บทที่ 1
บทนำ

1.1 การพัฒนาบุดเซอร์

ประเทศไทยมีพระพุทธศาสนาเป็นศาสนาประจำชาติ พุทธศาสนามีพระพุทธเจ้าเป็นพระบรมศาสดา พระองค์ได้เสด็จปรินิพพานไปเป็นเวลาร่วม 2532 ปี แล้ว ยังคงเหลือแต่พุทธวจนะ คือคำสั่งสอนของพระพุทธองค์เท่านั้น ซึ่งได้มีการสังคายนา หรือรวบรวมคำสอนเหล่านั้นไว้เป็นหมวดหมู่ เรียกว่า พระไตรปิฎก และนับแต่นั้นมา พระไตรปิฎก จึงถูกนำมาใช้เป็นหลักในการศึกษาปฏิบัติธรรม ของพุทธศาสนิกชนทุกหมู่เหล่า ตลอดจนใช้เป็นหลักฐานอ้างอิง ในการค้นคว้าของนักวิชาการในหลายสาขา ทั้งนี้เนื่องจากคำสอนและเรื่องราวที่จารึกไว้ในพระไตรปิฎกนั้น มีความเกี่ยวข้องกับแขนงวิชาต่าง ๆ แทบทุกสาขา โดยเฉพาะภาษาศาสตร์ ภูมิศาสตร์ ประวัติศาสตร์ มานุษยวิทยา โบราณคดี ปรัชญา จิตวิทยา รัฐศาสตร์ เศรษฐศาสตร์ และนิติศาสตร์

แต่อย่างไรก็ดี ในการศึกษาค้นคว้าพระไตรปิฎกนั้น มีอุปสรรคอยู่มาก ทั้งทางด้านภาษาบาลีที่ไม่คุ้นเคย และที่เป็นอุปสรรคสำคัญ คือ ด้านปริมาณของพระไตรปิฎก ซึ่งจากพระไตรปิฎกฉบับสยามรัฐ ที่ได้รับการจัดพิมพ์เป็นภาษาบาลีโดยใช้อักษรไทยนั้น มีจำนวนถึง 45 เล่มพิมพ์ จึงเป็นการยากที่จะทำการศึกษาค้นคว้าถึงคำสอนในเรื่องใดเรื่องหนึ่ง ที่มีปรากฏอยู่หลายแห่ง ในหลายเล่มของพระไตรปิฎกออกมาได้อย่างรวดเร็ว และครบถ้วนภายในเวลาอันสั้น นอกจากนั้น ยังต้องอาศัยการจดจำที่แม่นยำ และการจดจำได้มาก ๆ ซึ่งโดยทั่วไปแล้ว มนุษย์จะมีความสามารถที่จำกัด และไม่เท่าเทียมกันในด้านนี้ ดังนั้น หากนำเอาคอมพิวเตอร์ซึ่งมีคุณสมบัติในการเก็บบันทึกและจดจำข้อมูลได้จำนวนมาก ๆ เข้ามาเป็นเครื่องมือช่วยในการศึกษาค้นคว้าพระไตรปิฎกแล้ว ย่อมจะก่อให้เกิดประโยชน์อย่างยิ่ง

ในปัจจุบันได้มีการจัดตั้ง โครงการพัฒนาระบบคอมพิวเตอร์เพื่อการศึกษาพระไตรปิฎก ขึ้นใน มหาวิทยาลัยมหิดล โดยการนำของ ดร. ศุภชัย ตั้งวงศ์ศานต์ และได้พัฒนางานต่าง ๆ บนเครื่องไมโครคอมพิวเตอร์ประเภทไอบีเอ็มพีซี งานที่ได้ดำเนินการไปแล้ว มีดังนี้

1. จัดเก็บข้อมูลของพระไตรปิฎกฉบับภาษาบาลีอักษรไทย จำนวนทั้งหมด 45 เล่มพิมพ์ ซึ่งเป็นปริมาณข้อมูลถึง 24.23 ล้านตัวอักษรเข้าคอมพิวเตอร์
2. พัฒนาโปรแกรมคอมพิวเตอร์ชื่อ **บุดเซอร์** (BUDSIR ย่อมาจาก Buddhist Scripture Information Retrieval) สำหรับการจัดเก็บ และการสืบค้นสารสนเทศพระไตรปิฎก

3. ปริวรรตต้นฉบับพระไตรปิฎกคอมพิวเตอร์ภาษาบาลีอักษรไทย เป็นภาษาบาลีอักษรโรมัน โดยคงข้อความไว้ตามฉบับของอักษรไทยทั้งหมด พระไตรปิฎกคอมพิวเตอร์อักษรโรมันนี้ มีปริมาณของข้อมูลรวม 30.13 ล้านตัวอักษร
4. พัฒนาโปรแกรมคอมพิวเตอร์ชื่อ บุดเซอร์ทู (BUDSIR-II) เพื่อการสืบค้นสารสนเทศพระไตรปิฎกฉบับภาษาบาลีอักษรโรมัน โดยใช้หลักการเดียวกันกับ บุดเซอร์

บุดเซอร์ ได้ใช้หลักการจากสาขาวิชาการสืบค้นสารสนเทศ สร้างเป็นระบบสืบค้นสารสนเทศพระไตรปิฎกขึ้นมา บุดเซอร์ มีโครงสร้างฐานข้อมูลเป็นแบบ **แฟ้มข้อมูลย้อนกลับ** (inverted file) ประกอบด้วยแฟ้มข้อมูลหลัก 3 ประเภท คือ แฟ้มข้อมูลพระไตรปิฎก แฟ้มข้อมูลพจนานุกรม และแฟ้มข้อมูลย้อนกลับ ซึ่งได้เก็บรวบรวม คำ ต่าง ๆ จากสารสนเทศพระไตรปิฎก คำทุกคำจะมีตัวชี้ระบุถึง สารสนเทศทั้งหลายที่มีความสัมพันธ์ต่อกัน ซึ่งได้แก่ สารสนเทศที่มีคำนั้น ๆ ปรากฏอยู่ในข้อความ นอกจากนี้ ยังได้เก็บ ข้อมูลตำแหน่ง ของคำทุกคำในทุกตำแหน่งเป็นตัวเลขระบุ เล่ม/หน้า/บรรทัด/คำที่ อย่างละเอียด จากข้อมูลดังกล่าวนี้เอง บุดเซอร์ จึงสามารถสืบค้นสารสนเทศ ด้วยการระบุ คำสำคัญได้ใน 2 ลักษณะ คือ

1. <u>การสืบค้นสารสนเทศด้วย **คำเดี่ยว**</u> บุดเซอร์กำหนดให้ป้อนได้ 3 แบบ ได้แก่
   (i) คำเต็ม
   เช่น นิพฺพาน อานนฺท สมฺมาปฏิปทา
   (ii) ส่วนของคำ ที่ละส่วนท้ายไว้ ด้วยเครื่องหมาย "*" หรือ "?"
   เช่น นิพฺพาน* อานน?? สมฺมาปฏิ*
   (iii) ส่วนของคำ ที่ละต้นคำ และอาจจะละส่วนท้ายด้วย โดยใช้เครื่องหมาย "*"
   เช่น *กุมาร *กุมาร*
2. <u>การสืบค้นสารสนเทศด้วย **วลี**</u> บุดเซอร์กำหนดให้ป้อนได้ 2 แบบ ได้แก่
   (i) วลีที่ประกอบด้วย คำเต็ม ไม่เกิน 5 คำ
   เช่น "อตฺตานำ อุปมำ กตฺวา" "อาโรคฺยปรมา ลาภา"
   (ii) วลีที่ประกอบด้วย ส่วนของคำ และคำเต็ม รวมกันไม่เกิน 5 คำ
   เช่น "อตฺตา* อุปมำ กตฺว*" "อาโรคฺยปรมา ลาภ*"

โดยที่ "*" หมายถึง กลุ่มอักษรใด ๆ ไม่จำกัดความยาว ส่วน "?" หมายถึง อักษรใด ๆ 1 อักษร

1.2 ข้อจำกัดในระบบสืบค้นสารสนเทศของบุดเซอร์และแนวทางสู่ระบบใหม่

บุดเซอร์ นับเป็นก้าวแรกของความสำเร็จ ในการนำคอมพิวเตอร์มาช่วยงานด้านการศึกษาพระไตรปิฎก และเป็นที่ยอมรับของผู้ศึกษาค้นคว้าพระไตรปิฎกโดยถ้วนหน้า แต่อย่างไรก็ดี เนื่องจากบุดเซอร์ เป็นระบบสืบค้นสารสนเทศพระไตรปิฎกระบบแรก และใช้เวลาในการพัฒนาเพียง 7 เดือนเท่านั้น ฉะนั้น ความสามารถด้านการค้นของบุดเซอร์ จึงมีขอบข่ายที่จำกัดมาก คือ มีทิศทางการค้นด้วยคำสำคัญ ได้เพียง 2 ทิศทาง ได้แก่ ค้นด้วย คำเดี่ยว และ วลี เท่านั้น ซึ่งไม่อาจตอบสนองความต้องการค้นของผู้ใช้ ในระดับที่ซับซ้อนได้ นอกจากนั้น ยังขาดสิ่งอำนวยความสะดวกที่ผู้ใช้ต้องการ อีกหลายประการ ข้อจำกัดของบุดเซอร์และแนวทางแก้ไขปรับปรุง ซึ่งเป็นที่มาของการพัฒนาระบบสืบค้นสารสนเทศพระไตรปิฎกระบบใหม่ ที่เรียกว่า BUDSIR-III นั้น มีใจความสำคัญโดยสรุป ดังนี้

จากลักษณะของภาษาธรรมชาติที่ใช้บรรยายเรื่องราวต่าง ๆ นั้น การกล่าวถึงบทความเรื่องหนึ่งมักต้องอ้างอิง ด้วยคำสำคัญมากกว่า 1 คำ จึงจะสามารถเจาะจงถึงสารสนเทศในเรื่องนั้น ๆ ได้ตรงจุด เช่น information retrieval เป็นต้น ในการค้นสารสนเทศของพระไตรปิฎก ก็เช่นเดียวกัน หากใช้ คำเดี่ยว สำหรับสืบค้นสารสนเทศ มักจะได้สารสนเทศผลลัพธ์ที่มีขนาดใหญ่ และกระจัดกระจายมาก ดังนั้น หนทางแก้ไขคือ ใช้คำสำคัญตั้งแต่ 2 คำ ขึ้นไป ในการค้นหาเรื่องราวที่เฉพาะเจาะจง ลักษณะเช่นนี้ เป็นการกำหนดเงื่อนไขการค้นเพิ่มขึ้น ซึ่งทำให้ได้สารสนเทศผลลัพธ์ที่มีขนาดเล็กลง และตรงกับความต้องการหรือใกล้เคียงมากยิ่งขึ้น

การกำหนดเงื่อนไขในการค้นด้วย กลุ่มคำสำคัญ มีหลายวิธีด้วยกัน วิธีหนึ่ง คือ การระบุกลุ่มคำสำคัญเป็น **วลี** วิธีนี้ ต้องใช้ปฏิบัติการที่เรียกว่า **ปฏิบัติการประชิด** (Adjacent Operation) ซึ่งมีการตรวจสอบตำแหน่งของคำสำคัญเหล่านั้นว่า มีตำแหน่งใดที่เป็นตำแหน่งต่อเนื่องกันบ้าง สำหรับในบุดเซอร์ สามารถสืบค้นสารสนเทศด้วย วลี ได้เช่นกัน แต่ ยังจำกัด จำนวนคำ ที่จะประกอบเป็นวลีไว้ไม่เกิน 5 คำ ซึ่งในแง่ของการสืบค้นด้วย พุทธภาษิต อาจต้องระบุเกินกว่า 5 คำ เช่น
" **อปฺปมาโท อมตำ ปทำ ปมาโท มจฺจุโน ปทำ** " ดังนั้น จึงควรขยายขอบเขตการสืบค้นสารสนเทศพระไตรปิฎกด้วย วลี ออกไป โดยไม่จำกัดจำนวนคำ ที่นำมาประกอบเป็น วลีสำคัญ

นอกจากลักษณะของวลีแล้ว ยังสามารถกำหนดเงื่อนไขการค้นด้วย กลุ่มคำสำคัญ เพื่อจำกัดขอบเขตของสารสนเทศผลลัพธ์ให้แคบลง ได้อีก 2 วิธี คือ วิธีแรก กำหนดว่า กลุ่มคำสำคัญทั้งกลุ่มต้องปรากฏอยู่ในสารสนเทศเดียวกันโดยไม่จำเป็นต้องต่อเนื่องกัน และอีกวิธีหนึ่งคือ แบ่งกลุ่มคำสำคัญ เป็น 2 พวก แล้วนำไปค้นหาสารสนเทศที่ปรากฏคำศัพท์พวกแรก **แต่**ไม่ปรากฏคำศัพท์อีกพวกหนึ่งอยู่พร้อมกัน ปฏิบัติการค้นดังกล่าว มีชื่อเรียกตามลำดับว่า **ปฏิบัติการ AND** (AND Operation) และ **ปฏิบัติการ NOT** (NOT Operation) ซึ่งเป็นส่วนหนึ่งของ **วิธีปฏิบัติการแบบบูล** (Boolean Operations) ใน

วิชาการสืบค้นสารสนเทศ ซึ่งกำหนดว่าต้องมีโครงสร้างฐานข้อมูลเป็นแบบ แฟ้มข้อมูลย้อนกลับ ด้วย

จากลักษณะเฉพาะของภาษาบาลี ซึ่งเป็นภาษาที่มีวิภัติปัจจัย คือ คำศัพท์คำหนึ่งสามารถเปลี่ยนรูปไปได้หลาย ๆ แบบ ขึ้นอยู่กับหน้าที่ของคำที่จะเข้าไปประกอบในประโยค เช่น อานนฺท (อ่านว่า อา-นัน-ทะ) อานนฺทำ (อ่านว่า อา-นัน-ทัง) อานนฺโท (อ่านว่า อา-นัน-โท) หมายถึงพระอานนท์ทั้ง 3 คำ เป็นต้น ดังนั้น หากต้องการค้นสารสนเทศที่เกี่ยวข้องกับ บุคคล สัตว์ สิ่งของ หรือ สถานที่ใด ให้ได้ครบถ้วนสมบูรณ์จริง ๆ แล้ว จะต้องค้นด้วยกลุ่มคำศัพท์ทุกรูปแบบที่มีความหมายถึงสิ่งนั้น ในลักษณะต่าง ๆ ด้วยวิธีนี้ จะทำให้ได้สารสนเทศทั้งหมด ที่มีคำศัพท์ในกลุ่มนี้ปรากฏอยู่ แม้เพียงคำเดียว ลักษณะของปฏิบัติการค้นแบบนี้ มีชื่อเรียกว่า **ปฏิบัติการ OR** (OR Operation) ซึ่งเป็นอีกรูปแบบหนึ่งของปฏิบัติการแบบบูล

การนำเอาวิธีปฏิบัติการแบบบูล และปฏิบัติการประชิด เหล่านี้มาใช้ จะทำให้สามารถพัฒนาขึ้นไปสู่การค้นในระดับที่ซับซ้อนยิ่งขึ้นไปกว่านี้ได้อีก ด้วยการใช้หลายปฏิบัติการมาประกอบกัน ในการค้นคราวเดียว หรือภายในไม่กี่ขั้นตอน ทั้งนี้ จะต้องมีการกำหนดกฎเกณฑ์ของการดำเนินปฏิบัติการตามลำดับก่อนหลังไว้ด้วย และผู้ใช้ต้องเข้าใจในกฎเกณฑ์นี้ เพื่อจะได้ป้อนคำสั่งค้นได้ถูกต้องตรงกับความต้องการจริง ๆ หากผู้ใช้เติมปฏิบัติการในคำสั่งไว้ผิดลำดับแล้ว อาจได้รับสารสนเทศผลลัพธ์ที่คลาดเคลื่อนจากเป้าหมายไปได้

ในด้านการให้ความสะดวกประกอบการค้น แก่ผู้ใช้ บุดเซอร์ได้เปิดโอกาสให้ผู้ใช้สามารถค้นหากลุ่มคำศัพท์ โดยการป้อนเพียง ส่วนของคำ เข้าไป ซึ่งสะดวกแก่ผู้ใช้ที่จดจำคำได้ไม่แน่ชัด หรือผู้ใช้ที่สนใจคำศัพท์เหล่านั้นทั้งกลุ่ม ยิ่งกว่านั้น ยังเป็นประโยชน์แก่ผู้ใช้ในการค้นคำศัพท์ที่เกิดจากการสมาส-สนธิบางประเภทในภาษาบาลีได้ การป้อนส่วนของคำ ผู้ใช้ต้องเติม "*" ไว้ข้างหน้า และอาจเติม "?" หรือ "*" ต่อท้ายไว้ด้วยก็ได้ แต่อย่างไรก็ดี บุดเซอร์ก็ยังไม่สามารถค้นหา กลุ่มคำศัพท์ที่เกิดจากการสมาส-สนธิในอีกหลาย ๆ ลักษณะ ดังตัวอย่างของกลุ่มคำศัพท์ในลักษณะหนึ่ง คือ

ราโชวาท อาจาริโยวาท พุทฺโธวาท

กลุ่มคำศัพท์นี้อาจแทนด้วย "*โ?วาท" โดยที่ "?" ใช้แทนอักษรบาลีใด ๆ 1 อักษร และส่วนของคำที่กำหนดนี้จะหมายความถึง กลุ่มคำภาษาบาลี ที่ขึ้นต้นเป็นอะไรก็ได้ แต่ต้องตามด้วย "โ" และถัดมาเป็นอักษรบาลีใด ๆ 1 อักษร แล้วลงท้ายด้วย "วาท" การนำ "?" มาเติมระหว่างกลุ่มอักษร รวมทั้งกำหนดให้มีการใช้ "*" และ "?" เติมข้างหน้า และข้างท้ายของ ส่วนของคำ ได้นี้ จะทำให้สามารถ

แจกแจงรูปแบบการค้นกลุ่มคำศัพท์ได้ 3 ประเภท ดังนี้

1. ส่วนของคำ ที่มี "?" อยู่ภายในเทอม และอาจมี "*" หรือ "?" เติมอยู่ข้างท้ายด้วย
   เช่น "ปรินิพพานกา?ล" "อรห??ต??" "โ?วาท*" เป็นต้น
2. ส่วนของคำ ที่มี "?" อยู่ภายในเทอม และมี "*" นำหน้า รวมทั้งอาจมี "*" หรือ "?" เติมอยู่ข้างท้ายด้วย
   เช่น "???ปณฺฑ?ต" "???ปณฺฑ???ต??" "???ปณฺฑ???ต*" เป็นต้น
3. ส่วนของคำ ที่มี "?" อยู่ภายในเทอม และมี "?" นำหน้า รวมทั้งอาจมี "*" หรือ "?" เติมอยู่ข้างท้ายด้วย
   เช่น "*โ?วาทำ" "*โ?วาท???" "*โ?วาท*" เป็นต้น
   โดยที่ ? ใช้แทนอักษรใด ๆ 1 อักษร และอาจใส่ "?" ไว้กี่ตัวก็ได้
   * แทนกลุ่มอักษรใด ๆ ไม่จำกัดความยาว ใช้สำหรับเติมหัวหรือท้ายเท่านั้น

1.3 การออกแบบและพัฒนาระบบสืบค้นสารสนเทศพระไตรปิฎกระบบใหม่

ในการขยายขอบข่ายการสืบค้นสารสนเทศในระบบใหม่ ได้จัดให้มีการใช้ปฏิบัติการแบบบูล และปฏิบัติการประชิด เพื่อตอบสนองความต้องการสืบค้นสารสนเทศพระไตรปิฎก ในระดับที่ซับซ้อน โดยให้ผู้ใช้ระบุ คำสำคัญ มากกว่า 1 คำ และเลือกใช้ปฏิบัติการที่เหมาะสมในการสืบค้นสารสนเทศ แล้วป้อนเข้าสู่ระบบ และเพื่อให้สามารถดำเนินการสืบค้นด้วยปฏิบัติการต่าง ๆ เหล่านี้ ได้อย่างมีประสิทธิภาพ จึงจำเป็นต้องสร้างขบวนการเก็บบันทึกสารสนเทศผลลัพธ์ ของแต่ละขั้นตอนไว้ เพื่อให้สามารถนำไปประมวลผลในขั้นตอนถัด ๆ ไปได้ ด้วยวิธีการดังนี้เอง ไม่ว่าคำสั่งค้น จะเป็นรูปแบบใด มีความซับซ้อนมากน้อยแค่ไหน ก็จะสามารถประมวลผลออกมาได้หมด

จากลักษณะฐานข้อมูลพระไตรปิฎกที่เป็นแบบแฟ้มข้อมูลย้อนกลับ ทำให้สามารถเข้าถึงสารสนเทศใด ๆ ได้ทันที โดย ข้อมูลตำแหน่งคำ ที่เป็นตัวเลข จะระบุถึงสารสนเทศพระไตรปิฎก ด้วย หมายเลขเล่ม หมายเลขหน้า และยังสามารถระบุละเอียดไปถึงตำแหน่งของคำแต่ละคำที่ปรากฏในสารสนเทศนั้น ๆ ในแต่ละหน้า ว่า อยู่ใน บรรทัดใด และคำที่เท่าไรในบรรทัด คุณสมบัติของข้อมูลตำแหน่งดังกล่าว สามารถนำมาใช้ประโยชน์ ในการเก็บบันทึกสารสนเทศผลลัพธ์ที่สืบค้นมาได้ไว้ชั่วคราว โดยการเก็บข้อมูลตำแหน่ง ของคำสำคัญที่ใช้ค้นสารสนเทศเหล่านั้นไว้แทน และ **เซทของตำแหน่ง** เหล่านี้จะบ่งชี้ได้ว่า มีสารสนเทศใดบ้างที่เป็นผลลัพธ์ของการค้นนั้น ข้อมูลเซทของตำแหน่ง จะถูกเก็บไว้ในแฟ้มข้อมูลเฉพาะที่สร้างขึ้นมา มีชื่อเรียกว่า แฟ้ม F-Set และจะมีการกำหนด หมายเลขประจำ ให้กับแฟ้ม F-Set ที่ถูกสร้างขึ้นใหม่ จากการสืบค้นสารสนเทศทุกครั้ง จากหลักการเก็บข้อมูลในลักษณะพิเศษ

ดังกล่าวนี้เอง ทำให้ประหยัดเนื้อที่เก็บในหน่วยความจำสำรองได้มากกว่าอย่างเห็นได้ชัด เมื่อเทียบกับการเก็บเนื้อความสารสนเทศโดยตรง และในกรณีของการสืบค้นสารสนเทศที่ประกอบด้วยหลายขั้นตอน ก็สามารถดำเนินปฏิบัติการแบบบูล และปฏิบัติการประชิด กับแฟ้ม F-Set ได้ด้วย

นอกจากนั้น การเก็บสารสนเทศผลลัพธ์ไว้ได้ ยังให้ประโยชน์อีกบางประการ คือ ภายหลังการค้นยังสามารถเรียกสารสนเทศนั้น ๆ มาดูทางจอภาพ หรือพิมพ์ออกทางเครื่องพิมพ์ได้ทันทีโดยไม่ต้องค้นใหม่ทุกครั้งอย่างในบุดเซอร์ และยังสะดวกในการเปรียบเทียบผลลัพธ์ที่ได้จากการค้นแต่ละครั้ง

ในการเก็บแฟ้ม F-Set ดังกล่าว จะต้องมีการจัดทำระบบจัดการแฟ้มข้อมูลส่วนนี้ ซึ่งได้แก่ การสร้างแฟ้มข้อมูลเก็บเซทตำแหน่ง การลบทำลายแฟ้มข้อมูล และการเรียกดูรายการของแฟ้ม F-Set เหล่านี้ ที่ได้เก็บบันทึกไว้ในฐานข้อมูล เป็นต้น

ส่วนการประมวลผลเพื่อค้นหากลุ่มคำศัพท์ที่ถูกระบุให้เผียง ส่วนของคำ ที่จะจัดให้มีเพิ่มเติมขึ้นจากที่มีอยู่เดิมในบุดเซอร์ ซึ่งมีอยู่ 3 ลักษณะ ดังกล่าวมาแล้วนั้น ก็จะใช้ วิธีเทียบเคียงสายอักษร (string matching method) โดยนำ ส่วนของคำ ไปเทียบเคียงกับข้อมูล คำสำคัญ ในแฟ้มพจนานุกรมด้วยหลักการของ คนูท-มอริส-แพรท (Knuth-Morris-Pratt algorithm หรือ KMP algorithm) ที่ได้ปรับปรุงใหม่ เพื่อใช้ในการค้นหากลุ่มคำศัพท์ใน 3 ลักษณะนี้โดยเฉพาะ

จากลักษณะการเชื่อมโยงกับผู้ใช้ในบุดเซอร์ ซึ่งเป็นระบบเมนู จะจำกัดแนวทางการใช้งานของผู้ใช้ไว้ค่อนข้างแคบ เหมาะกับผู้ใช้เริ่มต้น แต่ในระบบสืบค้นที่พัฒนาขึ้นใหม่นี้ ได้มีการขยายขอบเขตการค้นให้มีประสิทธิภาพมากขึ้นด้วยปฏิบัติการแบบบูล ปฏิบัติการประชิด และยังสามารถเก็บบันทึกสารสนเทศจากการค้นเอาไว้ได้ ซึ่งมีความซับซ้อนมากขึ้น ในลักษณะเช่นนี้ ตัวเชื่อมโยงกับผู้ใช้ของระบบสืบค้นใหม่ต้องเป็นในแนวที่ให้ผู้ใช้เป็นผู้ควบคุมทิศทางการสืบค้นเอง จึงได้จัดให้มีลักษณะการใช้งานแบบดำเนินตามคำสั่ง (command-driven) โดยจัดกลุ่มคำสั่งสอบถามขึ้นชุดหนึ่ง เพื่อให้ผู้ใช้สามารถติดต่อสั่งงานต่าง ๆ กับระบบ ซึ่งนับว่าเหมาะสมกับลักษณะการทำงานของระบบสืบค้นสารสนเทศพระไตรปิฎกระบบใหม่นี้มาก แต่ทั้งนี้ทั้งนั้น ผู้ใช้งานต้องมีความคุ้นเคยกับระบบคอมพิวเตอร์มาก่อน และมีการใช้งานบ่อย ๆ จึงจะสามารถใช้ประโยชน์จาก BUDSIR-III ได้อย่างเต็มที่

กลุ่มคำสั่งสอบถามของ BUDSIR-III สำหรับให้ผู้ใช้สั่งงานกับระบบ ให้สืบค้นสารสนเทศ หรือเรียกใช้ฟังชั่นอำนวยความสะดวกต่าง ๆ นั้น มีดังนี้

HELP เป็น คำสั่งขอความช่วยเหลือ ซึ่งจะแสดงข้อความอธิบายวิธีใช้คำสั่งต่าง ๆ
ตัวอย่างการใช้ เช่น **HELP search**

SEARCH เป็น คำสั่งสืบค้นสารสนเทศหัวข้อใด ๆ แล้วเก็บบันทึกไว้ในเซท
การระบุความต้องการเพื่อค้นสารสนเทศจะประกอบด้วย 2 ส่วน คือ
ตัวปฏิบัติการ ได้แก่ & (ปฏิบัติการ AND), + (ปฏิบัติการ OR),
- (ปฏิบัติการ NOT), @ (ปฏิบัติการประชิด), () วงเล็บ
ตัวถูกปฏิบัติการ ได้แก่ คำบาลี และ **หมายเลขของ F-Set** (#n)
ตัวอย่างคำสั่งค้น เช่น **SEARCH อานนฺท + อานนฺโท + #1**

DISPLAY เป็น คำสั่งให้แสดงสารสนเทศผลลัพธ์ ซึ่งถูกเก็บไว้ในแฟ้ม F-Set ออกทางจอภาพ
และผู้ใช้สามารถใช้ปุ่มอำนวยความสะดวกต่าง ๆ เหมือนที่มีใช้ในบุดเซอร์
ตัวอย่างการใช้ เช่น **DISPLAY #1**

SDISPLAY เป็น คำสั่งเรียกดูรายการสารสนเทศผลลัพธ์ทั้งหมด ที่ยังมีเก็บบันทึกไว้ในแฟ้ม F-Set
ต่าง ๆ ในฐานข้อมูล
ตัวอย่างการใช้ เช่น **SDISPLAY**

FLUSH เป็น คำสั่งยกเลิกการเก็บสารสนเทศผลลัพธ์รายการหนึ่ง
ตัวอย่างการใช้ เช่น **FLUSH #2**

PRINT เป็น คำสั่งพิมพ์สารสนเทศผลลัพธ์ ที่ถูกเก็บไว้ในแฟ้ม F-Set ออกทางเครื่องพิมพ์
ตัวอย่างการใช้ เช่น **PRINT #3**

EXPAND เป็น คำสั่งค้นคำศัพท์อย่างเดียว ซึ่งอาจป้อนเป็น คำเต็ม หรือ ส่วนของคำ ในรูปแบบ
ต่าง ๆ และจะรายงาน รายการคำศัพท์ หมายเลขเล่ม จำนวนครั้งที่ปรากฏคำศัพท์
ตัวอย่างการใช้ เช่น **EXPAND *โ?วาท**

READ เป็น คำสั่งการสืบค้นสารสนเทศตามเล่ม/ข้อ/หน้า
ตัวอย่างการใช้ เช่น **READ**

PHRASE เป็น คำสั่งค้นหาวลีที่มีปรากฏอยู่จริง ๆ และรายงานจำนวนที่พบในแต่ละเล่ม
ตัวอย่างการใช้ เช่น **PHRASE อตฺตา หิ อตฺตโน นาโถ**

QUIT เป็น คำสั่งเลิกใช้งาน ระบบ BUDSIR-III และกลับสู่ DOS
ตัวอย่างการใช้ เช่น **QUIT**

1.4 ขอบเขตของการวิจัยและเวลาที่ใช้

1. ในการทำวิทยานิพนธ์ฉบับนี้ ได้พัฒนาโปรแกรมบนเครื่องไมโครคอมพิวเตอร์ 16 บิต (IBM PC/AT Compatible) ซึ่งมีหน่วยความจำหลักขนาด 640 กิโลไบต์ หน่วยเก็บข้อมูลเสมือน (virtual disk) ขนาด 384 กิโลไบต์ และหน่วยเก็บข้อมูลสำรองเป็นฮาร์ดดิสค์ขนาดประมาณ 100 เม็กกะไบต์ (ซึ่งต้องมีซอฟท์แวร์ที่ทำหน้าที่เป็นตัวกลางระหว่างตัวขับจาน (disk drive) กับระบบงาน MS-DOS ด้วย) มีจอภาพแบบสีเดียว (monochrome) และมีเครื่องพิมพ์แบบดอทเมตริกซ์
2. การพัฒนาโปรแกรมระบบสืบค้นนี้ใช้ภาษา BASIC (QUICK BASIC ของบริษัทไมโครซอฟท์)
3. โปรแกรมที่สร้างขึ้นจะใช้ฐานข้อมูลพระไตรปิฎกของบุดเซอร์ที่มีอยู่เดิม และจะสร้าง กลุ่มแฟ้มข้อมูล F-Set เพิ่มขึ้นในฐานข้อมูล ซึ่งต้องสามารถเชื่อมโยงกับฐานข้อมูลที่มีอยู่เดิม เพื่อให้สามารถทำงานได้อย่างมีประสิทธิภาพตามที่ออกแบบ

งานวิจัยนี้ ได้ใช้เวลาในการดำเนินงาน ตั้งแต่การศึกษาระบบของ BUDSIR การค้นคว้าหาแนวทางใหม่ ๆ การออกแบบ และพัฒนาระบบสืบค้นสารสนเทศพระไตรปิฎกระบบใหม่ จนสำเร็จสามารถใช้งานได้นี้ รวมเป็นเวลาทั้งสิ้น 2 ปี 4 เดือน

สำหรับในบทต่อ ๆ ไป ได้บรรยายรายละเอียด ในหัวข้อเรื่องต่าง ๆ ไว้ ดังนี้

ในบทที่ 2 จะกล่าวถึงบุดเซอร์โดยละเอียด ตั้งแต่ประวัติความเป็นมา โครงสร้างของภาษาบาลีและการจัดเก็บเข้าคอมพิวเตอร์ ฐานข้อมูลพระไตรปิฎก ตลอดจนขบวนการค้นคืนสารสนเทศพระไตรปิฎก และท้ายบทจะกล่าวถึง ข้อจำกัดในระบบสืบค้นสารสนเทศพระไตรปิฎกของบุดเซอร์

ในบทที่ 3 จะกล่าวถึง แนวความคิด หลักการต่าง ๆ ที่นำมาใช้ในการออกแบบ ระบบสืบค้นสารสนพระไตรปิฎกระบบใหม่ ซึ่งได้แก่ ลักษณะโครงสร้างของระบบสืบค้นสารสนเทศทั่วไป หลักการของขบวนการสืบค้นสารสนเทศกับความต้องการค้นของผู้ใช้ ในระดับพื้นฐานและระดับที่ซับซ้อน

บทที่ 4 จะกล่าวถึง การออกแบบระบบสืบค้นสารสนเทศพระไตรปิฎกระบบใหม่ เริ่มจากหลักการของระบบใหม่ โครงสร้างแฟ้ม F-Set ที่จะเพิ่มเข้าในฐานข้อมูลเดิม และการออกแบบปฏิบัติการแบบบูล และปฏิบัติการประชิด ในขบวนการค้นคืนสารสนเทศพระไตรปิฎกระบบใหม่

บทที่ 5 เป็นการบรรยายถึง ผลของการพัฒนาระบบใหม่ ที่มีชื่อเรียกว่า BUDSIR-III โดยกล่าวถึง การทำงานของคำสั่งแต่ละคำสั่ง ในกลุ่มคำสั่งสอบถาม ตัวอย่างการใช้งานระบบสอบถาม คุณลักษณะเฉพาะของของ BUDSIR-III และความเร็วในการค้น

ในบทสุดท้าย คือบทที่ 6 จะสรุปถึงประสิทธิภาพ ความเร็ว ความถูกต้อง รวมทั้งสิ่งอำนวยความสะดวก ของ BUDSIR-III และให้ข้อเสนอแนะถึงแนวทางต่าง ๆ ที่สามารถจะนำคอมพิวเตอร์มาช่วยในการศึกษาและเผยแพร่พระไตรปิฎก อันจะเป็นคุณประโยชน์อย่างยิ่ง