



บทที่ 1

บทนำ

ความเป็นมาของปัญหา

ปัจจุบันนี้ ในประเทศไทยมีผู้นิยมนำเครื่องไมโครคอมพิวเตอร์ไปใช้งานต่างๆ เป็นจำนวนมากเห็นได้จากยอดขายในปีหนึ่งๆ ขายได้เป็นจำนวนนับหมื่นเครื่อง การที่มีผู้นิยมมาใช้เนื่องจากมีราคาถูก โปรแกรมใช้งานต่างๆ อยู่ในรูปของโปรแกรมสำเร็จรูป สามารถเลือกซื้อได้ง่าย ตรงตามวัตถุประสงค์ของงานและราคาไม่แพง ไมโครคอมพิวเตอร์จึงนับเป็นอุปกรณ์ ที่มีบทบาทอย่างยิ่งในอันที่จะเสริมสร้างคนในประเทศ ให้มีความรู้ความสามารถและประสบการณ์เกี่ยวกับการประมวลผลข้อมูลด้วยเครื่องจักร อย่างไรก็ตามความแพร่หลายในการใช้ยังคงจำกัดอยู่แต่ในเมืองหลวงและหัวเมืองใหญ่ ๆ เท่านั้น การกระจายการใช้เครื่องไมโครคอมพิวเตอร์ยังไม่เต็มที่เกิดจากปัญหาหลาย ๆ ประการ เช่น ปัญหาการเรียนรู้การใช้งานปัญหาเกี่ยวกับเอกสารคู่มือต่าง ๆ ไม่ครบและเป็นภาษาอังกฤษ ปัญหาอีกประการหนึ่งที่สำคัญไม่น้อยกว่าปัญหาอื่น ๆ คือปัญหาการขาดแคลนโปรแกรมที่ทำงานกับข้อมูลภาษาไทย รวมถึงการสั่งงานโปรแกรม หรือการเขียนโปรแกรมด้วยภาษาไทย

รูปแบบคำสั่ง ที่ใช้ในโปรแกรมสำเร็จรูปหรือตัวแปลภาษาต่าง ๆ เป็นรูปแบบที่จำกัด ผู้ใช้จะต้องจำกฎข้อบังคับของคำสั่งแต่ละคำสั่งได้อย่างเคร่งครัด ตัวแปลภาษาหรือโปรแกรมจะไม่ยอมรับคำสั่งที่ผิดกฎแม้เพียงเล็กน้อย รูปแบบคำสั่งบางคำสั่งเป็นรูปแบบในเชิงคณิตศาสตร์ ที่ยากแก่การเข้าใจของคนทั่ว ๆ ไป ในงานบางประเภทที่ไม่สลับซับซ้อน เป็นงานที่เกี่ยวข้องกับ งานส่วนบุคคลหรืองานธุรกิจทั่ว ๆ ไป เช่นงานจัดการข้อมูล ประวัติบุคคลากร ประวัติลูกค้า รายละเอียด

เกี่ยวกับสินค้า ถ้าหากผู้ใช้โปรแกรมสามารถออกคำสั่งโดยใช้ภาษาไทย ที่เราใช้พูดหรือเขียนในชีวิตประจำวัน ย่อมเป็นสิ่งที่ส่งผลให้ผู้ใช้งาน เข้าใจวิธีการสั่งและใช้งาน โปรแกรมได้ในเวลาอันรวดเร็ว ก่อให้เกิดการกระจายการประยุกต์ใช้งานคอมพิวเตอร์ไปสู่สังคม ได้กว้างขวางและรวดเร็วยิ่งขึ้น ภาษาในแบบที่มีรูปแบบคำสั่งจำกัดส่วนใหญ่นิยมใช้สำหรับงานด้านการเขียนโปรแกรม เป็นภาษาที่มีไวยากรณ์อิสระจากเนื้อหา (Context free Grammar) ส่วนภาษาธรรมชาติที่ใช้นักงานในชีวิตประจำวัน มีไวยากรณ์ขึ้นกับเนื้อหา (Context Sensitive Grammar)

ภาษาสอบถาม (Query Language)

เป็นภาษาที่ใช้สำหรับการสอบถามปัญหาและ มีรูปแบบที่คล้ายกับภาษาในแบบ อินเตอร์พรีตเตอร์ (Interpreter) คือเมื่อเวลาที่ผู้ใช้ พิมพ์คำสั่งเข้าสู่เครื่อง ตัวแปลภาษาจะทำงานให้ทันที ภาษาสอบถามส่วนใหญ่มักจะเป็นแบบที่มีไวยากรณ์อิสระจากเนื้อหา

ตัวอย่างภาษาในแบบไวยากรณ์อิสระจากเนื้อหา เช่น ส่วนหนึ่งของคำสั่ง LIST ของโปรแกรมดีเบสทู (dBASE II) และเป็นคำสั่งในแบบภาษาสอบถามที่มีไวยากรณ์ดังนี้

<คำสั่งLIST> ::= LIST ; LIST FOR <lexp> (1)

<lexp> ::= <lterm> ; <lexp><lop><lterm> ;
'(' <lexp> ')'; <ulop> '(' <lexp> ')'

<lterm> ::= <rexp> ; <ulop><rexp> (3)

<rexp> ::= <aexp><rop><aexp> (4)

<aexp> ::= <aterm> ; <aexp><aop><aterm> ;
'(' <aexp> ')'; <uaop> '(' <aexp> ')'

<aterm> ::= <arg> ; <uaop><arg> (6)

<arg> ::= <var> ; <con> ; <field> (7)

$\langle \text{uop} \rangle ::= \text{.NOT.} \quad (8)$

$\langle \text{lop} \rangle ::= \text{.AND.} ; \text{.OR.} \quad (9)$

$\langle \text{rop} \rangle ::= = ; < ; > ; <= ; >= \quad (10)$

$\langle \text{uaop} \rangle ::= + ; - \quad (11)$

$\langle \text{aop} \rangle ::= + ; - ; * ; / \quad (12)$

สมมติให้

$\langle \text{con} \rangle ::= 3000 ; 20 \quad (13)$

$\langle \text{field} \rangle ::= \text{SALARY} ; \text{AGE} \quad (14)$

เราสามารถสร้างคำสั่ง LIST ตามกฎในไวยากรณ์ดังนี้

$\langle \text{aterm} \rangle ::= \text{SALARY} \quad \text{แทน(14)ใน(6)}$

$\langle \text{aexp} \rangle ::= \text{SALARY} \quad \text{แทน(6)ใน(5)}$

$\langle \text{aterm} \rangle ::= 3000$

$\langle \text{aexp} \rangle ::= 3000$

$\langle \text{rexp} \rangle ::= \text{SALARY} > 3000 \quad \text{แทน (10),(5)ใน(4)}$

ในท่านองเดียวกันจะได้

$\langle \text{rexp} \rangle ::= \text{AGE} = 20$

$\langle \text{lexp} \rangle ::= \text{SALARY} > 3000 \text{ .AND. } \text{AGE} = 20$

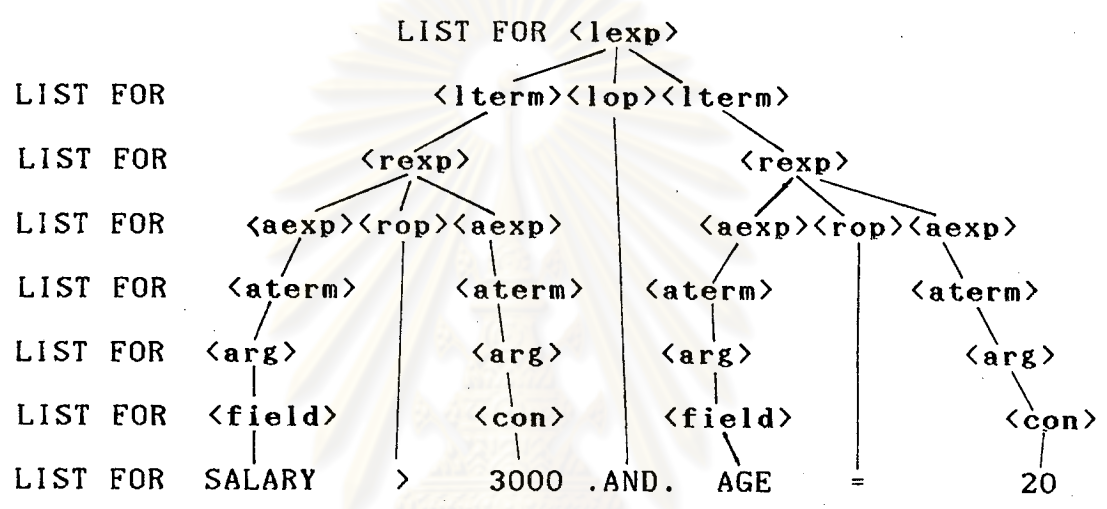
แทน(4),(3)ใน(2)

$\langle \text{คำสั่งLIST} \rangle ::= \text{LIST FOR SALARY} > 3000 \text{ .AND. } \text{AGE} = 20$

แทน(2)ใน(1)

จุฬาลงกรณ์มหาวิทยาลัย

แสดงแผนผังเพื่อแสดงขั้นตอนการสร้างคำสั่งดังนี้



จะเห็นได้ว่า ภาษาที่มีไวยากรณ์อิสระจากเนื้อหา ในการแทนหน่วย
 ใดหน่วยหนึ่ง ในกฎด้วยนิยามของหน่วยนั้น เราสามารถแทนได้โดยเสรี โดย
 ที่หน่วยอื่นๆในกฎเดียวกันไม่ส่งอิทธิพลมากระทบ เช่นจากกฎข้อที่(4) ซึ่งเป็นกฎ
 ของ <rexp> (Relational Expression) เราสามารถแทนค่าของ <aexp>
 (Arithmetic Expression) โดยไม่ต้องคำนึงถึง <rop> (Relational
 Operator) เป็นต้น

ตัวอย่างภาษาในแบบไวยากรณ์ขึ้นกับเนื้อหา ได้แก่ภาษาไทยที่เราใช้
 พูด หรือ เขียนกันในชีวิตประจำวัน ไวยากรณ์แบบง่ายๆของประโยคในภาษาไทย
 เขียนได้ดังนี้



- <ประโยค> ::= <หน่วยประธาน><หน่วยกรรม> (1)
- <หน่วยประธาน> ::= จัน : สุนัข (2)
- จัน<หน่วยกรรม> ::= จัน<หน่วยกรรม 1> (3)
- สุนัข<หน่วยกรรม> ::= สุนัข<หน่วยกรรม 2> (4)
- <หน่วยกรรม 1> ::= พุด : ร้องไห้ (5)
- <หน่วยกรรม 2> ::= เहां : หอน (6)

เมื่อเราเลือก 'จัน' เป็นหน่วยประธานจะได้ประโยคดังนี้

- <ประโยค> ::= จัน<หน่วยกรรม> แทน(2)ใน(1) (7)
- ::= จัน<หน่วยกรรม 1> แทน(3)ใน(7) (8)
- ::= จันพุด : จันร้องไห้ แทน(5)ใน(8) (9)

เห็นได้ว่าในกฎข้อ(1)ไม่สามารถแทนค่าของแต่ละหน่วยเข้าไปในกฎได้อย่างอิสระจากกัน <หน่วยประธาน> แต่ละหน่วยถูกบังคับให้ใช้คู่กับ <หน่วยกรรม> ที่เหมาะสม เช่น 'จัน' สามารถคู่กับ 'พุด' หรือ 'ร้องไห้' ได้เท่านั้น ดังนั้นกฎ(3) และ (4) จึงถูกกำหนด ขึ้นบังคับการใช้<หน่วยกรรม> ใดที่เหมาะสมกับ<หน่วยประธาน> ภาษาในแบบไวยากรณ์อิสระจากเนื้อหาจะไม่มีกฎในลักษณะข้อ(3)และ(4)นี้ การที่เป็นเช่นนี้เพราะว่า ต้องคำนึงถึงเนื้อหาหรือความหมายของประโยคที่ถูกสร้างขึ้นว่าถูกต้องตรงกับความเป็นจริงหรือไม่

ข้อดีและข้อเสียสำหรับภาษาทั้งสองแบบในการนำมาใช้เป็นภาษา

คอมพิวเตอร์

ภาษาในแบบไวยากรณ์อิสระจากเนื้อหา

ข้อดี

1. ประโยคคำสั่งกระชับรัดกุม ไม่มีคำขยายที่เกินความจำเป็น
2. มีกฎ (SYNTAX) ในการสร้างประโยคจำกัดและไม่ยุ่งยากซับซ้อน
3. การตีความคำสั่งทำได้ง่ายและรวดเร็ว เพราะเป็นภาษาที่ใช้กฎการสร้างประโยคในการกำหนดความหมายของประโยคเพียงอย่างเดียว ไม่จำเป็นต้องมีการกำหนดความสัมพันธ์ทางความหมาย
4. คำสั่งแต่ละคำสั่งตีความได้เพียงอย่างเดียว ไม่มีรูปแบบคำสั่งที่มีความหมายกำกวม

ข้อเสีย

1. รูปแบบคำสั่งบางรูปแบบเป็นรูปแบบทางคณิตศาสตร์ ซึ่งยากแก่ความเข้าใจของคนทั่วไป เช่น กฎเกี่ยวกับการสร้างนิพจน์ (Expression)
2. ขาดความยืดหยุ่นของคำสั่ง เนื่องจากประโยคคำสั่งแต่ละคำสั่ง มีรูปแบบเพียงอย่างเดียว และมีกฎที่หลีกเลี่ยงไม่ได้ ดังนั้นถ้าใช้คำสั่งในรูปแบบอื่นถือว่าผิด แม้มองในแง่ความหมายจะถูกต้องก็ตาม
3. ไม่ใช้ภาษาธรรมชาติที่ใช้ในชีวิตประจำวัน ทำให้ผู้ใช้งานเสียเวลาในการเรียนรู้

ภาษาในแบบไวยากรณ์ขึ้นกับเนื้อหา

ข้อดี

1. ภาษาที่มีความยืดหยุ่นสูง ประโยคคำสั่งถูกสร้างขึ้นโดยไวยากรณ์ (SYNTAX) จำนวนหลายกฎที่สุดเท่าที่จะทำได้ เช่น สมมติว่าเราต้องการให้โปรแกรมแสดงรายชื่อลูกค้าที่มียอดหนี้มากกว่า 50,000 บาท เราสามารถใช้ประโยคคำสั่งดังตัวอย่างต่อไปนี้

ก. แสดงรายชื่อลูกค้าที่มียอดหนี้มากกว่า 50,000 บาท

ข. ต้องการรายชื่อผู้ที่คิดหนี้มากกว่า 500,000 บาท

ข้อเสีย

1. ประโยคคำสั่งขาดความกระชับรัด บางครั้งประกอบด้วยคำขยายที่ไม่จำเป็นต่อการตีความ

2. กฎในการสร้างประโยคคำสั่งยุ่งยากซับซ้อน

3. การตีความคำสั่งโดยเครื่องทำได้ใจความที่ครบ

สมบูรณ์ทำได้ยาก

4. ไม่เหมาะสำหรับใช้ เป็นภาษาในการเขียนโปรแกรม

วัตถุประสงค์ของการวิจัย

1. สร้างภาษาคอมพิวเตอร์ที่ช่วยเหลือผู้ใช้งาน

2. สร้างภาษาคอมพิวเตอร์ในรูปแบบใหม่ ๆ ที่ทำงานกับภาษาไทย

3. อ่านวยความสะดวกให้ผู้ใช้งาน สามารถสั่งงานโปรแกรมได้โดยสะดวกและง่าย โดยใช้ภาษาสอบถาม (Query Language) ที่มีข้อจำกัดน้อย และมีรูปแบบที่ใกล้เคียงกับภาษาไทยที่ใช้งานในชีวิตประจำวัน

4. เพื่อให้คนทั่วไปเรียนรู้การสั่งงานโปรแกรม ได้รวดเร็วขึ้น โดยไม่จำเป็นต้องมี พื้นฐานความรู้ทางคอมพิวเตอร์ หรือมีเพียงเล็กน้อย

ประโยชน์ที่คาดว่าจะได้รับการวิจัยนี้

1. ส่งเสริมให้เกิดการเรียนรู้ และการประยุกต์ใช้งานคอมพิวเตอร์แพร่หลายไปสู่สังคมได้กว้างขวางยิ่งขึ้น
2. เป็นพื้นฐานหรือเป็นแนวทาง สำหรับการค้นคว้าวิจัยเรื่องคำสั่งงานลักษณะเดียวกันให้มีประสิทธิภาพและความสามารถมากยิ่งขึ้น
3. ใช้เป็นเครื่องช่วยการสอน การสั่งงานและการประยุกต์ใช้งานคอมพิวเตอร์
4. นำไปประยุกต์หรือดัดแปลง เพื่อใช้งานกับข้อมูลภาษาไทยได้สะดวกขึ้นกว่าโปรแกรมที่ดัดแปลงแก้ไขมาจากโปรแกรมของต่างประเทศ

ขอบเขตการวิจัย

1. ประเภทของฐานข้อมูลที่ใช้

1.1 เป็นภาษาสอบถามภาษาไทย (Thai Query Language) เพื่อใช้สั่งงานโปรแกรมสำเร็จรูปสำหรับงานจัดการข้อมูล โดยอาศัยฐานข้อมูลในแบบสัมพันธ์ (Relational Data Base) การที่เลือกใช้ฐานข้อมูลในแบบนี้ก็เพราะใช้งานได้ง่าย นิยมใช้กันแพร่หลาย หาซื้อได้ทั่วไป ราคาไม่แพง (โปรแกรมในการจัดการฐานข้อมูล ไม่ใช้งานที่อยู่ในส่วนของวิทยานิพนธ์นี้)

1.2 ใช้โปรแกรมจัดการฐานข้อมูลแบบสัมพันธ์ที่ชื่อบีทรีฟ (Btrieve) ทำหน้าที่ในการจัดการฐานข้อมูล เพราะเป็นโปรแกรมที่ทำงานได้ดี และหาซื้อได้ง่ายในตลาดเมืองไทย

2. หน้าที่ของโปรแกรมที่สร้าง

ระบบโปรแกรมที่ถูกสร้างขึ้นเรียกว่าตัวประมวลผล

ภาษาสอบถาม

ตัวประมวลผลที่จะถูกสร้างขึ้นเป็นระบบที่ทำหน้าที่ เป็นส่วน ประสาน ระหว่างผู้ใช้งานกับระบบจัดการฐานข้อมูล คือ ทำหน้าที่ในการ รับคำสั่ง จากผู้ใช้ ตีความคำสั่ง และทำตามคำสั่ง โดยงานส่วนใดที่เกี่ยวข้องกับข้อมูลใน แฟ้มก็จะเป็นหน้าที่ของบีทรีฟ ส่วนงานที่เกี่ยวข้องกับการรับข้อมูลทางแบ่นพิมพ์ การ แสดงข้อมูลทางจอภาพ ทางเครื่องพิมพ์ การทำรายงานก็จะเป็นหน้าที่ของระบบที่ ถูกสร้างขึ้น

3. เครื่องคอมพิวเตอร์ที่ใช้

จะสร้างตัวประมวลผลภาษาสอบถามตามงานวิจัยนี้โดยอาศัย เครื่องในตระกูลไอบีเอ็มพีซีหรือ เครื่องที่มีคุณสมบัติที่ทำงานได้ในลักษณะเดียวกัน

4. ความสามารถและข้อจำกัดของภาษาสอบถามภาษาไทย

ภาษาสอบถามภาษาไทยที่ถูกสร้างขึ้นแบ่งคำสั่งออกเป็น 2 ชนิด คือ

- ก. คำสั่งชนิดไวยากรณ์อิสระจากเนื้อหา
- ข. คำสั่งชนิดกึ่งภาษาไทยธรรมชาติ

ความสามารถและข้อจำกัดของภาษาสอบถามชนิดกึ่งภาษา
ไทยธรรมชาติ

เนื่องจากในเรื่องเกี่ยวกับภาษาไทยธรรมชาติ หรือแม้แต่กึ่งภาษาไทยธรรมชาติ เป็นเรื่องที่กว้างขวางมาก ประกอบกับกฎเกณฑ์ ของเรื่องนี้ยังไม่มีผู้รวบรวมไว้ชัดเจน ดังนั้นจึงไม่สามารถทำให้ตัวประมวลผล ภาษาสอบถาม มีความสามารถในการวิเคราะห์ประโยคคำสั่งภาษาไทยธรรมชาติ ได้ผลสมบูรณ์ จึงต้องเรียกว่าภาษาสอบถามชนิดกึ่งภาษาไทยธรรมชาติ ซึ่งจะมี ข้อจำกัดของภาษาดังต่อไปนี้

4.1 เป็นภาษาสอบถามภาษาไทยที่มีข้อจำกัดเกี่ยวกับรูปแบบของคำสั่งน้อย และ ำที่มีรูปแบบของคำสั่งใกล้เคียงกับภาษาไทยที่ใช้งานชีวิต

ประจำวัน โดยจะเน้นเฉพาะคำสั่งประเภทใช้ค้นข้อมูลตามเงื่อนไข

4.2. ค่า (Data Value) ซึ่งทำหน้าที่ในการกำหนดคุณสมบัติ ของระเบียบที่ถูกค้นหาเช่นชื่อคน ที่อยู่ ประเภทกิจการ จะต้องเขียนอยู่ภายใต้เครื่องหมายผมหอง(‘) หรือพันธู (“) ตัวอย่างเช่น

‘อยากทราบอายุของคนชื่อ ‘สมศักดิ์’
‘สมศักดิ์’อายุเท่าไร

จากตัวอย่าง ข้อความที่อยู่ระหว่างเครื่องหมาย ‘ ทำหน้าที่ในการ กำหนดคุณสมบัติของระเบียบ หรือกลุ่มระเบียบที่ต้องการค้นหา เครื่องหมาย ‘ ทำให้รูปแบบของประโยคคำสั่งแตกต่างไปจากรูปแบบของประโยคในภาษาไทยเล็กน้อย

ความสามารถของภาษาสอบถามชนิดไวยากรณ์อิสระ

จากเนื้อหา

ภาษาที่มีไวยากรณ์ชนิดนี้ยอมเป็นภาษาที่มีรูปแบบของคำสั่งจำกัด งานวิจัยนี้ทำให้มีความยืดหยุ่นมากขึ้นโดยการ ทำให้คำสั่งมีรูปแบบมากกว่าหนึ่งรูปแบบ ใช้รูปแบบคำสั่งที่ง่ายต่อการเข้าใจ

5. คำสั่งของภาษาสอบถามภาษาไทย

สามารถใช้ภาษาในการสั่งงานเกี่ยวกับการจัดการข้อมูลโดยทำงานต่าง ๆ เหล่านี้คือ

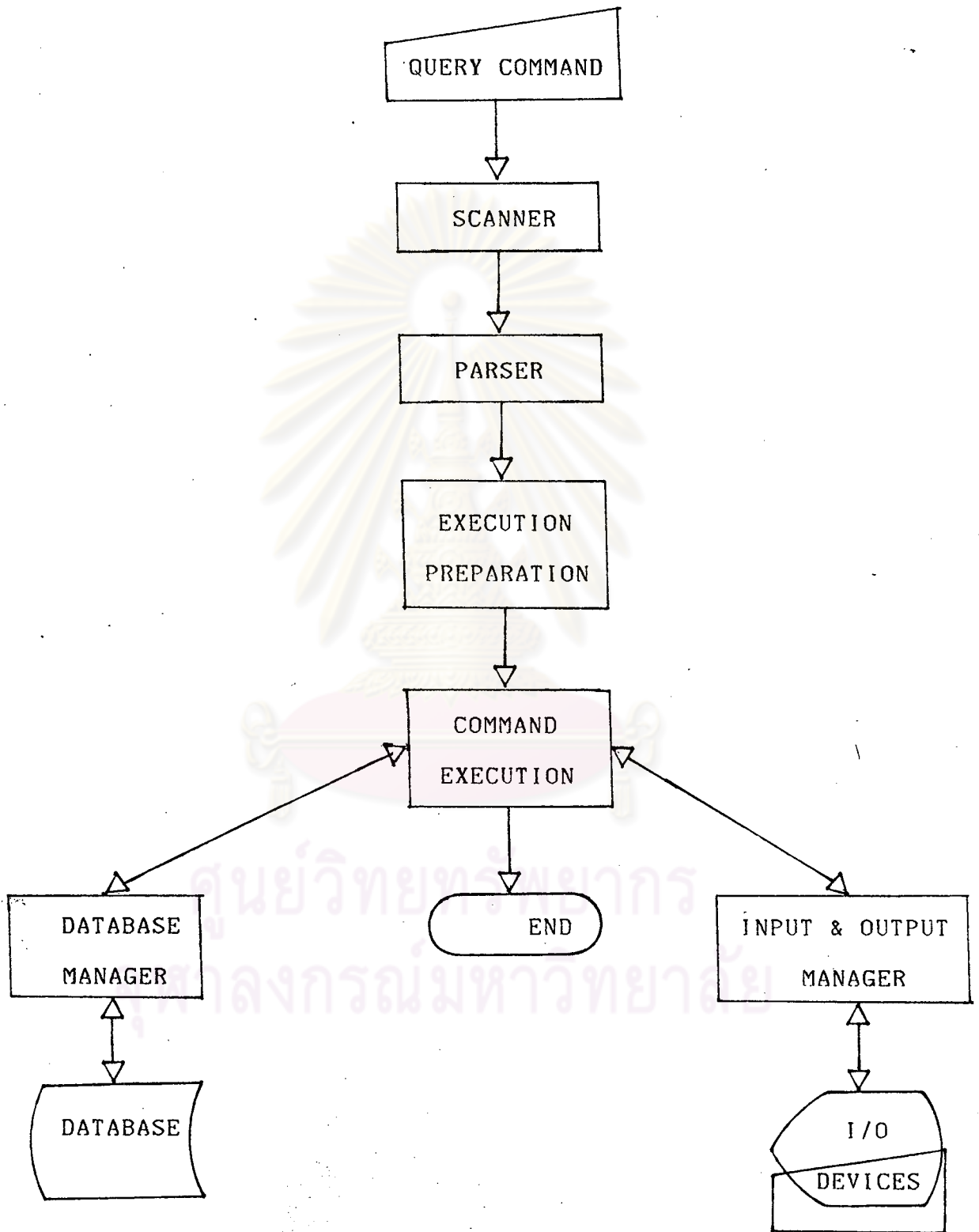
- 5.1 คัดเลือกข้อมูลจากแฟ้มข้อมูลตามเงื่อนไขที่ต้องการ
- 5.2 ใสข้อมูลเพิ่มเข้าแฟ้ม
- 5.3 แก้ไขข้อมูลในแฟ้ม
- 5.4 เปิดแฟ้มข้อมูลที่ต้องการใช้งาน

- 5.5 ลบระเบียบข้อมูลในแฟ้ม
- 5.6 สร้างแฟ้มข้อมูลให้มีโครงสร้างตามที่ต้องการ
- 5.7 แสดงโครงสร้างของแฟ้มข้อมูลที่ถูกเปิดใช้งาน
- 5.8 ย้ายระเบียบปัจจุบัน
- 5.9 ปิดแฟ้มข้อมูล

โครงสร้างของตัวประมวลผลภาษาสอบถามภาษาไทย

โครงสร้างภายในของตัวประมวลผลภาษาสอบถามภาษาไทย ประกอบด้วยส่วนประกอบต่าง ๆ ที่ทำหน้าที่ต่างๆดังนี้

1. หน่วยวิเคราะห์คำ หรือเล็กซิคอลแอนนาไลเซอร์ (Lexical Analyzer) หรือสแกนเนอร์(Scanner) ทำหน้าที่ในการวิเคราะห์คำ ในคำสั่งว่าเป็นคำสำคัญ ชื่อของเขตข้อมูล หรือค่าคงที่ จากนั้นเปลี่ยนให้อยู่ในรูปของรหัสสั้นๆที่เรียกว่าโทเคน(Token)
2. หน่วยวิเคราะห์วากยสัมพันธ์ (Syntax Analyzer) หรือพาสเซอร์(Parser) ทำหน้าที่ในการวิเคราะห์ไวยากรณ์ของประโยคคำสั่งว่า ถูกต้องหรือไม่ โดยทำการวิเคราะห์ผลลัพธ์ที่เกิดจากโปรแกรมในขั้นที่หนึ่ง
3. หน่วยเตรียมการทำงาน (Execution Preparation Module) เป็นโปรแกรมที่ทำหน้าที่ในการเตรียมองค์ประกอบต่างๆสำหรับการทำงานตามคำสั่ง เช่น การจัด(SET)ค่าของพารามิเตอร์ (Parameter)ต่างๆ
4. หน่วยปฏิบัติการตามคำสั่ง(Command Execution Module) ทำหน้าที่ในการทำงานตามคำสั่งแต่ละคำสั่ง เช่นการสร้างแฟ้ม การรับข้อมูล ฯลฯ โดยงานบางส่วนทำด้วยตัวมันเอง บางส่วนทำโดยการเรียกโปรแกรมในส่วนอื่นๆ
5. หน่วยจัดการอุปกรณ์รับและส่งข้อมูล (Input&Output Manager) ทำหน้าที่เกี่ยวกับการรับ แก่ไขข้อมูลทางจอภาพ การแสดงผลทางเครื่องพิมพ์
6. หน่วยจัดการฐานข้อมูล(Data Base Manager) ทำงานต่าง ๆ ที่เกี่ยวข้องกับแฟ้มข้อมูล เช่น การสร้างแฟ้ม การค้นข้อมูล หน้าที่นี้ทำโดยบีทีเอฟ



ผังที่ 1.1 แสดงขั้นตอนการทำงานของภาษาสอบถาม