

บทที่ 6

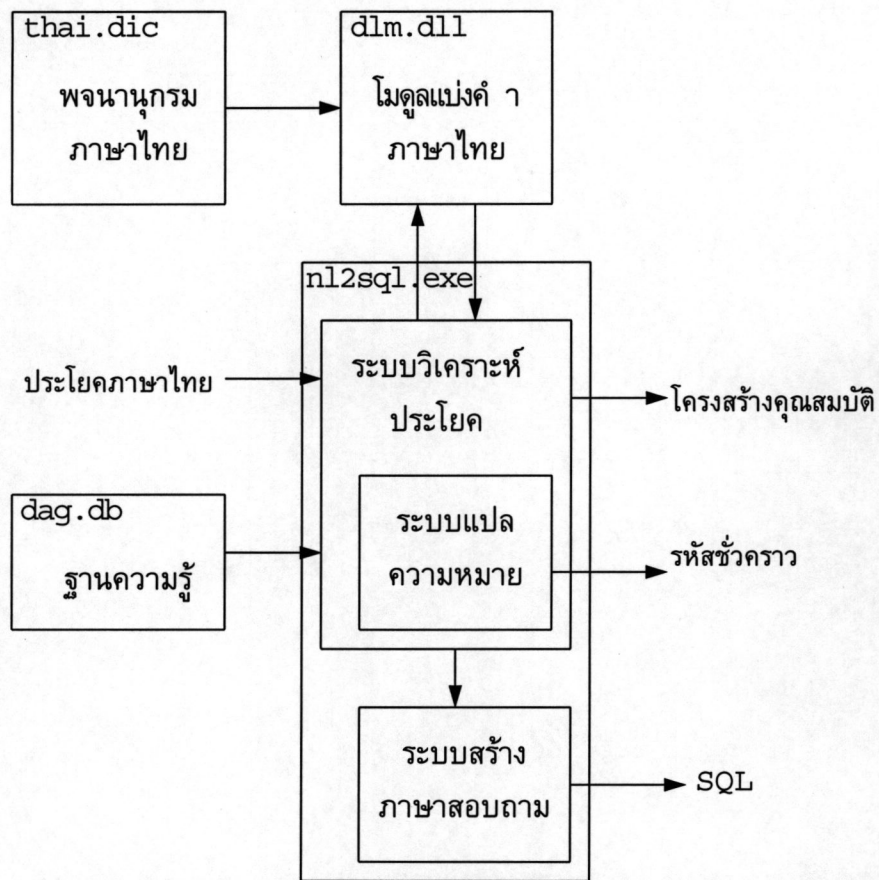
การทำงานของโปรแกรม

ในบทนี้จะกล่าวถึงส่วนประกอบต่างๆของโปรแกรม ขั้นตอนการทำงานของแต่ละส่วน และอัลกอริทึมในการสร้างภาษาสอบถาม โดยได้มีการอธิบายถึงการประยุกต์ใช้หลักการและทฤษฎีพร้อมไปกับการแสดงผลที่ได้จากการประมวลผลประโยคในรูปแบบต่างๆ ทั้งนี้ได้มีการดัดแปลงระบบจากรูปที่ 2.1 แต่โดยหลักแล้วโปรแกรมที่สร้างขึ้นมาจะทำการประมวลผลประโยคภาษาไทย โดยทำการวิเคราะห์ประโยคและสร้างภาษาสอบถามเอสคิวแอล นอกจากนี้ยังสามารถแสดงผลที่ได้จากการวิเคราะห์ คือ โครงสร้างคุณสมบัติ และรหัสชั่วคราว เพื่อเป็นการตรวจสอบว่าการทำงานมีข้อผิดพลาดหรือถูกต้องตามที่คาดการณ์ไว้หรือไม่

ส่วนประกอบของโปรแกรม

มีแฟ้มที่ใช้กับโปรแกรมทั้งหมด 4 แฟ้ม ดังนี้

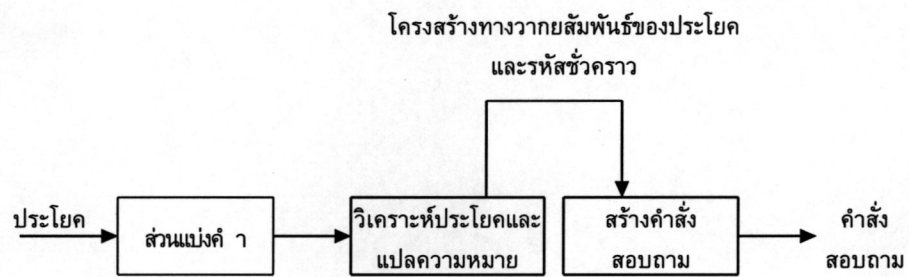
1. nl2sql.exe เป็นแฟ้มที่อยู่ของโปรแกรมหลัก ประกอบด้วยระบบวิเคราะห์ประโยคภาษาไทย ระบบแปลความหมาย และระบบสร้างภาษาสอบถาม
2. dlm.dll เป็นแฟ้มที่อยู่ของโมดูลแบ่งคำด้วยพจนานุกรม โมดูลนี้ถูกเรียกใช้โดยระบบวิเคราะห์ประโยคใน nl2sql.exe
3. thai.dic เป็นพจนานุกรมภาษาไทยที่เก็บไว้ในรูปของโครงสร้าง trie ซึ่งจะถูกบรรจุไว้ในหน่วยความจำเมื่อโปรแกรมเริ่มต้นทำงาน
4. dag.db เป็นแฟ้มที่เก็บข้อมูลและความหมายของคำ ข้อมูลเหล่านี้จะไม่ถูกบรรจุไว้ในหน่วยความจำทั้งหมด แต่จะมีการหาดัชนี (Index) ตำแหน่งของข้อมูลเหล่านี้ไว้ก่อนเพื่อสะดวกต่อการค้นหา



รูปที่ 6.1 แสดงหน้าที่ของแต่ละแฟ้มข้อมูลของโปรแกรม

ส่วนของการแบ่งคำในแฟ้ม dlm.dll ได้สร้างตามหลักการการแบ่งคำด้วยพจนานุกรมที่ใช้ trie ในการเก็บคำ และมีตัวรู้จำสถานะจำกัดในการช่วยตัดสินใจในการแบ่งคำ ซึ่งตัวแบ่งคำที่ใช้สามารถแบ่งคำได้อย่างถูกต้องเมื่อคำคำนั้นมีอยู่ในพจนานุกรม เนื่องจากประโยคที่ใช้ไม่มีคำที่เป็นข้อจำกัดของการแบ่งคำด้วยวิธีนี้

ส่วนวิเคราะห์ประโยคในแฟ้ม nl2sql.exe ได้มีการเปลี่ยนแปลงระบบจากรูปที่ 2.1 ที่มีส่วนของการแปลความหมายแยกออกมาต่างหากมาผนวกเข้ากับส่วนวิเคราะห์ประโยค ดังรูปที่ 6.2 ทั้งนี้เนื่องจากการนำเสนอข้อมูลที่ใช้อ้างอิงกับระบบฐานข้อมูลมีลักษณะเป็นโครงสร้างคุณสมบัติแบบเดียวกับโครงสร้างของคลังคำ ฉะนั้นจึงทำการผนวกเข้าไปในส่วนของการวิเคราะห์ประโยค เพื่อให้การประมวลผลเร็วขึ้น เนื่องจากไม่จำเป็นต้องสร้างโมดูลในการวิเคราะห์ที่มีการเคลื่อน (traverse) ไปตามโครงสร้างองค์ประกอบของประโยคอีกครั้งหนึ่ง

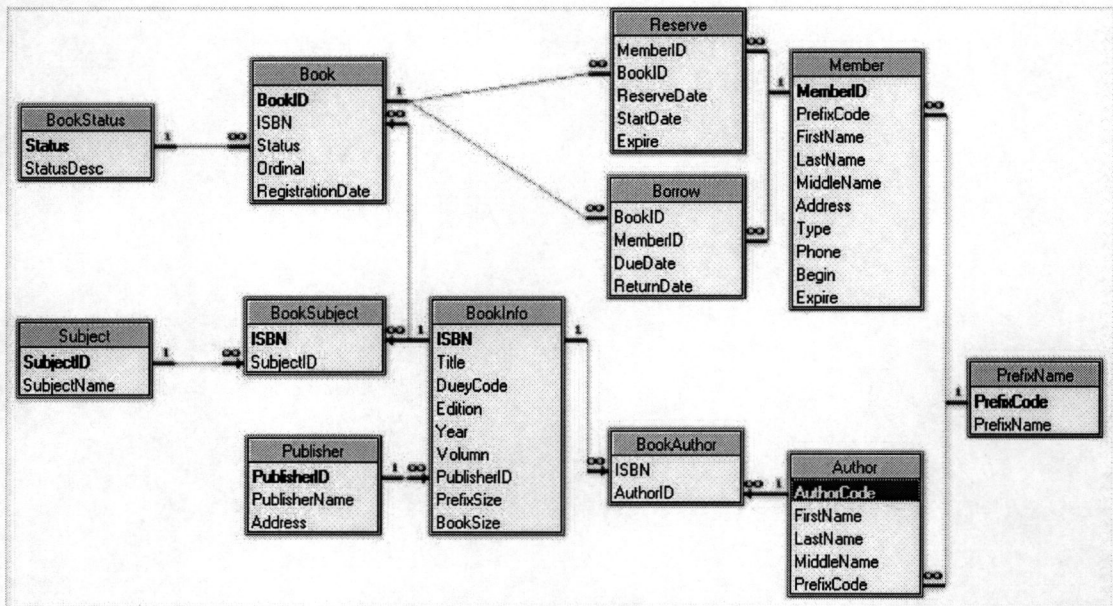


รูปที่ 6.2 โครงสร้างของระบบหลังจากการดัดแปลงแล้ว

ขั้นตอนการทำงาน

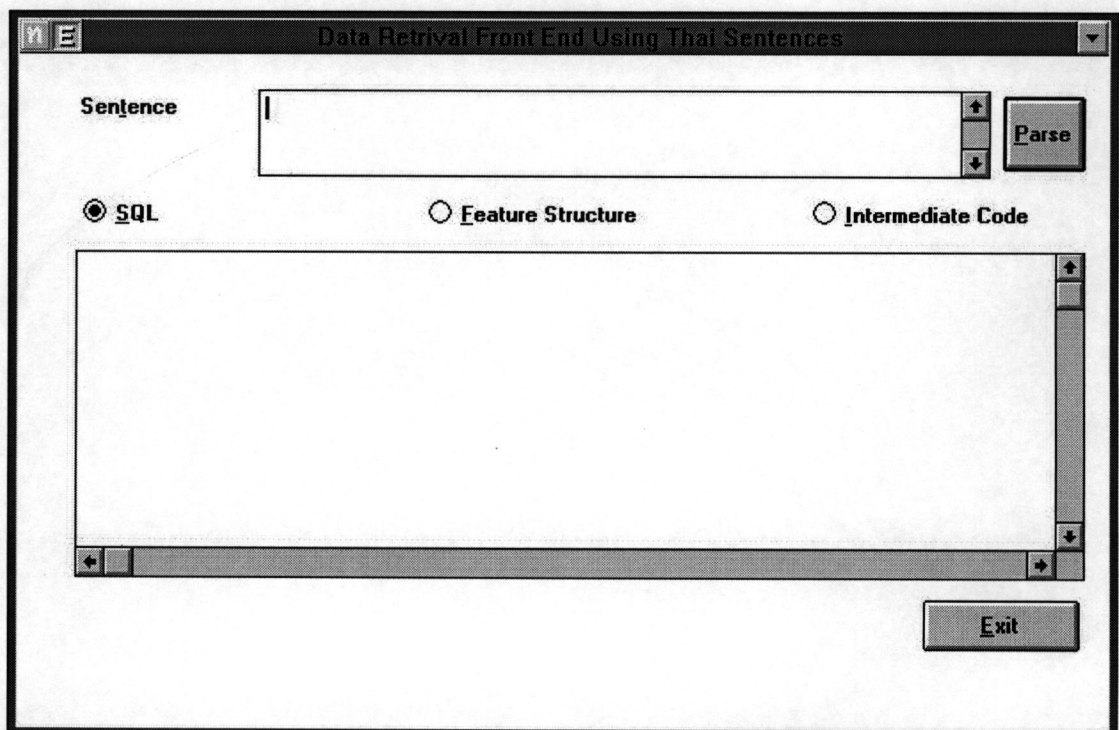
ในงานวิจัยนี้ได้ทำการจำลองระบบฐานข้อมูลของห้องสมุดมาใช้เป็นระบบฐานข้อมูลตัวอย่าง โดยมีรูปแบบข้อมูลและความสัมพันธ์ดังรูปที่ 6.3 ซึ่งมีตารางข้อมูลดังนี้

- Book ข้อมูลของหนังสือแต่ละเล่ม
- BookStatus คำอธิบายของสถานะหนังสือ
- BookInfo รายละเอียดของหนังสือ
- BookSubject หมวดของหนังสือ
- Subject ชื่อของหมวดหนังสือ
- Publisher รายละเอียดของผู้พิมพ์ หรือสำนักพิมพ์
- BookAuthor ผู้แต่งหนังสือแต่ละเรื่อง
- Author รายละเอียดของผู้แต่ง
- Member รายละเอียดของสมาชิกห้องสมุด
- PrefixName รายชื่อคำนำหน้าชื่อ
- Borrow ข้อมูลการยืมหนังสือ
- Reserve ข้อมูลการจองหนังสือ



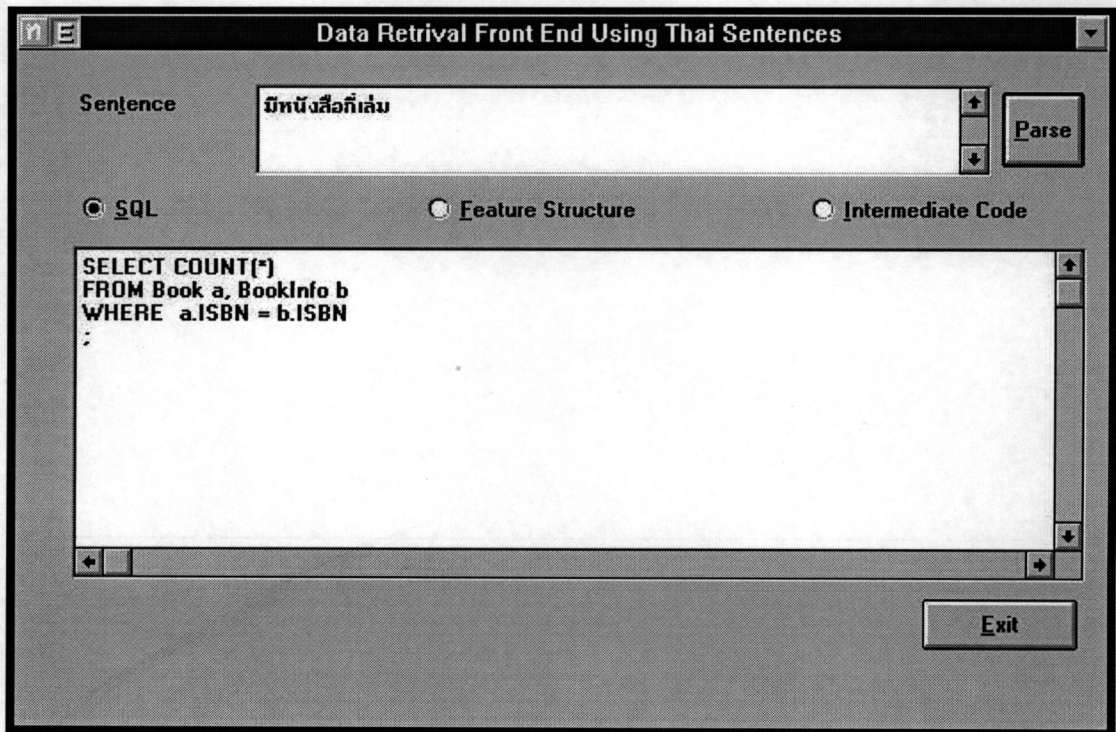
รูปที่ 6.3 แสดงความสัมพันธ์ของฐานข้อมูลจำลองของห้องสมุด

เมื่อเริ่มต้นโปรแกรมจะทำการเรียกใช้ส่วนแบ่งคำให้ทำการดึงพจนานุกรมที่อยู่ในรูปแบบโครงสร้าง trie ขึ้นมาไว้ในหน่วยความจำ และทำการสร้างดัชนีสำหรับการค้นคืนคีย์คำในแฟ้มที่ชื่อ dag.db ซึ่งเปรียบเสมือนฐานข้อมูลคีย์คำที่ใช้ในการวิเคราะห์ประโยค หลังจากนั้นตัวโปรแกรมก็พร้อมที่รับประโยคที่จะนำมาแปลงเป็นเอสคิวแอล ซึ่งมีหน้าตาดังรูปที่ 6.4

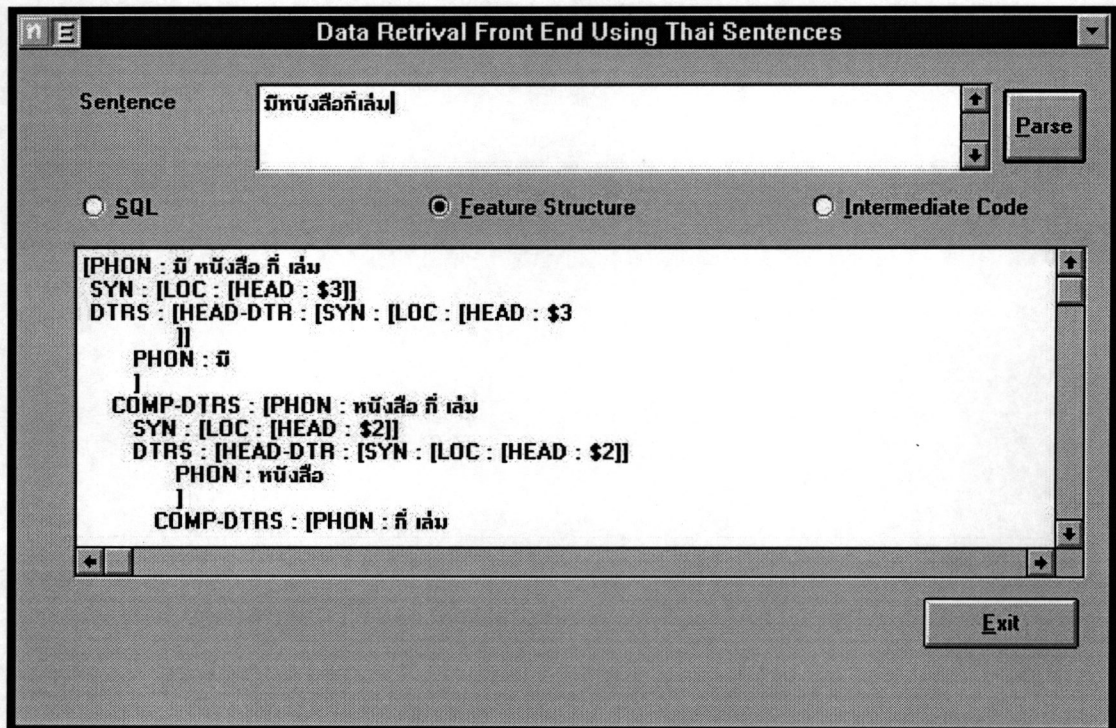


รูปที่ 6.4 แสดงหน้าจอของโปรแกรมเมื่อเริ่มต้นทำงาน

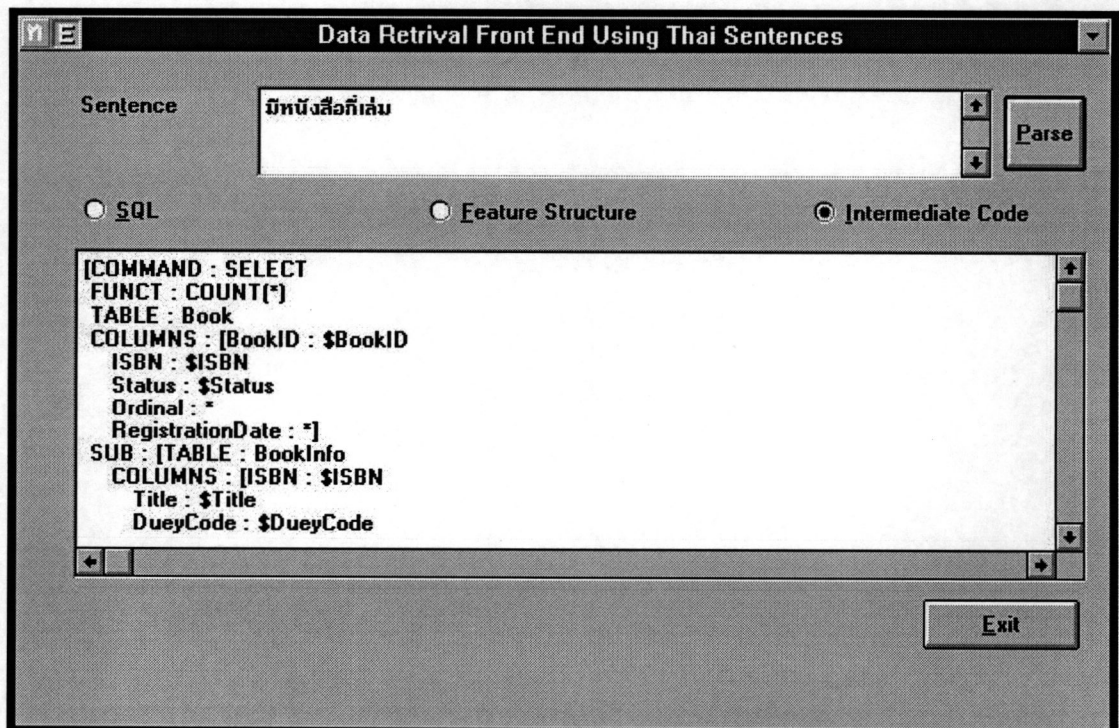
มีส่วนนำเข้าประโยคภาษาไทยอยู่ที่ส่วนของ Sentence ซึ่งเมื่อพิมพ์ประโยคที่ต้องการแปลงเสร็จเรียบร้อยแล้วกดปุ่มที่มีชื่อว่า Parse หรือกดปุ่ม Enter ที่เป็นพิมพ์ ตัวโปรแกรมจะเริ่มทำการวิเคราะห์ประโยค สร้างโครงสร้างคุณสมบัติ และรหัสชั่วคราว ดังรูปที่ 6.5-6.7 ตามลำดับ โดยที่เมื่อต้องการดูภาษาเอสคิวแอลที่ได้จากการแปลงประโยคให้เลือกปุ่มที่ระบุเป็น SQL และถ้าหากต้องการเลือกดูโครงสร้างคุณสมบัติหรือรหัสชั่วคราวก็ให้เลือกปุ่ม Feature Structure หรือ Intermediate Code ตามลำดับ



รูปที่ 6.5 ผลลัพธ์ของการแปลงเป็นประโยค “มีหนังสือกี่เล่ม” เป็นเอสคิวแอล

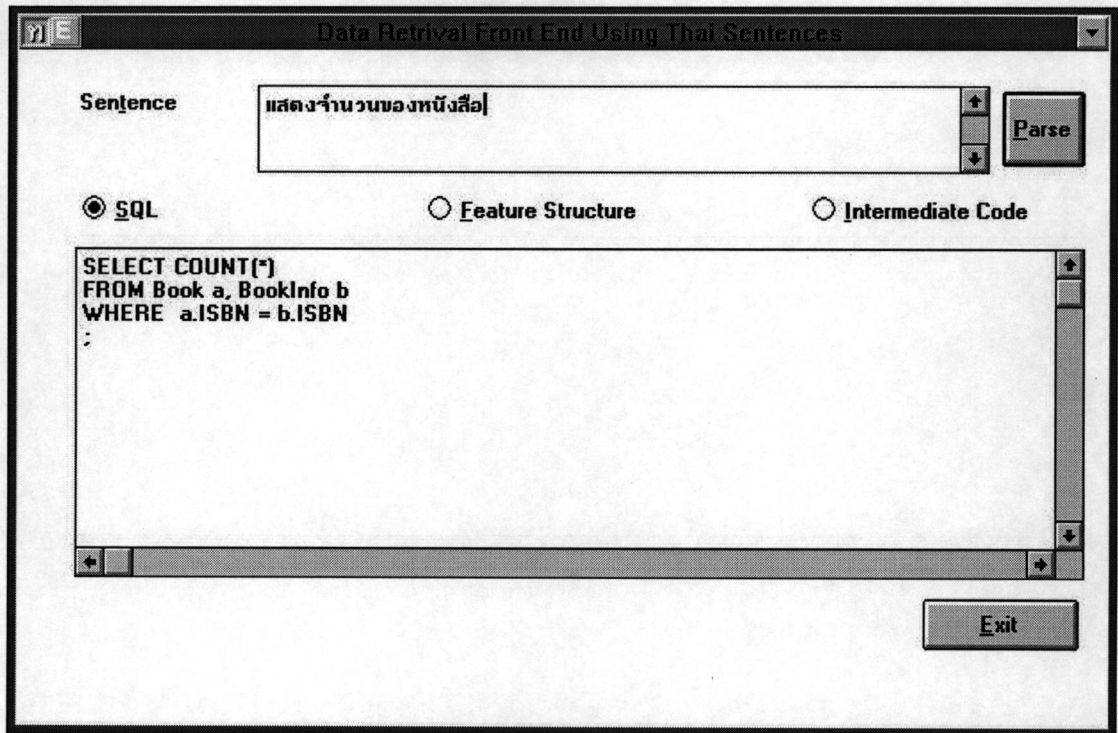


รูปที่ 6.6 แสดงโครงสร้างคุณสมบัติของประโยค “มีหนังสือที่เล่ม”



รูปที่ 6.7 แสดงโครงสร้างคุณสมบัติที่เป็นรหัสชั่วคราวของประโยค “มีหนังสือที่เล่ม”

ในการแปลงประโยค ตัวโปรแกรมสามารถแปลงประโยคได้มากกว่า 1 รูปแบบที่มีความหมายของประโยคที่เป็นการอ้างถึงคำสั่งเอสคิวแอลที่เหมือนกันตัวอย่างเช่น มีหนังสือที่เล่ม จะมีความหมายเช่นเดียวกับประโยค แสดงจำนวนของหนังสือ เป็นต้น ซึ่งโปรแกรมสามารถสร้างคำสั่งเอสคิวแอลได้แบบเดียวกัน ดังรูปที่ 6.5 และ 6.8



รูปที่ 6.8 ผลลัพธ์ของการแปลงประโยค “แสดงจำนวนของหนังสือ” เป็นเอสคิวแอล