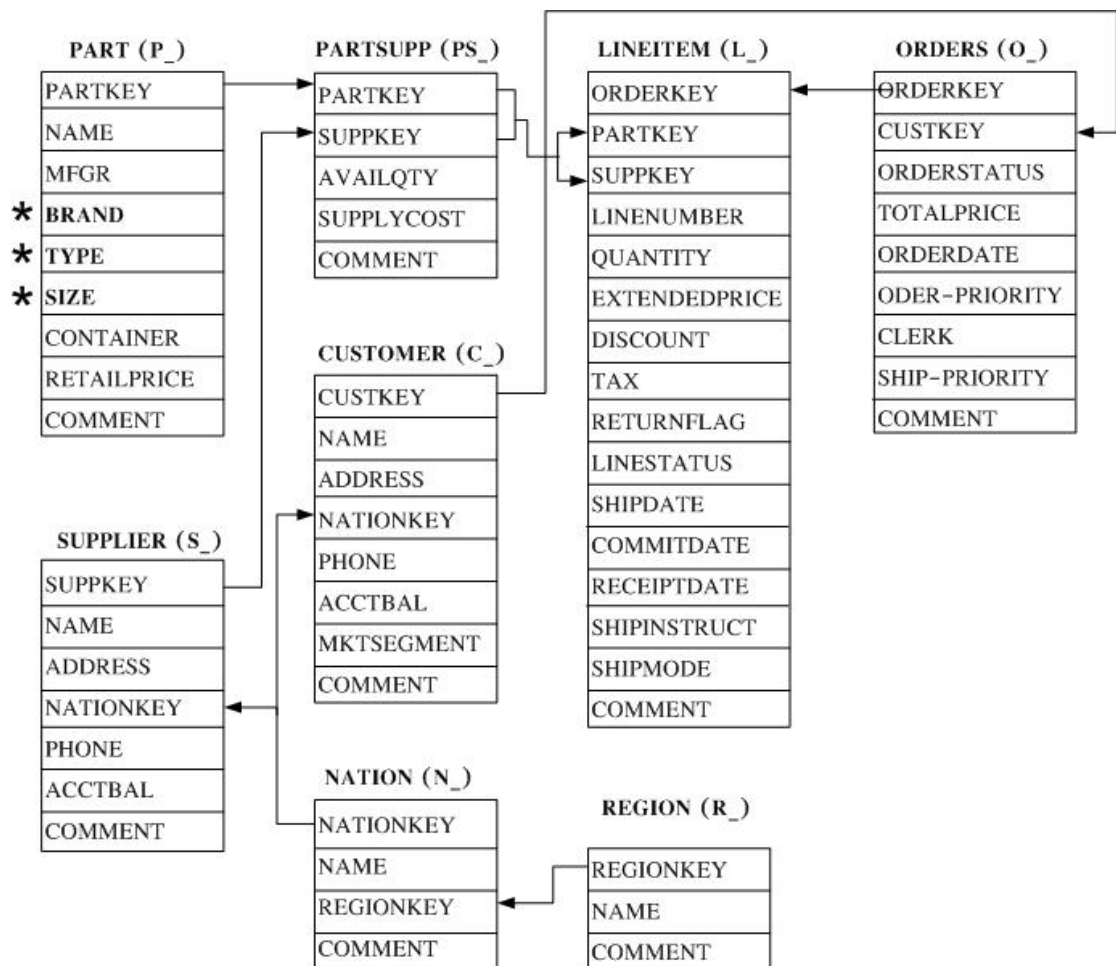


ภาคผนวก

ก. การเตรียมข้อมูลเพื่อใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลมาตรฐานจาก TPC-H Benchmark [25] ซึ่งเป็นตัววัดเปรียบเทียบประสิทธิภาพการประมวลผลการสอบถามที่ซับซ้อนในระบบสนับสนุนการตัดสินใจ (Decision Support) มีโครงสร้าง ดังนี้



ข้อมูลที่ใช้ในการทดลองมี 3 ชุด คือ

- ข้อมูลชุดที่ 1 แอทริบิวต์ BRAND บน PART Table
มี 1,000,000 แถว คาร์ดินอลิตี้เท่ากับ 25 ($C = 25$)
- ข้อมูลชุดที่ 2 แอทริบิวต์ SIZE บน PART Table
มี 1,000,000 แถว คาร์ดินอลิตี้เท่ากับ 50 ($C = 50$)
- ข้อมูลชุดที่ 3 แอทริบิวต์ TYPE บน PART Table
มี 1,000,000 แถว คาร์ดินอลิตี้เท่ากับ 150 ($C = 150$)

สำหรับแต่ละแอทริบิวต์ที่เลือกมาสร้างดัชนีบิตแมป จะต้องมีการเตรียมข้อมูลตามขั้นตอนต่อไปนี้

- 1) เลือกเฉพาะแอทริบิวต์ที่จะนำมาสร้างดัชนีโดยใช้คำสั่ง

```
awk -F\| '{print $number}' input_file > output_file
```

โดยที่ | คือ อักขระที่ใช้คั่นแต่ละค่าของแอทริบิวต์, number คือ แอทริบิวต์ที่ต้องการ, input_file คือ ชื่อไฟล์ตารางที่มีแอทริบิวต์ที่ต้องการ และ output_file คือ ชื่อไฟล์ที่ใช้เก็บข้อมูลของแอทริบิวต์ที่เลือก

ตัวอย่างเช่น `awk -F\| '{print $2}' part.tbl > attr_nation.txt` เป็นการเลือกแอทริบิวต์ที่ 2 จาก part.tbl เก็บไว้ใน attr_nation.txt

- 2) เปลี่ยนรูปค่าของแอทริบิวต์ให้เป็นจำนวนเต็มต่อเนื่องกัน เริ่มตั้งแต่ค่า 0 ดังนั้น ค่าของแอทริบิวต์ที่เปลี่ยนรูปแล้ว คือ 0,1,2,..., $C-1$ โดยการสร้าง file.awk ซึ่งเขียนเป็นชุดคำสั่งสำหรับเปลี่ยนค่าในไฟล์หนึ่งแล้วเก็บไว้ในอีกไฟล์หนึ่ง

ตัวอย่างชุดคำสั่งใน file.awk

```
/THAI/ {print 0}      (เปลี่ยนคำว่า THAI เป็น 0)
/CHINESE/ {print 1}  (เปลี่ยนคำว่า CHINESE เป็น 1)
/JAPANESE/ {print 2} (เปลี่ยนคำว่า JAPANESE เป็น 2)
```

จากนั้นทำให้ชุดคำสั่งใน file.awk ทำงาน โดยใช้คำสั่ง

```
awk -f file.awk input_file > output_file
```

โดยที่ file.awk คือ ไฟล์ชุดคำสั่ง, input_file คือ ชื่อไฟล์ที่ใช้เก็บข้อมูลของแอทริบิวต์ที่เลือก, output_file คือ ชื่อไฟล์ที่ใช้เก็บข้อมูลของแอทริบิวต์ที่เลือกที่มีการเปลี่ยนรูปเป็นจำนวนเต็มต่อเนื่องกันแล้ว

ข. การเขียนโปรแกรมเพื่อทดลองสอบถามแบบค่าเท่ากันบนดัชนีบิตแมป

ดำเนินการเขียนโปรแกรมด้วยตัวแปลภาษาซี (C Compiler) โดยมีการกำหนดค่าที่นิยาม, ตัวแปร, ฟังก์ชัน และไฟล์ที่สำคัญดังต่อไปนี้

○ การกำหนดค่า (define)

CADINALLITY 50 เป็นการกำหนดค่าให้กับคาร์ดินอลิตี้ ในที่นี้เท่ากับ 50

NUM_RECORD 1000000 เป็นการกำหนดจำนวนแถวข้อมูล ในที่นี้เท่ากับ 1000000

○ ตัวแปร (variable)

value ใช้สำหรับเก็บค่าที่ต้องการสอบถาม ($0 \leq v \leq C-1$)

vector_use ใช้สำหรับเก็บค่าลำดับของบิตแมปเวกเตอร์ที่ต้องการอ่านมาเพื่อดำเนินการตรรกะ

bitsvector ใช้สำหรับเก็บค่าของบิตแมปเวกเตอร์ที่อ่านได้จากตารางดัชนี

ansvector ใช้สำหรับเก็บค่าของบิตแมปเวกเตอร์ผลลัพธ์จากการดำเนินการตรรกะ ก่อนนำเขีนลงในไฟล์คำตอบของการสอบถาม

○ ฟังก์ชัน (function)

QuerySimple(value) เป็นฟังก์ชันที่ดำเนินการสอบถามบนดัชนีบิตแมปแบบพื้นฐาน

QueryInterval(value) เป็นฟังก์ชันที่ดำเนินการสอบถามบนดัชนีบิตแมปแบบช่วง

QueryScatter(value) เป็นฟังก์ชันที่ดำเนินการสอบถามบนดัชนีบิตแมปแบบกระจาย

QueryEncoded(value) เป็นฟังก์ชันที่ดำเนินการสอบถามบนดัชนีบิตแมปแบบเข้ารหัส

QueryDual(value) เป็นฟังก์ชันที่ดำเนินการสอบถามบนดัชนีบิตแมปแบบคู่กัน

WordPerVector(value) เป็นฟังก์ชันที่ดำเนินการหาจำนวน word ของความยาว 1 บิตแมปเวกเตอร์ ซึ่งในที่นี้เราให้ 1 word มีค่าเท่ากับ 32 (เท่ากับขนาดของ integer ในตัวแปลภาษาซีที่ใช้)

○ ไฟล์ (file)

simple.dat เป็นไฟล์ที่เก็บดัชนีบิตแมปแบบพื้นฐาน

interval.dat เป็นไฟล์ที่เก็บดัชนีบิตแมปแบบช่วง

scatter.dat เป็นไฟล์ที่เก็บดัชนีบิตแมปแบบกระจาย

encoded.dat เป็นไฟล์ที่เก็บดัชนีบิตแมปแบบเข้ารหัส

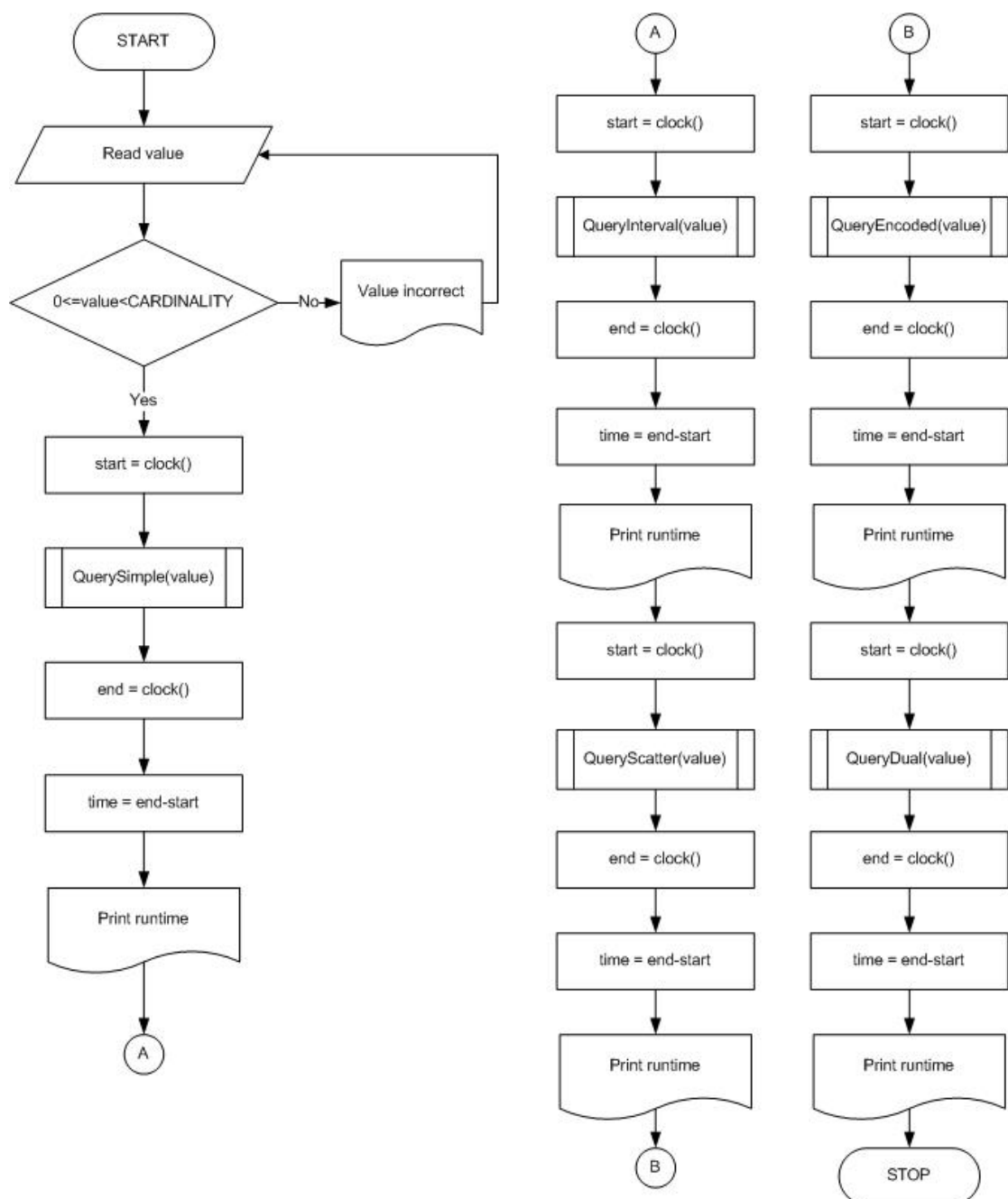
dual.dat เป็นไฟล์ที่เก็บดัชนีบิตแมปแบบคู่กัน

ans_simple.dat เป็นไฟล์ที่เก็บคำตอบของคำถามบนดัชนีบิตแมปแบบพื้นฐาน

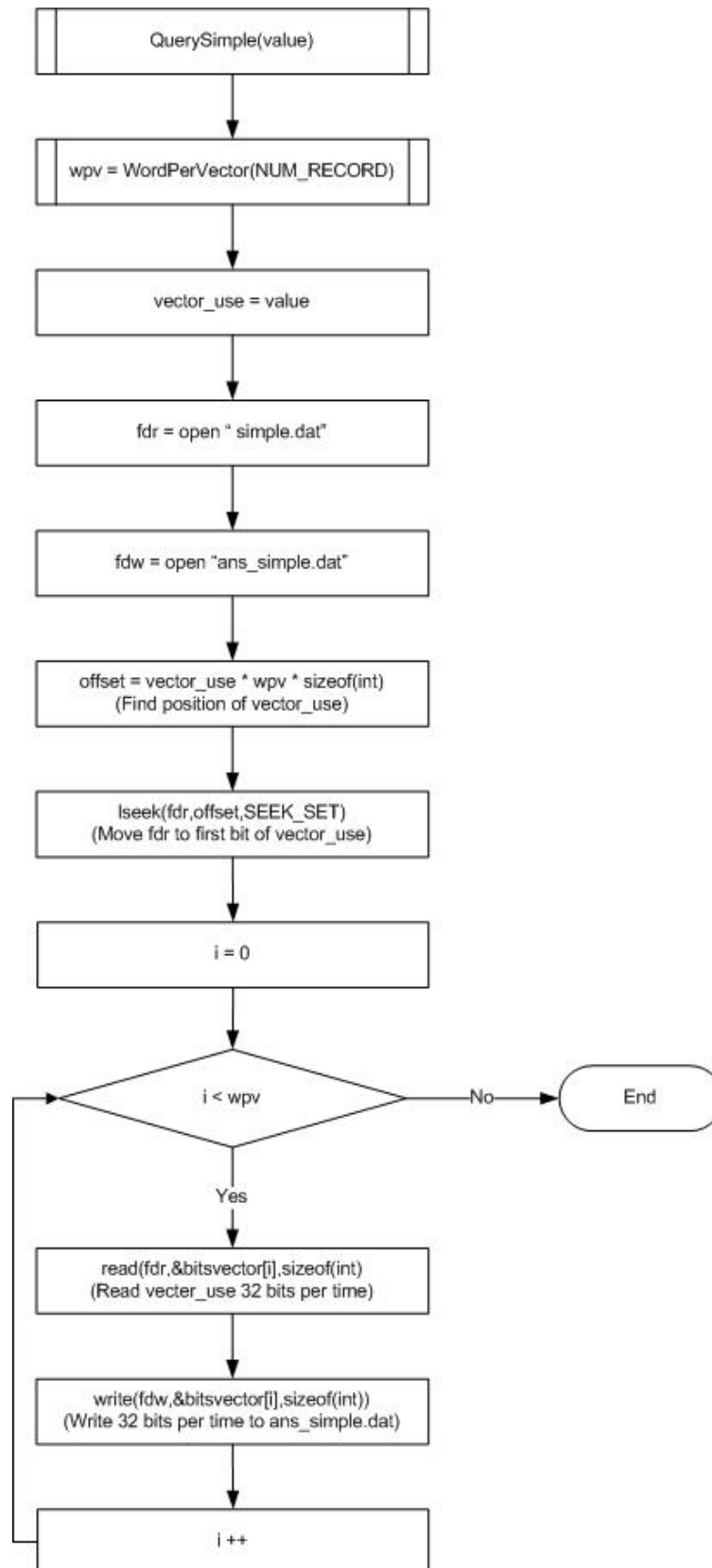
ans_interval.dat เป็นไฟล์ที่เก็บคำตอบของคำถามบนดัชนีบีตแมปแบบช่วง
 ans_scatter.dat เป็นไฟล์ที่เก็บคำตอบของคำถามบนดัชนีบีตแมปแบบกระจาย
 ans_encoded.dat เป็นไฟล์ที่เก็บคำตอบของคำถามบนดัชนีบีตแมปแบบเข้ารหัส
 ans_dual.dat เป็นไฟล์ที่เก็บคำตอบของคำถามบนดัชนีบีตแมปแบบคู่กัน

สำหรับขั้นตอนการทำงานของโปรแกรมเพื่อใช้ในการสอบถามมีดังต่อไปนี้

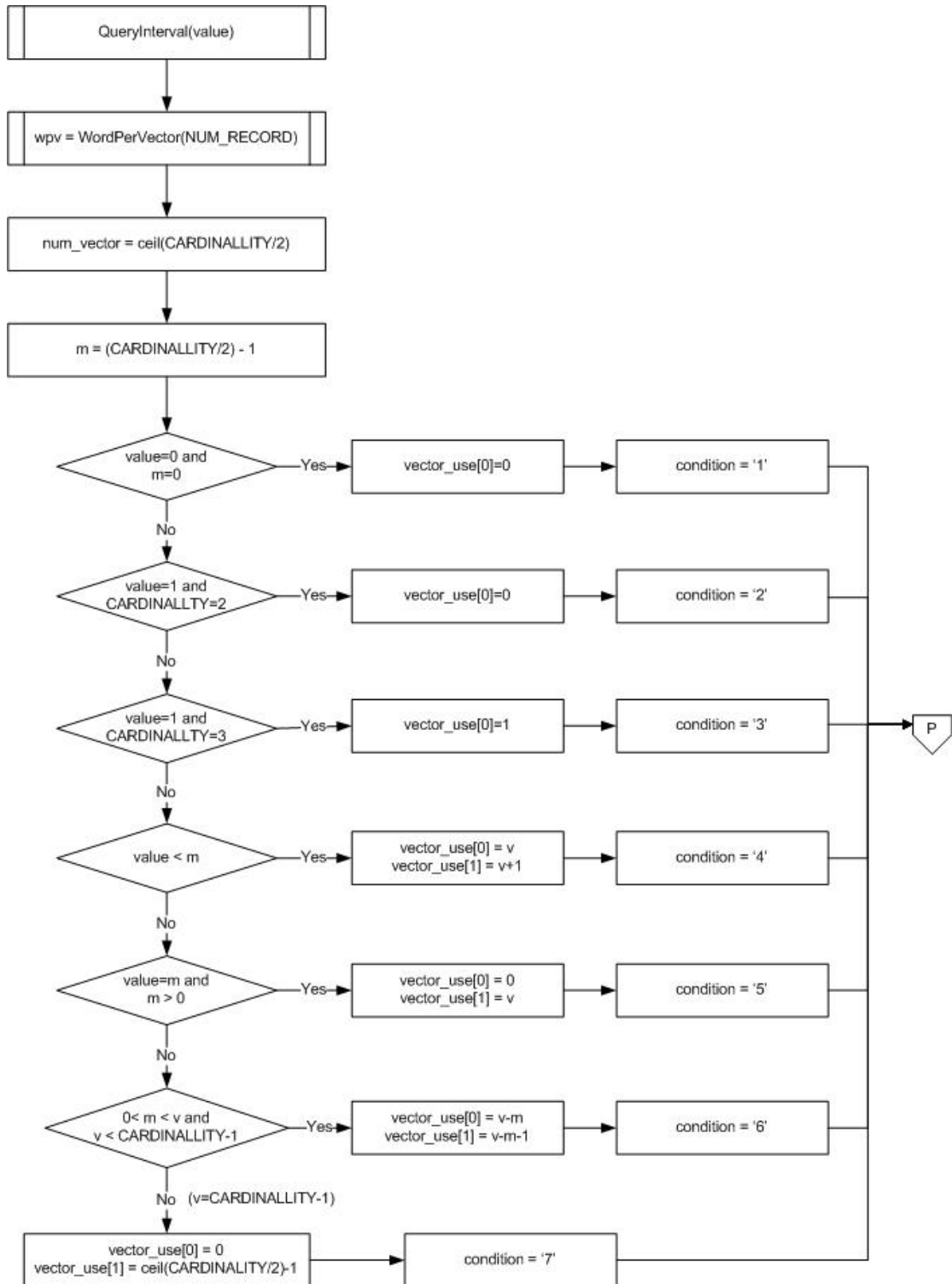
ข.1 ขั้นตอนวิธีทำงานหลักของการสอบถามบนดัชนีบีตแมปทั้ง 5 แบบ

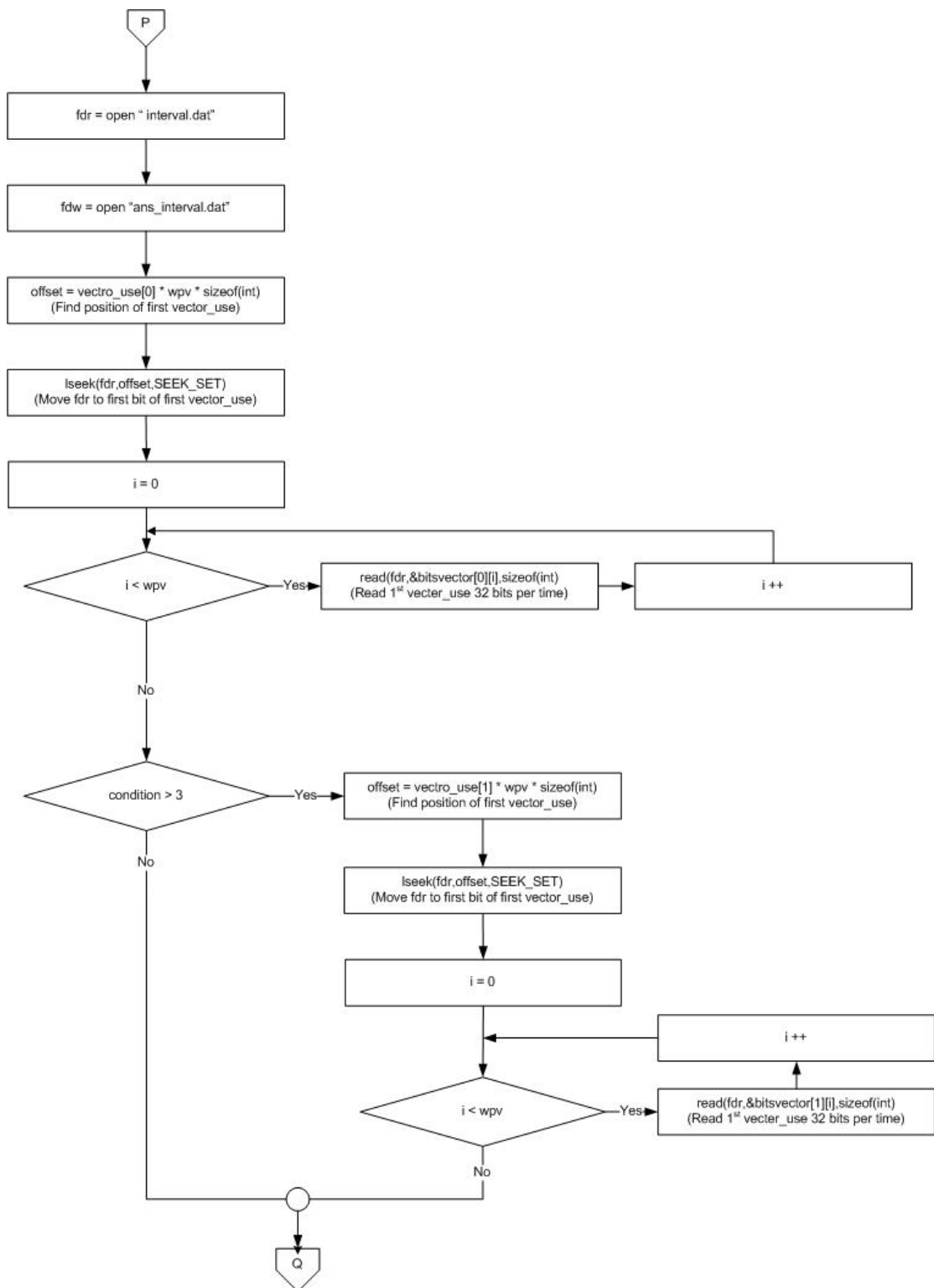


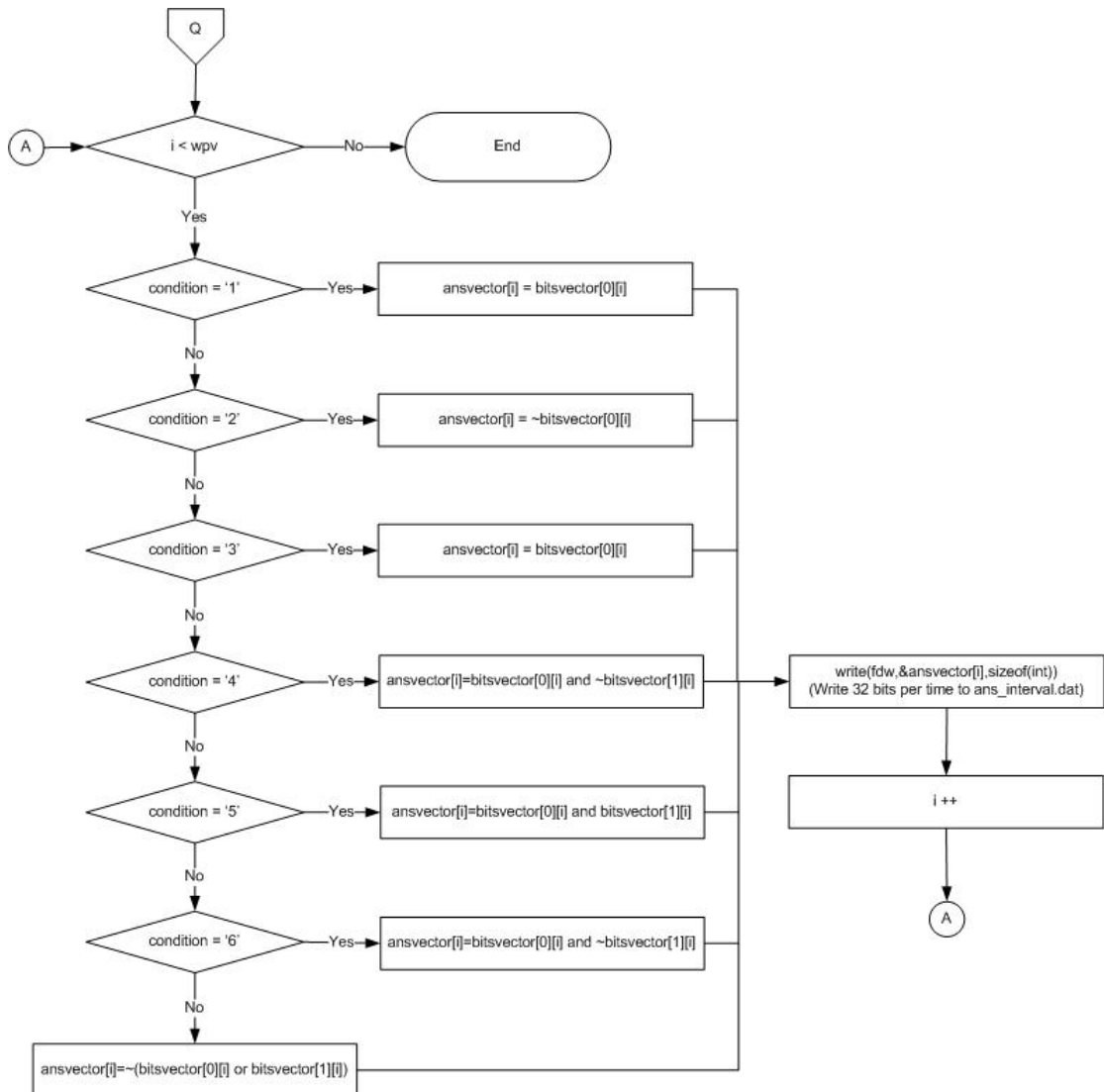
ข.2 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบพื้นฐาน



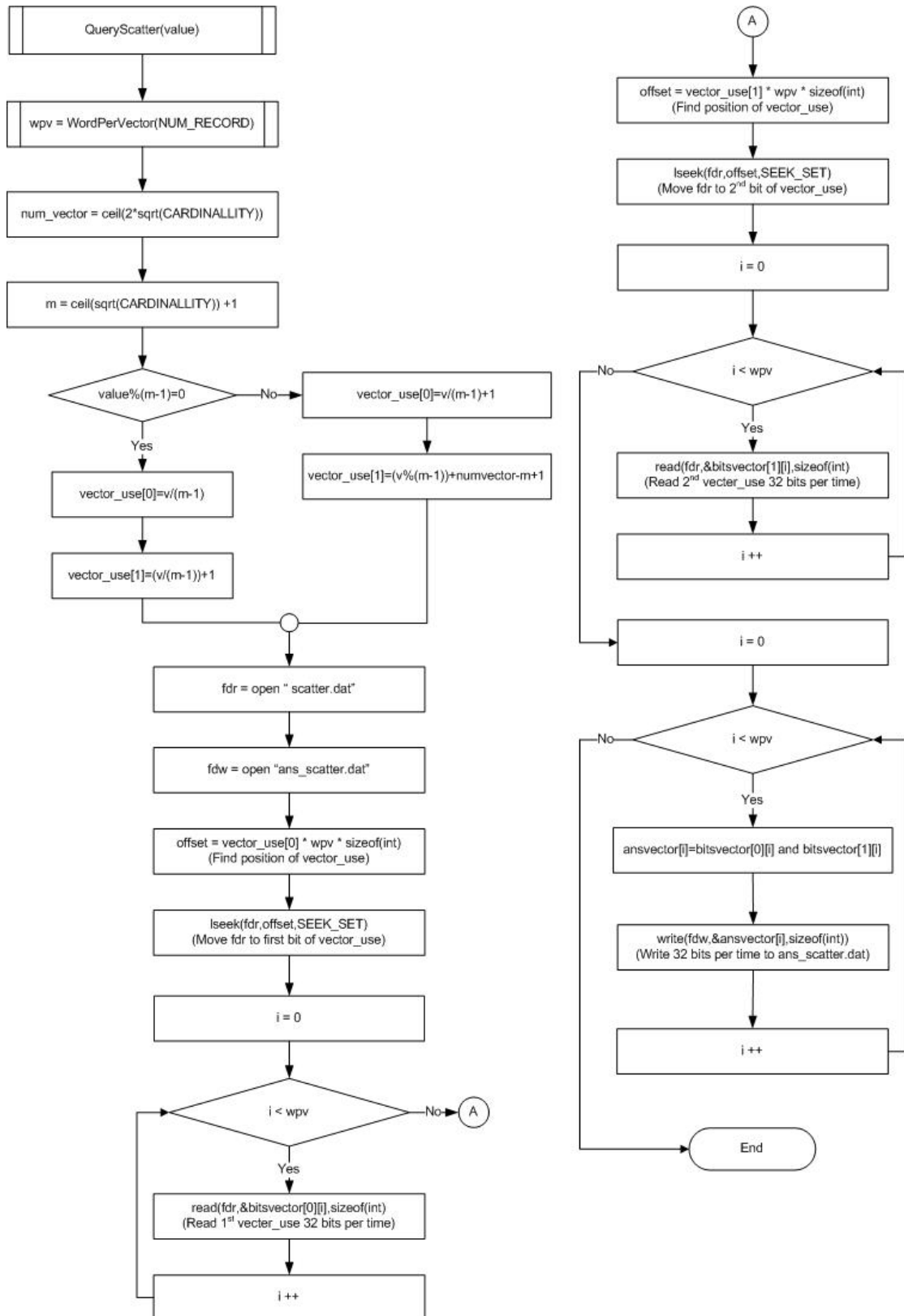
ข.3 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบช่วง



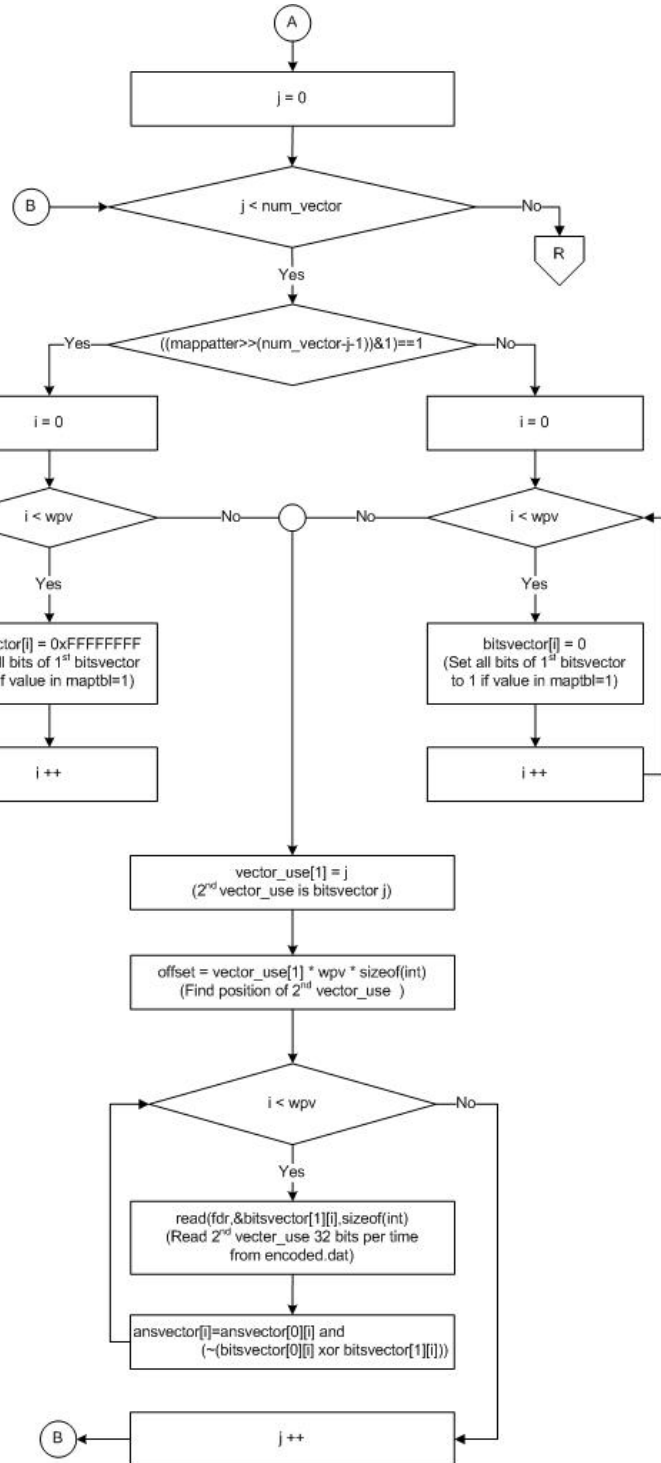
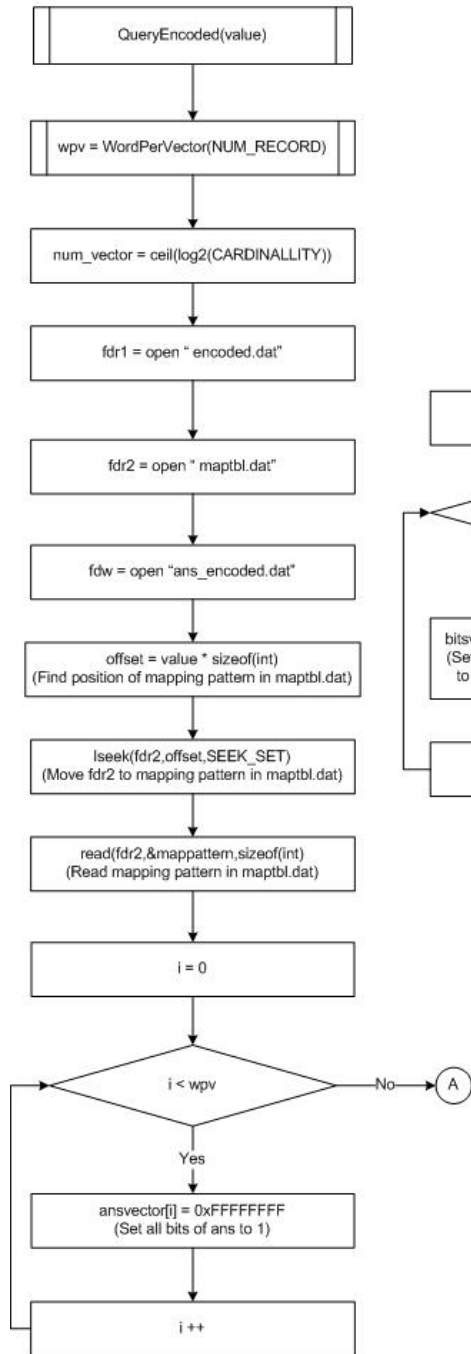


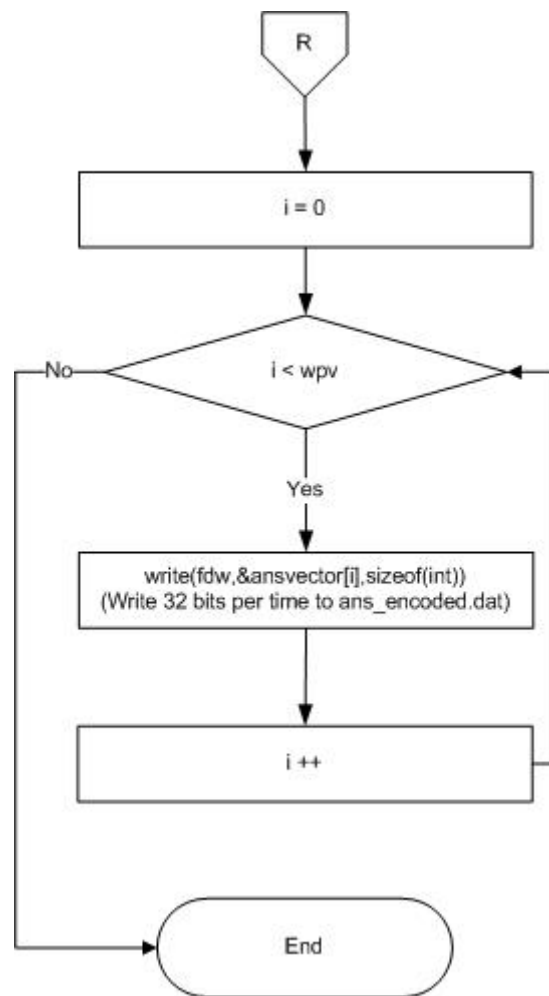


ข.4 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบกระจาย



ข.5 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบเข้ารหัส





ข.6 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบคู่กัน

