# บทที่ 4

## ผลการพัฒนาและการทดสอบโปรแกรม

### 4.1 ลักษณะระบบโปรแกรมและวิธีการใช้

ระบบโปรแกรมหลักที่พัฒนาขึ้นในงานวิจัยนี้ให้ชื่อว่า vecser ขนาด 109,857 ไบต์ มีลักษณะเป็นการรวมสิ่งแวดล้อมต่าง ๆ ของระบบเข้าไว้ด้วยกัน โดยนำเสนอเป็นวินโดว์ของแต่ละระบบย่อย ผู้ใช้สามารถดำเนินการในแต่ละระบบย่อยได้อย่างสะดวก โดยมีรายการทางเลือกกำกับในตอนล่างของแต่ละวินโดว์ เพื่อสามารถให้เลือกดำเนินการต่อไปได้

4.1.1 รายการเลือกหลัก จะเป็นกรอบแสดงโลโก้ของระบบ และรายการเลือกหลักของระบบ โดยมีทางเลือกหลักอยู่ 3 ทางเลือก คือ นำเอกสารเข้า สอบถามสารสนเทศ และออกจากระบบ ดังแสดงในรูปที่ 4.1

4.1.2 การนำเอกสารเข้าระบบ เป็นส่วนการนำเอกสารเข้าระบบ ซึ่งจะให้ผู้ใช้ใส่ชื่อแฟ้มเอกสารที่จะนำเข้าระบบ จากนั้นจะแสดงรายการเลือกตอนล่างของวินโดว์ 4 ทางเลือก คือ วิเคราะห์คำ ดำเนินการ กลับไปเมนูหลัก และออกจากระบบ ดังแสดงในรูปที่ 4.2

4.1.2.1 วิเคราะห์คำ เป็นส่วนแสดงให้เห็นคำที่ใช้เป็นดรรชนีของเอกสาร ดังแสดงในรูปที่ 4.3

4.1.2.2 ดำเนินการ เป็นส่วนดำเนินการกับเอกสารที่นำเข้าระบบ โดยจะมีการสร้างเวกเตอร์ของเอกสาร จากนั้นจะกลับไปวินโดว์รายการเลือกหลัก เป็นอันเสร็จสิ้นกระบวนการนำเอกสารเข้าระบบ

4.1.2.3 กลับไปเมนูหลัก ในกรณีเมื่อไม่ต้องการนำเอกสารนี้เข้าระบบ สามารถเลือกรายการนี้เพื่อกลับไปวินโดว์รายการเลือกหลักได้

4.1.2.4 ออกจากระบบ ในกรณีเมื่อไม่ต้องการนำเอกสารนี้เข้าระบบ และต้องการออกจากระบบโปรแกรม ก็สามารถใช้ทางเลือกนี้ได้เลย

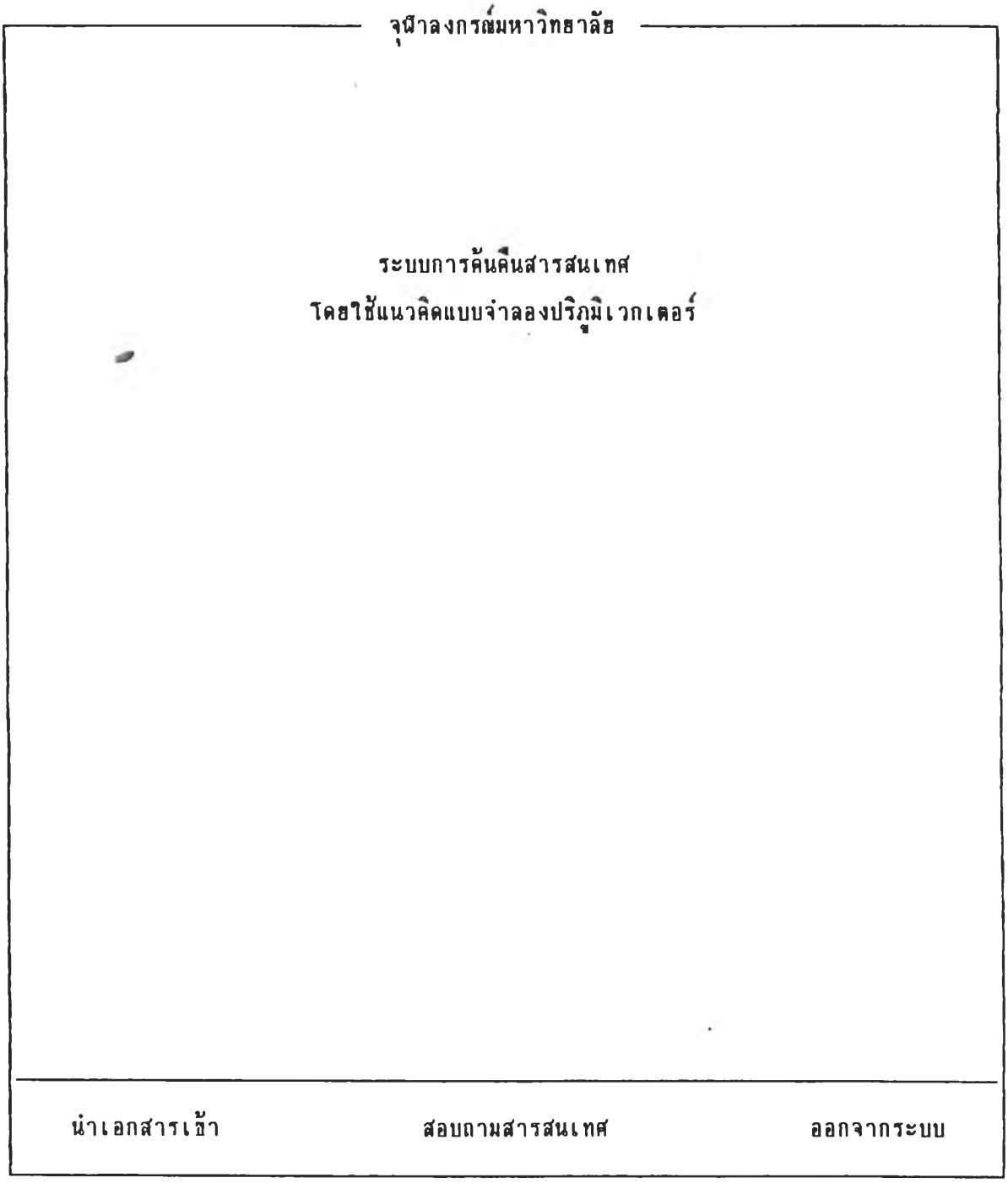
4.1.3 สอบถามสารสนเทศ เป็นส่วนการสอบถามสารสนเทศซึ่งเลือกจากรายการเลือกหลัก โดยจะแสดงกรอบให้ผู้ใช้ใส่ประโยคข้อคำถามได้ โปรแกรมจะใช้แถบสว่างแสดงที่จุดตัดคำให้โดยอัตโนมัติ แต่ผู้ใช้สามารถสลับการเป็นจุดตัดคำได้ โดยเลื่อนเคอร์เซอร์ไปที่จุดตัดคำแล้วใช้ปุ่ม F9 โดยปุ่ม F9 จะทำหน้าที่สลับสถานะการเป็นจุดตัดคำให้ คือหากเป็นแถบสว่างอยู่ก็จะสลับเป็นตัวปกติซึ่งไม่ใช้จุดตัดคำ หรือหากเป็นตัวปกติก็จะสลับเป็นแถบสว่างซึ่งแสดงการเป็นจุดตัดคำให้ เมื่อจบประโยคข้อคำถามให้ผู้ใช้กดปุ่ม Esc จากนั้นจะแสดงรายการเลือกตอนล่างของวินโดว์ 4 ทางเลือก คือ วิเคราะห์คำ ดำเนินการ กลับไปเมนูหลัก และออกจากระบบ ดังแสดงในรูปที่ 4.4

4.1.3.1 วิเคราะห์คำ เป็นส่วนแสดงให้เห็นคำที่ใช้เป็นดรรชนีของประโยคข้อคำถาม ดังแสดงในรูปที่ 4.5

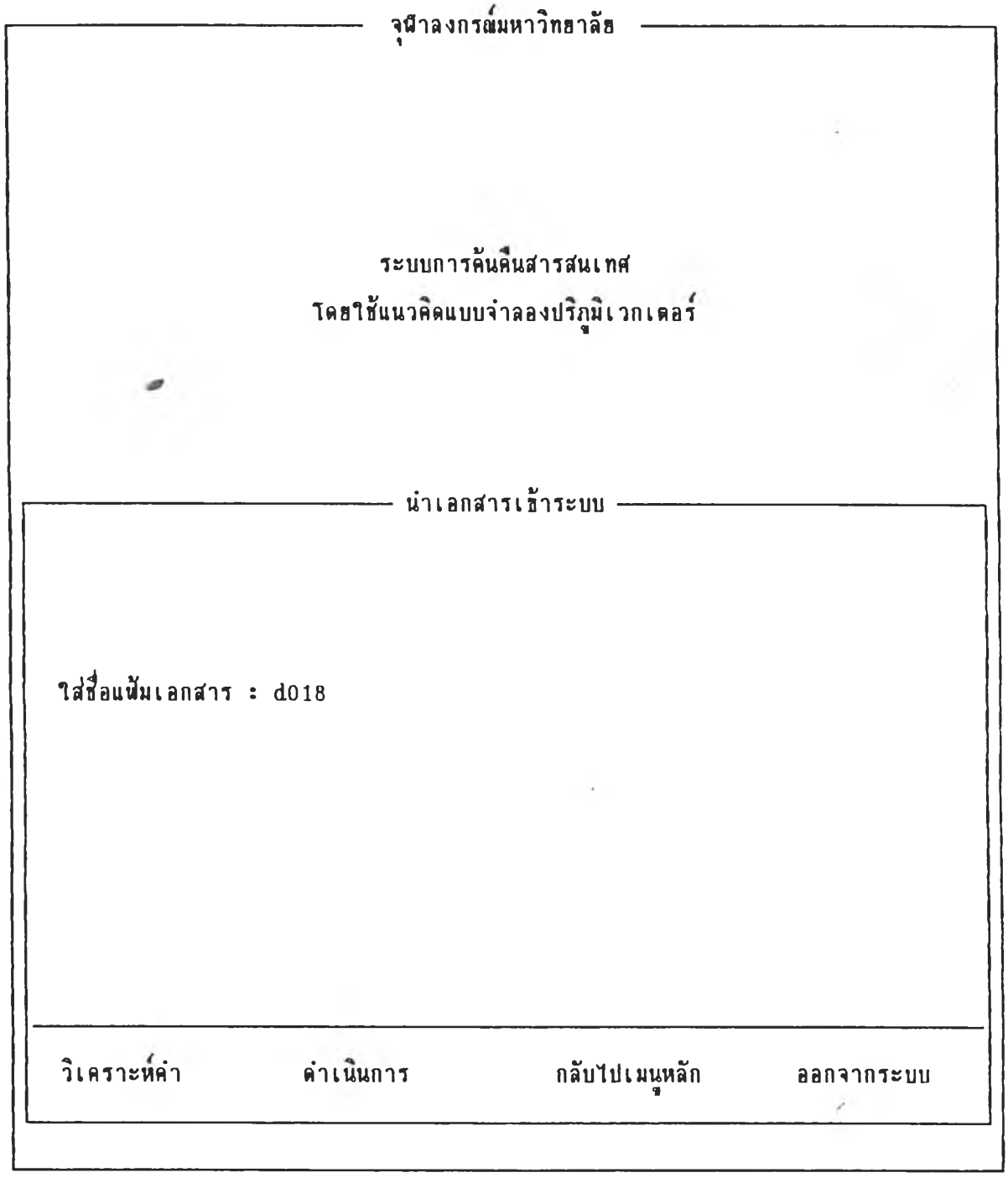
4.1.3.2 ดำเนินการ เป็นส่วนดำเนินการค้นหาเอกสาร จากประโยคข้อคำถาม โดยจะนำเวกเตอร์ของประโยคข้อคำถามไปคำนวณหาค่าความใกล้เคียงกับเวกเตอร์ของเอกสาร หากมีเอกสารที่ค่าความใกล้เคียงอยู่ในขอบเขตที่กำหนด ก็จะแสดงผลลัพธ์ออกมาตามลำดับค่าความใกล้เคียงจากมากไปหาน้อย ดังแสดงในรูปที่ 4.6 ผู้ใช้สามารถใช้ปุ่ม PgUp หรือ PgDn เลือกดูเอกสารที่ตรงกับกรณีได้ หรือใช้ปุ่ม Esc เมื่อต้องการเลิกดูเอกสาร และจะกลับไปเมนูหลัก

4.1.3.3 กลับไปเมนูหลัก ในกรณีเมื่อไม่ต้องการใช้ประโยคข้อคำถามนี้ สามารถเลือกรายการนี้เพื่อกลับไปวินโดว์รายการเลือกหลักได้

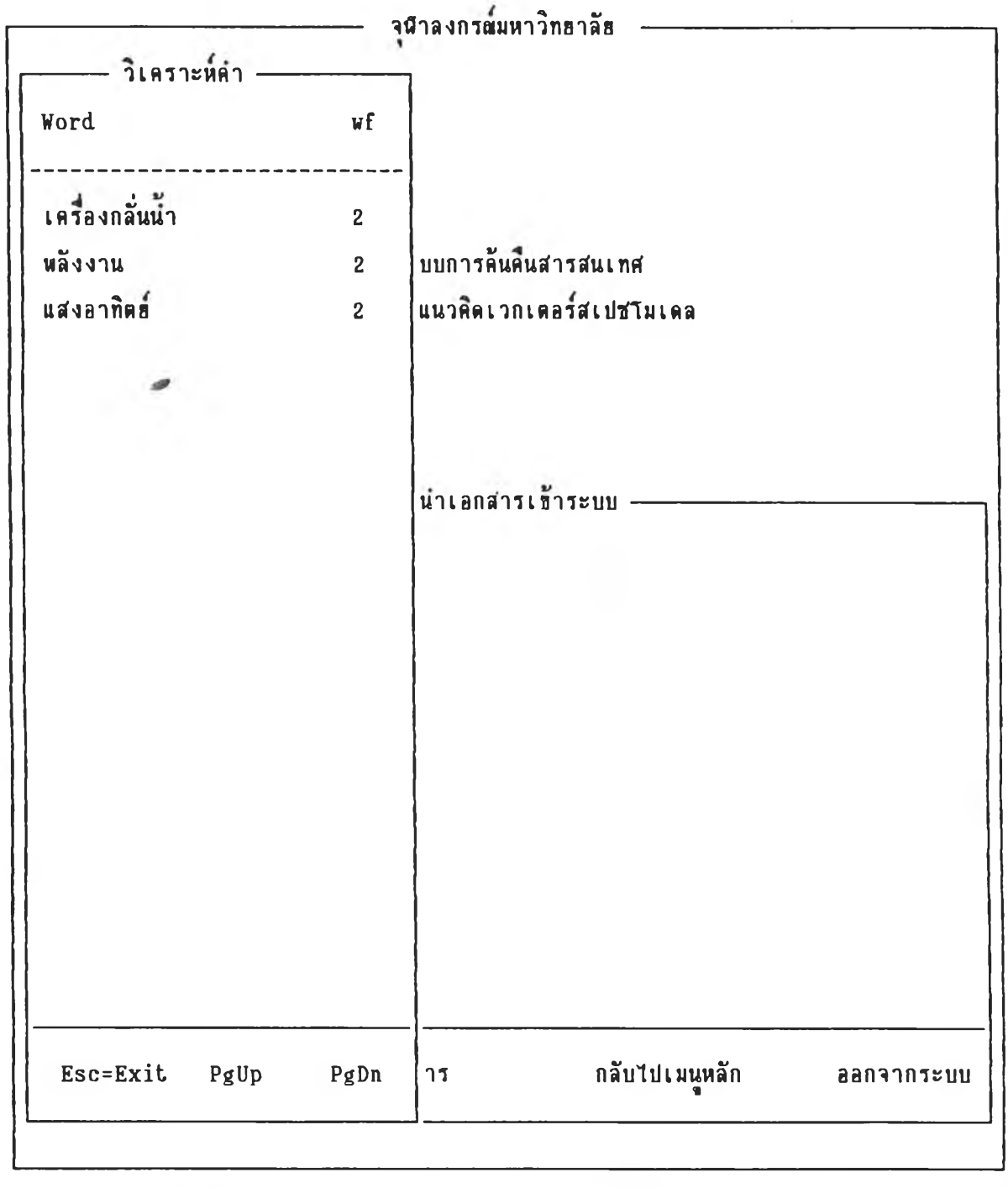
จุฬาลงกรณ์มหาวิทยาลัย



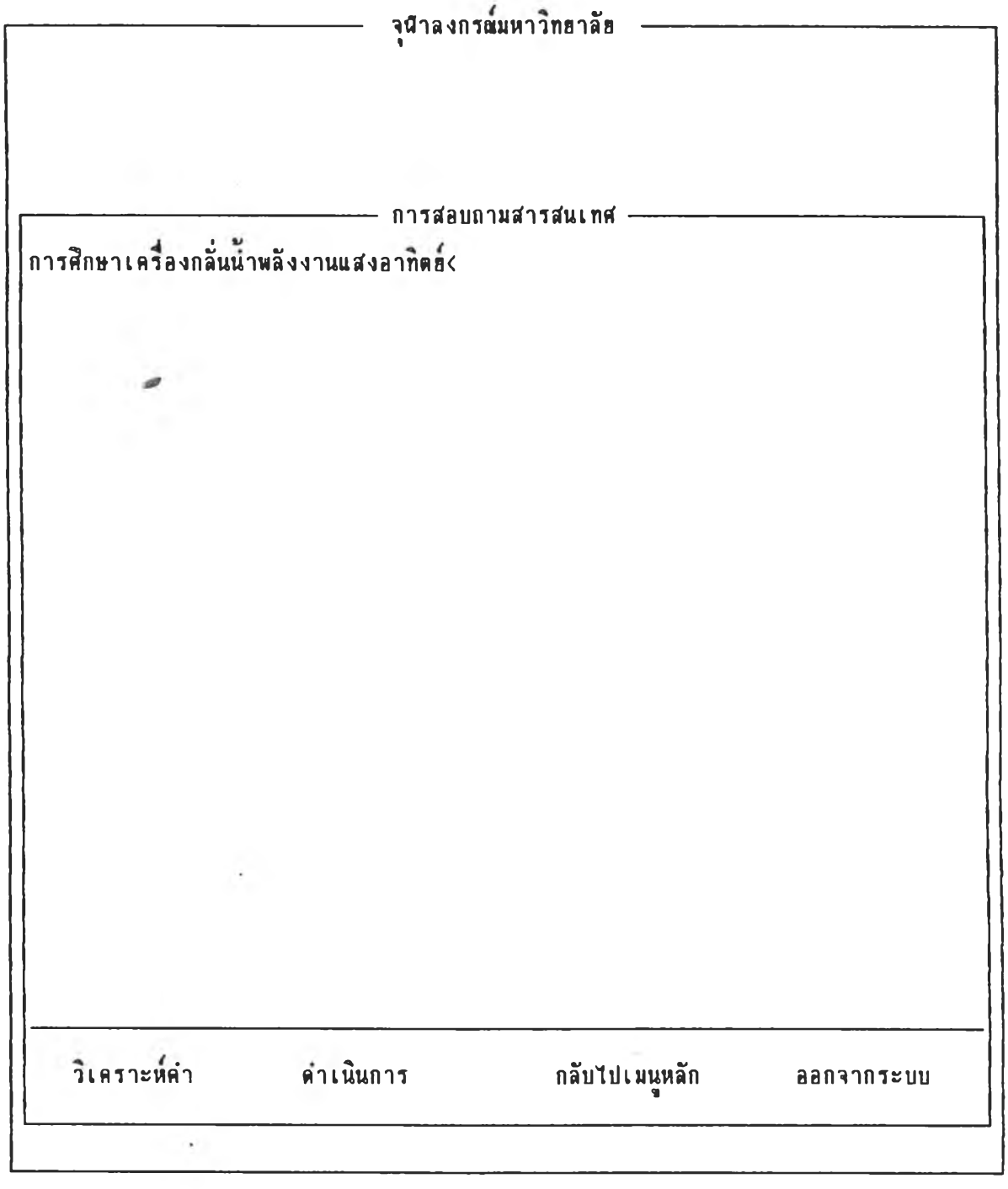

รูปที่ 4.1 ลักษณะจอภาพรายการเลือกหลักและกรอบแสดงโลโก้ของระบบ



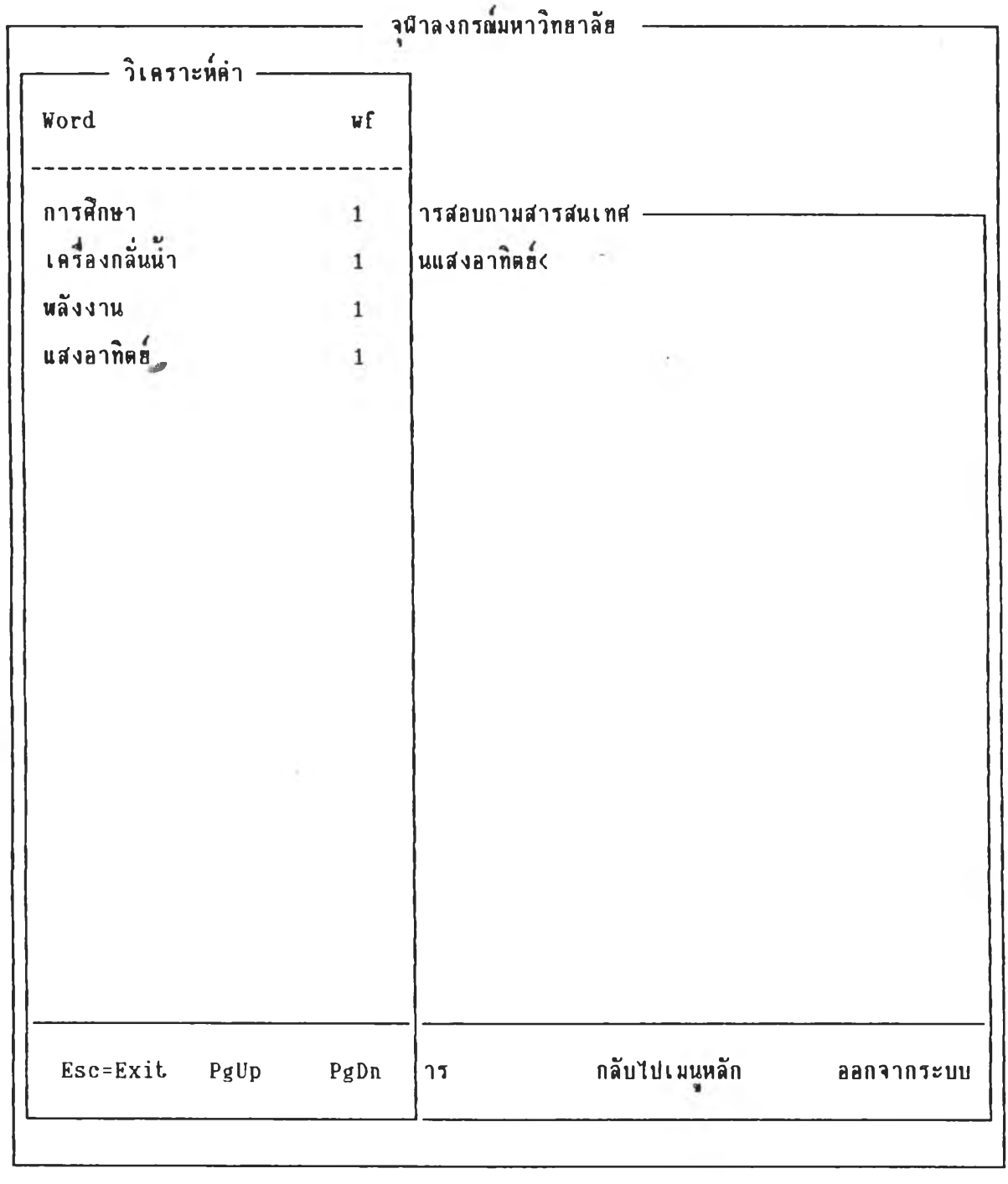

รูปที่ 4.2 ลักษณะจอภาพการนำเอกสารเข้าระบบ



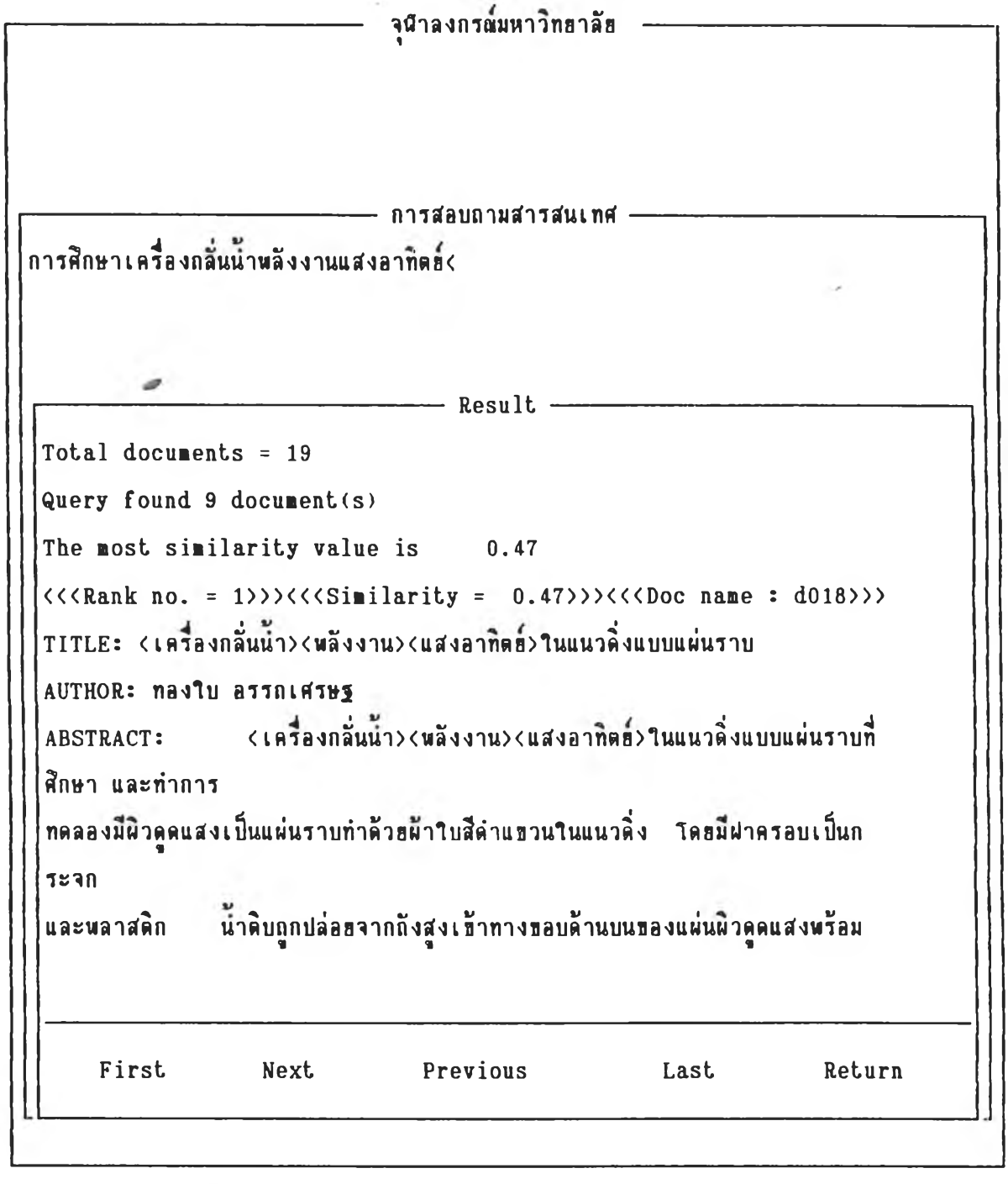

รูปที่ 4.3 ลักษณะจอภาพแสดงการวิเคราะห์คำของแฟ้มเอกสาร




รูปที่ 4.4 ลักษณะจอภาพการสอบถามสารสนเทศ




รูปที่ 4.5 ลักษณะจอภาพแสดงการวิเคราะห์ค่าของประโยคสอบถาม




รูปที่ 4.6 ลักษณะจอภาพแสดงผลลัพธ์จากการสอบถามสารสนเทศ

4.1.3.4 ออกจากระบบ ในกรณีเมื่อไม่ต้องการใช้ประโยคข้อคำถามนี้ และต้องการออกจากระบบโปรแกรม ก็สามารถใช้ทางเลือกนี้ได้เลย

4.1.4 ออกจากระบบ เป็นรายการเลือกที่อยู่บนทางเลือกหลัก เพื่อใช้เลือกออกจากระบบโปรแกรมนี้

นอกจากโปรแกรมหลัก vecser นี้แล้วยังได้พัฒนาโปรแกรม ให้ช่วยดำเนินการต่าง ๆ แยกออกมาจากโปรแกรมหลักอีก 2 โปรแกรมคือ

โปรแกรม adddoc ช่วยในการนำเอกสารเข้าระบบทีละหลาย ๆ แฟ้ม ในลักษณะของบรรทัดคำสั่ง (command line) โปรแกรมนี้มีขนาด 26,394 ไบต์

โปรแกรม wordtree ช่วยจัดการเกี่ยวกับแฟ้มข้อมูล wordtree ซึ่งเป็นแฟ้มข้อมูลที่ใช้เก็บค่าดรรชนีของเอกสารทั้งหมด มีลักษณะเป็นบีทรีที่จัดเก็บข้อมูลลงบนหน่วยความจำภายนอก โดยจะช่วยทำหน้าที่ลบ หรือเรียกค้นหาข้อมูลค่าดรรชนี โปรแกรมนี้มีขนาด 31,345 ไบต์

## 4.2 แฟ้มข้อมูลที่เกิดขึ้นในระบบ

ระบบที่พัฒนาขึ้นในงานวิจัยนี้ จะต้องตระเตรียมสร้างสารบบ (directory) ขึ้น 2 สารบบ ภายใต้สารบบที่ใช้เก็บโปรแกรม vecser adddoc และ wordtree คือ สารบบ DOC และ VEC โดยสารบบทั้งสองนี้จะใช้เก็บแฟ้มข้อมูลสำคัญของระบบโปรแกรม ซึ่งจะได้กล่าวในลำดับต่อไปนี้

4.2.1 แฟ้มข้อมูลเก็บเอกสาร ดังได้กล่าวแล้วว่าแฟ้มเอกสารที่ใช้ในงานวิจัยนี้เป็นแฟ้มที่สร้างขึ้นโดยใช้โปรแกรมประมวลคำ ดังนั้นการตั้งชื่อแฟ้มเอกสารจึงขึ้นกับความสามารถที่โปรแกรมประมวลคำนั้นให้ใช้ และจุดมุ่งหมายของการวิจัยนี้ ก็จะเป็นลักษณะให้ผู้ใช้กำหนดการตั้งชื่อแฟ้มเอกสารขึ้นเอง โดยอาจนำสิ่งที่เป็นตัวแทนของหนังสือมาตั้งเป็นชื่อแฟ้มข้อมูล เช่น เลขเรียกหนังสือ เลขที่หนังสือ เป็นต้น แฟ้มเอกสารที่จะนำเข้าระบบจะต้องจัดเก็บไว้ภายใต้สารบบ DOC

4.2.2 แฟ้มข้อมูลเวกเตอร์ ชื่อแฟ้มข้อมูลคือ vectors เก็บอยู่ภายใต้สารบบ vec

4.2.3 แฟ้มข้อมูลค่าดรรชนี ชื่อแฟ้มข้อมูลคือ wordtree เก็บอยู่ภายใต้สารบบ vec

## 4.3 ระบบคอมพิวเตอร์ที่นำมาใช้ในการพัฒนาระบบ

4.3.1 ฮาร์ดแวร์

- เครื่องไมโครคอมพิวเตอร์
  ซีพียู 80386SX 16 Mhz หน่วยความจำหลัก 1 เมกกะไบต์
- TRIDENT VGA Graphics Adapter with Color Monitor
  SuperSync 2A+ ของ TVM
- ฮาร์ดดิสก์ 43 เมกกะไบต์

4.3.2 ซอฟต์แวร์

- เอ็มเอส-ดอส เวอร์ชัน 3.30
- คอมไพเลอร์ Turbo C++ เวอร์ชัน 1.00
- ไดร์เวอร์ภาษาไทย Vthai เวอร์ชัน 2.01 โดยสถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย

## 4.4 การทดสอบโปรแกรม

แฟ้มเอกสารในการทดสอบ ใช้โปรแกรมประมวลคำ CU WRITER เวอร์ชัน 1.52 สำหรับการทดสอบโปรแกรมได้แบ่งออกเป็น 2 ส่วน คือ

4.4.1 การทดสอบความถูกต้อง ในส่วนนี้เป็นการทดสอบความถูกต้อง ของการคำนวณค่าความใกล้เคียงระหว่างเวกเตอร์ของเอกสาร และของประโยคสอบถาม นอกจากนั้นยังทดสอบการจัดลำดับของเอกสารที่ดึงออกมา กรณีที่ประโยคสอบถามตรงกับเอกสารหลายฉบับ

โดยเปรียบเทียบผลที่ได้จากโปรแกรมกับที่คำนวณด้วยมือ ในการทดสอบนี้ เพื่อให้สามารถควบคุมผลการคำนวณที่จะทำด้วยมือ จึงใช้แฟ้มเอกสารเพียง 10 ฉบับ ที่เลือกสรรแล้ว ทำการแปลงเอกสารเป็นเวกเตอร์ด้วยมือ ใช้ประโยคสอบถามที่กำหนดไว้ล่วงหน้า โดยพิจารณาและคำนวณด้วยมือแล้วว่าจะสามารถดึงเอกสารใดออกมาได้บ้าง จากการทดสอบดังกล่าวปรากฏว่าผลการทดสอบโปรแกรมเป็นไปตามผลที่คำนวณด้วยมือทุกประการ

4.4.2 การทดสอบระบบโดยรวม ข้อมูลที่ใช้ในการทดสอบ จากรายงานผลการวิจัยใน 4 สาขาวิชาการ คือ สาขาเคมีและเภสัชวิทยา สาขาวิทยาศาสตร์กายภาพ สาขาวิศวกรรมศาสตร์ และสาขาเกษตรศาสตร์และชีววิทยา โดยสร้างแฟ้มเอกสารเพื่อเป็นตัวแทนของรายงานผลการวิจัยดังกล่าว รายการข้อมูลที่เก็บมี 3 รายการ คือ ชื่อเรื่อง ชื่อผู้วิจัย และบทคัดย่อผลการวิจัย รายการข้อมูลที่ใช้ทำดรรชนีของเอกสารคือ ชื่อเรื่อง และบทคัดย่อผลการวิจัย โดยใช้สัญญลักษณ์ <...> คร่อมส่วนที่เลือกเป็นคำดรรชนี ตัวอย่างแฟ้มเอกสารแสดงในภาคผนวก ค. และเวคเตอร์ของเอกสารทั้งหมดที่คำนวณได้ดังแสดงในภาคผนวก ง. รายละเอียดข้อมูลต่าง ๆ ประกอบการทดสอบสรุปได้ดังนี้

| | |
|---|---|
| แฟ้มเอกสารที่ใช้ในการทดสอบรวม (ชื่อแฟ้ม d001-d165) | 165 แฟ้ม |
| ขนาดของแฟ้มเอกสารที่เล็กที่สุด | 565 ไบต์ |
| ขนาดของแฟ้มเอกสารที่ใหญ่ที่สุด | 1,935 ไบต์ |
| ขนาดโดยเฉลี่ยของแฟ้มเอกสาร | 965 ไบต์ |
| ขนาดรวมทั้งหมดของแฟ้มเอกสาร | 159,137 ไบต์ |
| ขนาดของแฟ้ม vectors ที่ได้ | 8,432 ไบต์ |
| ขนาดของแฟ้ม wordtree ที่ได้ | 35,198 ไบต์ |
| จำนวนคำดรรชนีทั้งหมดที่ใช้ | 399 คำ |

ได้ทำการทดสอบโดยใช้ประโยคสอบถามจากผู้ทำการทดสอบ 2 กลุ่ม คือ กลุ่มผู้ทราบ และไม่ทราบคำดรรชนีที่ใช้ในฐานข้อมูลนี้ จากผลการทดสอบปรากฏว่ากลุ่มผู้ไม่ทราบคำดรรชนีมีประโยค

สอบถามที่ไม่สามารถดึงเอกสารได้ 85 เปอร์เซ็นต์ สำหรับเอกสารที่ดึงออกมาได้ทุกเอกสารจะมีคำที่เป็นดรรชนีประกอบในประโยคสอบถาม และเมื่อตรวจสอบด้วยมือ คำดรรชนีที่ใช้ในประโยคสอบถามจะสามารถดึงได้ทุกเอกสารที่มีคำดรรชนีนี้บรรจุอยู่ เมื่อตรวจสอบความตรงกับกรณีระหว่างประโยคสอบถามและเอกสารที่ดึงออกมาได้ พบว่ากรณีประโยคสอบถามที่ใช้คำจำเพาะ เช่น "การวิจัยหนอนตายอยาก" "เภสัชวิทยาของไพล" "โรคเส้นด่าของยางพารา" จะสามารถดึงเอกสารที่ตรงกับกรณีได้ดี แต่ทั้งนี้คำดรรชนีของเอกสารต้องเป็นคำที่จำเพาะนี้ด้วย สำหรับลำดับของเอกสารที่ดึงออกมาได้ พบว่ากรณีประโยคสอบถามที่ใช้คำจำเพาะจะใช้ได้ดีเช่นกัน