**บทที่ 3 ขั้นตอนการดำเนินโครงการ**

ในบทนี้จะกล่าวถึงวิธีการออกแบบและขั้นตอนการพัฒนาการ การปริวรรตอักษรด้วยเครื่องสำหรับภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทย โดยมีวิธีการดำเนินโครงการดังต่อไปนี้

1. ศึกษาหลักการทฤษฎีและงานวิจัยรวมถึงเทคโนโลยีที่เกี่ยวข้อง
2. จัดเตรียมแหล่งข้อมูลไว้สำหรับการศึกษาและการทดสอบระบบ
3. วิเคราะห์ข้อมูล กำหนดภาพรวมของงาน เลือกรูปแบบการพัฒนา
4. ออกแบบและกำหนด กฎ (Rule-based) ขั้นตอนวิธี (Algorithm)
5. พัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)
6. พัฒนาส่วนตัวปริวรรตอักษร(Transliteration Machine)
7. พัฒนาส่วนต่อผู้ใช้งาน(User Interface)
8. ทดสอบและประเมินผลการใช้งานจากผู้เชี่ยวชาญภาษาสันสกฤต
   1. **การศึกษาหลักการทฤษฎีและงานวิจัยรวมถึงเทคโนโลยีที่เกี่ยวข้อง**

ผู้จัดทำได้ทำการศึกษาและทำความเข้าใจหลักการทฤษฎีและงานวิจัยรวมถึงเทคโนโลยีที่เกี่ยวข้อง ได้แก่ หลักการปริวรรตภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทย ตัวอย่างงานวิจัยการปริวรรตอักษรด้วยเครื่อง และเทคโนโลยีที่ใช้ในการพัฒนาทั้งที่เป็น เครื่องมือพัฒนา ภาษาโปรแกรม ชุดไลบรารี่ ฟังก์ชัน เฟรมเวิร์ค เพื่อให้ทราบถึงการทำงาน และแนวทางในการพัฒนาการปริวรรตอักษรด้วยเครื่อง โดยมีรายละเอียดต่อไปนี้

* + 1. **การศึกษาด้านทฤษฎีและงานวิจัยที่เกี่ยวข้อง**

1. **ศึกษาลักษณะการปริวรรตอักษรภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทย**

เป็นการศึกษารวบรวมและสรุปแนวทางการปริวรรตอักษรภาษาสันสกฤตเป็นอักษรไทยในลักษณะต่างๆ จากตัวอย่างที่เกี่ยวกับการปริวรรตหรือภาษาสันสกฤตในงานวิจัยและเอกสารออนไลน์ อีกทั้งจากตัวอย่างที่ใช้กันในสังคมออนไลน์ที่มีเนื้อหาเกี่ยวกับภาษาสันสกฤต และจากการสอบถามผู้เชี่ยวชาญด้านภาษาสันสกฤตอย่างไรก็ตามหลักการยังไม่เป็นกฎเกณฑ์ที่แน่นอนตายตัวบางตำราบางอาจารย์อาจจะแตกต่างไป หรือมีการปรับเปลี่ยนกันเล็กน้อย หรืออาจจะต้องอาศัยหลักการปริวรรตอักษรจากกลุ่มภาษาเดียวกันมีการปริวรรตที่เป็นแบบแผนและมาตรฐานอยู่แล้ว เช่น ภาษาบาลี ภาษาฮินดี และสามารถสรุปลักษณะการปริวรรตจากการศึกษาได้ดังนี้

1. การปริวรรตอักษรแบบคงรูป หรือ แบบแผน คือการปริวรรตอักษรแบบตรงตามรูปศัพท์เดิมโดยใช้เครื่องหมายพินทุด้วย เพื่อให้ถอดกลับเป็นอักษรโรมันหรืออักษรเทวนาครีได้ถูกต้องและได้ลักษณะอักขรวิธีได้ใกล้เคียงกับคำเดิม

2. การปริวรรตอักษรแบบปรับรูป หรือ แบบทั่วไป คือหมายถึงการปริวรรตอักษรโดยปรับให้เข้ากับอักขรวิธีไทย เช่น โดยไม่ใช้พินทุ และเพิ่มวิสรรชนีย์หรือใส่เครื่องหมายทัณฑฆาต เพื่อให้อ่านได้ง่ายขึ้นและรูปคำกลมกลืนกับภาษาไทย จึงนิยมใช้การทับศัพท์อย่างง่าย

1. **ศึกษาด้านทฤษฎีแนวทางการพัฒนาการปริวรรตอักษร**

เป็นการศึกษาทฤษฎีทางเทคโนโลยีสารสนเทศที่เกี่ยวกับการปริวรรตอักษรว่ามีลักษณะการพัฒนาอย่างไรหลักการและแนวทางมีอะไรเพื่อใช้เป็นหลักในการพัฒนาที่เหมาะสมกับ การปริวรรตอักษรภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทย ได้ทำการศึกษาแนวทางการปริวรรตอักษร(Machine Transliteration Approaches) และโมเดลปริวรรตอักษรด้วยเครื่อง (Machine Transliteration Models)

1. **ศึกษางานวิจัยอื่นที่เกี่ยวข้อง**

ศึกษาทางด้านเทคนิคการปริวรรต หรือกระบวนการที่ใกล้เคียงกันในงานวิจัยที่ผ่านมา เพื่อหาวิธีที่เหมาะสมต่อการพัฒนาและสามารถนำมาประยุกต์ใช้กับโครงการฉบับนี้

* + 1. **การศึกษาด้านเทคโนโลยีที่เกี่ยวข้อง**

ผู้จัดทำได้ทำการศึกษาและทำความเข้าใจเทคโนโลยีที่เกี่ยวข้อง ได้แก่ เครื่องมือพัฒนา ซอฟแวร์ ภาษาโปรแกรม ชุดไลบรารี่ ฟังก์ชัน เฟรมเวิร์ค ที่ช่วยพัฒนาส่วนเสริมการปริวรรตและสำหรับการทดสอบโดยมีรายละเอียดต่อไป

1. **ไลบรารี่การปริวรรตชุดอักษรอินเดียเป็นอักษรโรมันมาตรฐานต่างๆ**

เพื่อเพิ่มความสามารถให้ระบบเพื่อรองรับอักษรต้นฉบับภาษาสันสกฤตให้หลากหลายยิ่งขึ้น ผู้จัดทำได้ค้นหาไลบรารี่การปริวรรตชุดอักษรอินเดียเป็นอักษรโรมันมาตรฐานต่างๆ เนื่องจากมีซอฟแวร์และชุดไลบรารี่อยู่หลายชุด โดยผู้ศึกษาจะเลือกได้เลือกชุดไลบรารี่ ที่เป็นโอเพนซอร์ส โดยสามารถทำการปริวรรตอักษรภาษาสันสกฤต อักษรอินเดียให้เป็นอักษรโรมันมาตรฐานต่างๆ และสามารถแปลงไปกลับได้(Forward And Backward Transliteration) และทำการศึกษาวิธีใช้งานชุดไลบรารี่นั้น

1. **ซอฟแวร์สำหรับการเปรียบเทียบข้อความ**

ในการทดสอบความถูกต้องของการแปลนั้น มีส่วนการเปรียบเทียบผลลัพธ์จากระบบ นั้นต้องเปรียบเทียบกับตัวอย่างการปริวรรตซึ่งมีข้อความอยู่เป็นจำนวนมาก ดังนั้นผู้จัดทำจึงต้องค้นหา ซอฟแวร์สำหรับการเปรียบเทียบข้อความที่สามารถบอกความแตกต่างในระดับตัวอักษรได้

* 1. **การจัดเตรียมแหล่งข้อมูลไว้สำหรับการศึกษาและการทดสอบระบบ**

ผู้จัดทำได้ทำการจัดเตรียม จัดหาแหล่งข้อมูลและเก็บรวบรวมข้อมูลไว้สำหรับการศึกษาและการทดสอบระบบ ได้แก่ ตัวอย่างสำเนาเอกสารต้นฉบับภาษาสันสกฤต ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทย โดยมีรายละเอียดต่อไป

* + 1. **แหล่งข้อมูลตัวอย่างสำเนาเอกสารต้นฉบับภาษาสันสกฤต**

ผู้จัดทำได้ทำการสืบค้นหาและเลือกแหล่งข้อมูลตัวอย่างสำเนาเอกสารต้นฉบับภาษาสันสกฤต เพื่อใช้ทดสอบการปริวรรต และใช้ประกอบการศึกษา โดยมีเงื่อนไขคือ เป็นข้อมูล ที่จัดเก็บเป็นข้อมูลอิเล็กทรอนิกส์ชนิด TEXT มีข้อมูลจัดเรียงกันเป็นหมวดหมู่และมีเอกสารอยู่หลากหลายชนิด โดยผู้จัดทำได้เลือกใช้แหล่งข้อมูลตัวอย่างสำเนาเอกสารต้นฉบับภาษาสันสกฤต จาก 2 แหล่ง เป็นแหล่งข้อมูลหลักในการศึกษาค้นคว้า ได้แก่

1. **Digital Sanskrit Buddhist Canon(DSBC)**

DSBC [1] ทำงานอยู่บนอินเตอร์เน็ตที่เว็บไซต์ www.dsbcproject.org โดยเป็นคลังเก็บข้อมูลเอกสารอิเล็กทรอนิกส์ที่เกี่ยวกับพระพุทธศาสนาในภาษาสันสกฤตหรือปรากฤต โดยที่ DSBC เป็นโครงการที่ทำการถอดความภาษาสันสกฤตหรือปรากฤตจากต้นฉบับตัวเขียน (Manuscript) และจัดเก็บเป็นข้อมูลอิเล็กทรอนิกส์ชนิด TEXT เป็นตัวอักษรเทวนา-ครีและตัวอักษรโรมันแบบ IAST โครงการ DSBC เป็นโครงการของมหาวิทยาลัยออฟ-เดอะเวสต์ (University of the West) รัฐแคลิฟอร์เนีย สหรัฐอเมริกา

1. **Göttingen Register of Electronic Texts in Indian Languages(GRETIL)**

GRETIL [1] ทำงานอยู่บนอินเตอร์เน็ตที่เว็บไซต์ gretil.sub.uni-goettingen.de โดยเป็นคลังเก็บข้อมูลเอกสารอิเล็กทรอนิกส์ในสาขาภารตวิทยา (Indology) โดยมีเอกสารทั้งภาษาสันสกฤต ปรากฤตและภาษาอินเดียอื่นๆ จากแหล่งข้อมูลจากในอินเดีย เอเชียกลาง และเอเชียตะออกเฉียงใต้ GRETIL เป็นโครงการของห้องสมุดกลางแห่งรัฐนีเดอร์ซัคเซินและมหาวิทยาลัยเกิททิงเงิน\* สหพันธ์สาธารณรัฐเยอรมนี

*\*หมายเหตุ ห้องสมุดกลางแห่งรัฐนีเดอร์ซัคเซินและมหาวิทยาลัยเกิททิงเงิน ใช้ชื่อภาษาอังกฤษ: Göttingen State and University Library ภาษาเยอรมัน: Niedersächsische Staats- und Universitätsbibliothek Göttingen หรือเรียกโดยย่อว่า SUB Göttingen[2]*

*มหาวิทยาลัยเกิททิงเงิน ใช้ชื่อในภาษาอังกฤษ : University of Göttingen ภาษาเยอรมัน :* *Georg-August-Universität Göttingen[3]*

* + 1. **ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทย**

ผู้จัดทำได้รวบรวมตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นอักษรไทย เพื่อไว้ทำการศึกษาเปรียบเทียบ ถอดกฎการปริวรรต และทดสอบระบบเพื่อเปรียบเทียบผลลัพธ์ ตัวอย่างการปริวรรตส่วนหนึ่งมาจากภาคผนวกของงานวิจัยที่เกี่ยวข้องในด้านภาษาสันสกฤตและจากแหล่งข้อมูลออนไลน์ โดยมากเป็นงานที่ปริวรรตมาจากต้นฉบับอักษรเทวนาครี บางส่วนนั้นไม่แสดงข้อความต้นฉบับ ผู้จัดทำจำเป็นต้องหาต้นฉบับในแหล่งข้อมูลข้างต้น ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทยที่ผู้จัดทำนำมาเป็นแนวทางหลักแบ่งเป็น 2 ประเภทได้แก่

1. **ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทยแบบคงรูปหรือแบบแผน**

ผู้จัดทำใช้ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทยแบบคงรูปหรือแบบแผน จาก 5 คัมภีร์เป็นแนวทางหลัก แต่ละคัมภีร์มีที่มาดังต่อไปนี้

* 1. **คัมภีร์มูลสรวาสติวาทวินยวัสตุ จีวรวัสตุและกรรมวัสตุ**

จากภาคผนวก ก ปริวรรตอักษรไทยของจีวรวสฺตุและกรฺมวสฺต ในวิทยานิพนธ์ของมหาวิทยาลัยศิลปากร ปี 2552 เรื่องการศึกษาเปรียบเทียบคัมภีร์มูลสรวาสติวาทวินยวัสตุกับพระวินัยปิฎก ศึกษาเฉพาะกรณีจีวรวัสตุและกรรมวัสตุ โดย พระมหาฉัตร ชัยมูลสาร[]

* 1. **คัมภีร์ทิวยาวทาน**

จากภาคผนวก ข บทปริวรรตทิวยาวทาน ในวิทยานิพนธ์ของมหาวิทยาลัยศิลปากรปี 2556 เรื่องการศึกษาวิเคราะห์คัมภีร์ทิวยาวทานเรื่องที่ 1-19 โดย นางสาวปัทมา นาควรรณ[1]

* 1. **คัมภีร์โพธิจรรยาวตาร**

จากภาคผนวก ก คัมภีร์โพธิจรรยาวตาร ฉบับปริวรรต ในวิทยานิพนธ์ของมหาวิทยาลัยศิลปากร ปี 2549 เรื่อง การศึกษาเชิงวิเคราะห์คัมภีร์โพธิจรรยาวตาร โดย พระมหาวิชาญ กำเหนิดกลับ[2]

* 1. **คัมภีร์มูลมัธยมกการิกา**

จากภาคผนวก ก คัมภีร์มูลมัธยมกการิกา ฉบับปริวรรต ในวิทยานิพนธ์ของมหาวิทยาลัยธรรมศาสตร์ ปี 2554 เรื่อง การศึกษาเชิงวิเคราะห์การใช้เหตุผลแบบวิภาษวิธีในคัมภีร์มูลมัธยมกการิกา โดย นายกฤษฎา ภูมิศิริรักษ์[3]

* 1. **คัมภีร์ภควัทคีตา**

จากหนังสือภควัทคีตา(ฉบับโศลก )เกียรติขจร ชัยเธียร[4]

* 1. **คัมภีร์วัชรเฉทิกปรัชญาปารมิตาสูตร ฉบับอักษรไทย**

จาก Internet Archive: Digital Library โดย กรกิจ ดิษฐาน []

1. **ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทยแบบปรับรูปหรือแบบทั่วไป**

เนื่องจากเอกสารประเภทนี้มีอยู่น้อยมาก และไม่ค่อยแพร่หลายเนื่องจากมีผู้ปริวรรตน้อยมาก ไม่ได้รับความนิยมเท่าภาษาบาลีที่มีผู้ปริวรรตเป็นบทสวดมนต์สำหรับบุคคลทั่วไปอยู่เป็นจำนวนมาก หรือหากพบก็พบเป็นข้อความหรือบทสั้นๆ ที่บุคคลทั่วที่สนใจปริวรรตกันขึ้นเองซึ่งปรากฏในสังคมออนไลน์และเว็บบอร์ดในกลุ่มที่สนใจในภาษาและวรรณคดีสันสกฤต กลุ่มที่สนใจหรือนับถือศาสนาพราหมณ์-ฮินดูและกลุ่มที่สนใจ ศาสนาพุทธแบบมหายาน ซึ่งส่วนมากก็เป็นในลักษณะการถอดเสียงไม่ใช่การปริวรรตอักษร ดังนั้นผู้จัดทำจึงใช้วิธีการ ใช้เอกสารที่เป็นบทสวดมนต์ภาษาบาลีที่มีอยู่ทั่วไปประกอบเข้าบางส่วนแต่ในที่นี้จะไม่กล่าวถึง

ส่วนเอกสารการปริวรรตประเภทนี้ที่ผู้จัดทำนำมาเป็นตัวอย่างเปรียบเทียบนั้นเป็นเอกสารที่เป็นผลงานการปริวรรตของคุณ กรกิจ ดิษฐาน ผู้ช่วยบรรณาธิการต่างประเทศและกีฬา หนังสือพิมพ์ M2F(2559) [] ที่เผยแพร่ในอินเตอร์เน็ต ผู้จัดทำใช้ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทยแบบปรับรูปหรือแบบทั่วไป 3 ตัวอย่างดังต่อไปนี้

* 1. **คัมภีร์ลลิตวิสตระ อัธยายที่ 26 ธรรมจักรประวรรตนสูตร**

จากห้องสมุดเสรีออนไลน์ วิกิซอร์ซ โดย กรกิจ ดิษฐาน []

* 1. **คัมภีร์วัชรเฉทิกปรัชญาปารมิตาสูตร ฉบับคำอ่านอักษรไทย**

จาก Internet Archive: Digital Library โดย กรกิจ ดิษฐาน []

* 1. **คัมภีร์ปรัชญาปารมิตาหฤทัยสูตร**

จากสารานุกรมเสรีออนไลน์ วิกิพีเดีย โดย กรกิจ ดิษฐาน []

* 1. **การวิเคราะห์ข้อมูล กำหนดภาพรวมของงาน เลือกรูปแบบการพัฒนา**

ผู้จัดทำได้ทำการวิเคราะห์ข้อมูลที่เก็บรวบรวมมาได้และกำหนดภาพรวมของงาน นำข้อมูลที่วิเคราะห์ได้มา เลือกรูปแบบการพัฒนาและเทคโนโลยี ให้เหมาะสม โดยมีรายละเอียดดังต่อไปนี้

* + 1. **สรุปวิธีการปริวรรตอักษรที่จะนำมาพัฒนา**

เนื่องจากกฎการปริวรรตอักษรภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทยที่รวบรวมมานั้นบางส่วนมีความแตกต่างในการปริวรรตอยู่มาก อาจมีกฎอื่นๆที่ใช้กันแต่ผู้จัดทำยังรวบรวมมาไม่หมด หรือไม่กฎเหล่านั้นก็ยังคลุมเครืออยู่ยังไม่มีแนวทางหลักในการปริวรรตโดยเฉพาะการปริวรรตแบบปรับรูปมีปัญหามากที่สุด และเพื่อการพัฒนาให้เหมาะสมกับข้อจำกัดทางเวลา ผู้จัดทำจัดเป็นไปได้ยากที่จะจัดทำได้ครบทุกกฎ ดังนั้นจึงเลือกกฎจากทฤษฎีการปริวรรตอักษรภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทยดังสรุปได้ต่อไปนี้

**วิธีการปริวรรตเป็นอักษรไทยแบบคงรูป หรือ แบบแผน**

* 1. การถอดรูปพยัญชนะ สระ เครื่องหมาย ตัวเลข หากมีวิธีเดียวให้ใช้ให้ใช้วิธีนั้น(อ้างอิงในบทที่ 2)
  2. ในกรณีถอดรูปพยัญชนะอักษรโรมันที่ประสมด้วยสระ ṛ ṝ ḷ ḹ (ฤ ฤๅ ฦ ฦๅ) มี 2 วิธี ให้ใช้ แบบไม่ใส่พินทุ
  3. รูปเครื่องหมายจันทรพินทุหรืออนุนาสิกะ มี 3 วิธี ให้ใช้ แบบไม้หันอากาศและมีนิคหิตข้างบน
  4. รูปเครื่องหมายวิสรรคะ มี 2 วิธี ให้ใช้ แบบวิสรรชนีย์ (ะ)
  5. รูปเครื่องหมายอวครหะ ไม่ถอดใช้ตามเดิม
  6. รูปเครื่องหมาย ทัณฑะและทวิทัณฑะ ถอดเป็น ฯ (อังคั่นเดี่ยว) ๚ (อังคั่นคู่)
  7. การถอดรูปพยัญชนะสังโยค หรือ พยัญชนะสังยุกต์ ถอดเฉพาะชุดพยัญชนะอักษรโรมัน ที่เป็นพยัญชนะสังโยคซ้อนหรือควบพยัญชนะโรมัน r (รูปพยัญชนะ ร)

**วิธีการปริวรรตเป็นอักษรไทยแบบปรับรูป หรือ แบบทั่วไป**

* 1. การถอดรูปพยัญชนะ สระ เครื่องหมาย ตัวเลข หากมีวิธีเดียวให้ใช้ให้ใช้วิธีนั้น(อ้างอิงในบทที่ 2)
  2. กรณี ร-เรผะ (รฺ) คือพยัญชนะอักษรโรมันประสมกับสระ a (รูปสระ อะ) ตามด้วยพยัญชนะอักษรโรมันตัว r ที่ไม่มีสระอักษรโรมันกำกับหรือตามมา (ar) มี 2 วิธี ให้ใช้ แบบ ร- หัน
  3. กรณีพยัญชนะอักษรโรมันประสมกับสระ a (รูปสระ อะ) ตามด้วยพยัญชนะอักษรโรมันตัว v ที่ไม่มีสระอักษรโรมันกำกับหรือตามมา (av) มี 2 วิธี ให้ใช้ แบบ วิสรรชนีย์
  4. กรณีพยัญชนะอักษรโรมันประสมกับสระ a (รูปสระ อะ) ตามด้วยพยัญชนะอักษรโรมันตัว h ที่ไม่มีสระอักษรโรมันกำกับหรือตามมา (ah) มี 2 วิธี ให้ใช้ แบบ ตัดทิ้ง
  5. ถอดรูปเครื่องหมายอนุสวาระและเครื่องหมายจันทรพินทุ พยัญชนะวรรคโรมันตามติดมา ให้ใช้ แบบมีการสนธิ
  6. ถอดรูปเครื่องหมายอนุสวาระและเครื่องหมายจันทรพินทุและมีพยัญชนะอวรรคโรมันที่เป็นอูษมะ ได้แก่ ś ṣ s h ตามติดมา ให้ใช้ แบบไม่มีการสนธิ
  7. รูปเครื่องหมายวิสรรคะ แบบ ห ใส่ทัณฑฆาต
  8. รูปเครื่องหมาย ทัณฑะและทวิทัณฑะ ถอดเป็น ฯ (อังคั่นเดี่ยว) ๚ (อังคั่นคู่)
  9. การถอดรูปพยัญชนะสังโยค หรือ พยัญชนะสังยุกต์ ที่ประสมกับรูปสระ e ai o au (รูปสระ เอ ไอ โอ เอา) ไม่มีแบบแน่นอน แต่ผู้จัดทำเห็นว่าสมควรว่าใช้แบบคงรูปคือ ถอดเฉพาะชุดพยัญชนะอักษรโรมัน ที่เป็นพยัญชนะสังโยคซ้อนหรือควบพยัญชนะโรมัน r (รูปพยัญชนะ ร)
     1. **สรุปแนวทางและโมเดลที่จะนำมาพัฒนา**

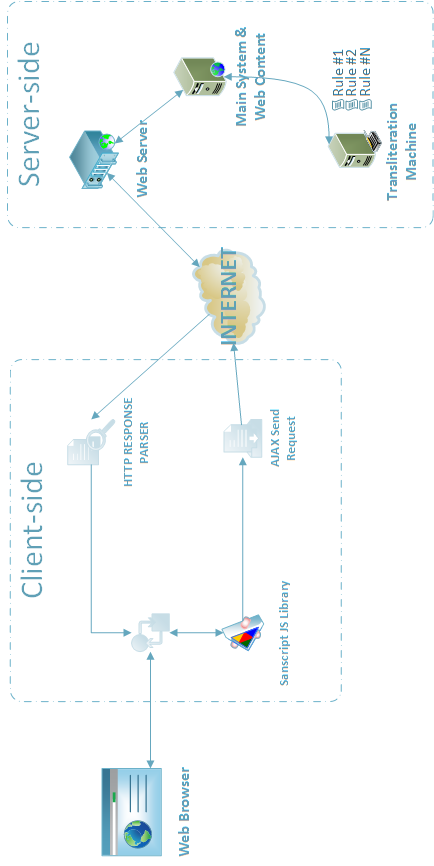
ในการเลือกแนวทางการปริวรรตอักษร(Machine Transliteration Approaches)เนื่องการศึกษาค้นคว้าจนได้รูปแบบกฎเกณฑ์การปริวรรต ผู้จัดทำจึงเลือกการปริวรรตอักษรด้วยเครื่องแบบใช้ฐานกฎ (Rule-based Approach)และในการเลือกโมเดลการปริวรรตอักษรด้วยเครื่อง (Machine Transliteration Models) ผู้จัดทำได้พิจารณาเห็นว่า ตัวอักษรโรมันแบบ IAST กับอักษรไทยนั้นครบคู่กัน อาจจะมาจากสาเหตุที่ไทยได้รับอิทธิพลภาษาสันสกฤตอย่างมาก ถึงแม้ว่าอักษรไทยบางตัวออกเสียงไม่เหมือนเสียทีเดียว แต่มองในรูปศัพท์แล้วจะให้ความหมายเดียวกัน ผู้จัดทำจึงเลือก โมเดลแบบใช้ตัวอักษร (Grapheme Based Model) ในการพัฒนา

* + 1. **สรุปเทคโนโลยีที่ช่วยพัฒนาส่วนเสริมการปริวรรตและการทดสอบ**

ไลบรารี่การปริวรรตชุดอักษรอินเดียเป็นอักษรโรมันมาตรฐานต่างๆ นั้นผู้จัดทำเลือกไลบรารี่ Sanscript [10],[11] เป็นไลบรารี่จาวาสคริปท์ที่สามารถทำการปริวรรตอักษรภาษาสันสกฤต อักษรอินเดียหลากหลายตัวอักษรให้เป็นอักษรโรมันมาตรฐานต่างๆ และสามารถแปลงไปกลับได้(Forward And Backward Transliteration) แม้ว่า Sanscript มุ่งเน้นการปริวรรตอักอักษรในภาษาสันสกฤต แต่ก็ยังสนับสนุนการปริวรรตสำหรับภาษาอินเดียอื่น ๆ อีกด้วยและยังสามารถติดตั้งและเรียกใช้งานโดยง่าย สำหรับซอฟแวร์สำหรับการเปรียบเทียบข้อความ ผู้จัดทำเลือก Pretty Diff v2.0.0 []ที่สามารถเปรียบเทียบข้อความเป็นจำนวนมากได้และสามารถเปรียบเทียบข้อความที่สามารถบอกความแตกต่างในระดับตัวอักษร พร้อมทั้งออกรายงานได้

* + 1. **สรุปภาพรวมของงาน**

สรุปภาพรวมของงานระบบจะทำงานในลักษณะ Server-Client โดยทำงานบนเว็บบราวเซอร์ โดยกระบวนการฝั่ง Client และการส่ง Request นั้น จะทำโดยผู้ใช้สามารถป้อนข้อความภาษาสันสกฤตโดยการเลือกชนิดตัวอักษรจากนั้น ไลบรารี่ Sanscript จะทำการปริวรรตอักษรชนิดต่าง ๆ และแสดงบนเว็บบราวเซอร์ได้ทันที่เนื่องจากเป็นไลบรารี่จาวาสคริปท์ที่จะทำงานในฝั่ง Client แต่หากจะต้องการปริวรรตอักษรชนิดอื่น ๆ เป็นภาษาไทยข้อความจะถูกส่งให้ไลบรารี่ Sanscript และปริวรรตชนิดตัวอักษรเหล่านั้นให้เป็นอักษรโรมันมาตรฐาน IAST แล้วส่งไปให้ Server โดยวิธีการ AJAX (Asynchronous JavaScript and XML) เมื่อระบบหลัก(Main System)รับคำขอดังกล่าวแล้วจะส่งข้อความให้กับ Transliteration Machine เพื่อทำการประมวลผลต่อไป ส่วนกระบวนการทางฝั่ง Server และ Response นั้น เมื่อ Transliteration Machine ได้ทำการประมวลผลจะนำข้อความไปผ่านกระบวนการตามกฎที่ได้กำหนดไว้ และคืนค่าเป็นผลการปริวรรตแบบคงรูปหรือแบบแผนและแบบปรับรูปหรือแบบทั่วไป โดยออกมาในรูปแบบ JSON (JavaScript Object Notation) จากนั้นก็ส่งต่อให้ระบบหลัก และส่งกลับไปยัง Client โดยมีตัวแปลง Response โดยจะทำหน้าที่แปลง JSON ที่ได้รับมาก แปลงให้เป็นตารางเปรียบเทียบข้อความตามชนิดตัวอักษร และแสดงผลลัพธ์การปริวรรตบนเว็บบราวเซอร์



รูปที่ 2.5 ผลลัพธ์หลังจากการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)

* 1. **การออกแบบและกำหนด กฎ (Rule-based) ขั้นตอนวิธี (Algorithm)**

ผู้จัดทำได้ทำการออกแบบและกำหนด กฎ (Rule-based) ขั้นตอนวิธี (Algorithm) ในการปริวรรตอักษรกฎ (Rule-based) วิธีการปริวรรตเป็นอักษรไทยแบบคงรูป หรือ แบบแผน

* + 1. **สัญลักษณ์ประกอบขั้นตอนวิธีของกฎการปริวรรตอักษรไทย**

ตารางที่ 2.2 การเปรียบเทียบรูปพยัญชนะวรรค

|  |  |
| --- | --- |
| **สัญลักษณ์** | **ชุดข้อมูลหรืออำอธิบายการทำงาน** |
| **TA** | พยัญชนะและสระอักษรไทยใด ๆ |
| **TS** | อักษรไทยใด ๆ ที่ไม่ใช้พยัญชนะและสระ |
| **TC** | พยัญชนะอักษรไทยใด ๆ |
| **TV** | สระอักษรไทยใด ๆ |
| **RNP** | ชุดตัวเลขและเครื่องหมายวรรคตอนอักษรโรมันได้แก่ {0,1 ,2 ,3 ,4 ,5 ,6 ,7 ,8 ,9 ,॥,।} |
| **TNP** | ชุดตัวเลขและเครื่องหมายวรรคตอนอักษรไทยได้แก่  {๐, ๑, ๒, ๓, ๔, ๕, ๖, ๗, ๘, ๙, ๚, ฯ} |
| **RMC** | ชุดพยัญชนะอักษรโรมันประกอบกัน 2 ตัวอักษรได้แก่  { kh,ch,ṭh,th,ph,gh,jh,ḍh,dh,bh } |
| **TMC** | ชุดพยัญชนะอักษรไทยที่แปลงมาจาก RMC ได้แก่  {ข,ฉ,ฐ,ถ,ผ,ฆ,ฌ,ฒ,ธ,ภ} |
| **RMV** | ชุดสระอักษรโรมันประกอบกัน 2 ตัวอักษรได้แก่ {ai,au} |
| **TMV** | ชุดสระอักษรไทยที่แปลงมาจาก RMV ได้แก่{ไ,เา} |
| **RSC** | ชุดพยัญชนะอักษรโรมันที่ตัวอักษรเดี่ยวได้แก่  {k,g,ṅ,c,j,ñ,ṭ,ḍ,ṇ,t,d,n,p,b,m,y,r,l,v,ḻ,ś,ṣ,s,h,ṁ,ṃ,ḥ,' } |
| **TSC** | ชุดพยัญชนะอักษรไทยที่แปลงมาจาก **RSC** ได้แก่  {ก,ค,ง,จ,ช,ญ,ฏ,ฑ,ณ,ต,ท,น,ป,พ,ม,ย,ร,ล,ว,ฬ,ศ,ษ,ส,ห,ँ,ํ,ะ,'} |
| **TSC2** | ชุดพยัญชนะอักษรไทยที่มาจาก **TSC** ใช้ในแบบปรับรูป แตกต่างที่ตำแหน่งวิสรรคะ(ห์)  {ก,ค,ง,จ,ช,ญ,ฏ,ฑ,ณ,ต,ท,น,ป,พ,ม,ย,ร,ล,ว,ฬ,ศ,ษ,ส,ห,ँ,ํ,ห์,'} |
| **RSV** | ชุดสระอักษรโรมันที่เป็นตัวอักษรเดี่ยวได้แก่  { a,ā,i,ī,u,ū,ṛ,ṝ,ḷ,ḹ,e,o } |
| **TSV** | ชุดสระอักษรไทยที่แปลงมาจาก **RSV** ใช้ในแบบคงรูป ได้แก่  { a,า,ิ,ี,ุ,ู,ฤ,ฤๅ,ฦ,ฦๅ,เ,โ } |
| **TSV2** | ชุดสระอักษรไทยที่มาจาก **TSV** ใช้ในแบบปรับรูป แตกต่างที่ตำแหน่งแรก ได้แก่{ะ,า,ิ,ี,ุ,ู,ฤ,ฤๅ,ฦ,ฦๅ,เ,โ } |
| **TIF** | ชุดสระอักษรไทยเมื่ออยู่ต้นคำต้องทำการเปลี่ยนแปลง ใช้ในแบบแบบคงรูป ได้แก่{ A ,า,ิ,ี,ุ,ู,เ,โ,ไ} |
| **TIFR** | ชุดสระอักษรไทยเมื่ออยู่ต้นคำเมื่อเปลี่ยนแปลงแล้ว ใช้ในแบบแบบคงรูปได้แก่ {อ, อา, อิ, อี, อุ, อู, เอ, โอ, ไอ} |
| **TIF2** | ชุดสระอักษรไทยชุดเดียวกันกับ **TIF-F** แต่ใช้ในแบบปรับรูป แตกต่างที่รูปตำแหน่งแรก ได้แก่ { ะ,า,ิ,ี,ุ,ู,เ,โ,ไ} |
| **TIFR2** | ชุดสระอักษรไทยชุดเดียวกันกับ **TIFR-F** แต่ใช้ในแบบปรับรูป แตกต่างที่รูปตำแหน่งแรก ได้แก่ {อะ, อา, อิ, อี, อุ, อู, เอ, โอ, ไอ} |
| **F-RS**(X,Y,Z)  Replace String Function | X คือค่าพารามิเตอร์ที่เป็นตัวอักษรที่จะเปลี่ยน Y คือค่าพารามิเตอร์ที่เป็นตัวอักษรที่ไปแทนที่  Z คือค่าพารามิเตอร์ที่เป็นข้อความ  **F-RS**() คือฟังก์ชั่นการการแทนที่ตัวอักษร X ในข้อความ Zทั้งหมดด้วยตัวอักษร Y เช่น  X = C , Y = R , Z = CAT ดังนั้น RS(X,Y,Z) = RAT |
| **F-RSA** (X,Y,Z)  Replace String with Array Function | X คือค่าพารามิเตอร์ที่เป็นชุดตัวอักษรที่จะเปลี่ยน  Y คือค่าพารามิเตอร์ที่เป็นชุดตัวอักษรที่ไปแทนที่  X และ Y ต้องมีสมาชิกจำนวนเท่ากัน  Z คือค่าพารามิเตอร์ที่เป็นข้อความ  **F-RSA**() คือฟังก์ชั่นการการแทนที่ชุดตัวอักษร X ในข้อความ Z ทั้งหมดด้วยชุดตัวอักษร Y เช่น X คือพารามิเตอร์ของชุดข้อมูลที่จะ  X = {A,B} , Y = {APP,L} , Z = ABE เช่น ดังนั้น RSA(X,Y,Z) = APPLE |

* + 1. **กำหนดกฎการปริวรรตอักษรไทยแบบคงรูป**

ตารางที่ 2.2 การเปรียบเทียบรูปพยัญชนะวรรค

|  |  |  |
| --- | --- | --- |
| **ลำดับ** | **ชื่อกฎ** | **หน้าที่** |
| 1. | RomanChandrabinduToSingle | เปลี่ยนเครื่องหมายจันทรพินทุให้เป็นแบบตัวเดียว |
| 2. | RomanNumberAndPunctuate | เปลี่ยนตัวเลขและเครื่องหมายวรรคตอน |
| 3. | RomanMixConsonant | เปลี่ยนพยัญชนะโรมันแบบตัวผสม |
| 4. | RomanMixVowel | เปลี่ยนสระโรมันแบบตัวผสม |
| 5. | RomanSingleConsonant | เปลี่ยนพยัญชนะโรมันแบบตัวเดี่ยว |
| 6. | RomanSingleVowel | เปลี่ยนสระโรมันแบบตัวเดี่ยว |
| 7. | Bindu | ใส่จุดพินทุให้พยัญชนะไทย |
| 8. | ThaiVowelInFist | เปลี่ยนสระกรณีอยู่ต้นคำ โดยการเติมรูปพยัญชนะ อ ผสมเข้าไป |
| 9. | ThaiVowelPrefix | เลื่อนสระที่เป็นสระหน้าไปหน้าพยัญชนะ |
| 10. | ARemove | ลบสระ A |
| 11. | SwapAnusvaraAndChandrabindu | เลื่อนเครื่องหมายอนุสาวระอักษรไทยและจันทรพินทุอักษรโรมันไปข้างหน้ากรณีอยู่หลัง ลากข้าง(า) |
| 12. | Chandrabindu | แปลงเครื่องหมายจันทรพินทุ |
| 13. | ThaiAAInFist | เปลี่ยนลากข้าง(า) โดยผสมรูปพยัญชนะ อ กรณีอยู่ถัดรูปสระอื่น ซึ่งเป็นสระตัวสุดท้ายที่จะแปลง เนื่องจากต้องแปลงรูปสระเอา ก่อน |

* + 1. **กำหนดกฎการปริวรรตอักษรไทยแบบปรับรูป**

ตารางที่ 2.2 การเปรียบเทียบรูปพยัญชนะวรรค

|  |  |  |
| --- | --- | --- |
| **ลำดับ** | **ชื่อกฎ** | **หน้าที่** |
| 1. | AvagarahaRemove | ลบอวครหะ |
| 2. | RomanChandrabinduToSingle | เปลี่ยนเครื่องหมายจันทรพินทุให้เป็นแบบตัวเดียว |
| 3. | RomanNumberAndPunctuate | เปลี่ยนตัวเลขและเครื่องหมายวรรคตอน |
| 4. | RomanMixConsonant | เปลี่ยนพยัญชนะโรมันแบบตัวผสม |
| 5. | RomanMixVowel | เปลี่ยนสระโรมันแบบตัวผสม |
| 6. | RomanSingleConsonant | เปลี่ยนพยัญชนะโรมันแบบตัวเดี่ยว |
| 7. | RomanSingleVowel | เปลี่ยนสระโรมันแบบตัวเดี่ยว |
| 8. | AnusvaraAndChandrabindu | เข้าสนธิอนุสวาระและจันทรพินทุ |
| 9. | ThaiVowelInFist | เปลี่ยนสระกรณีอยู่ต้นคำ โดยการเติมรูปพยัญชนะ อ ผสมเข้าไป |
| 10. | ThaiVisarga | เปลี่ยนข้อความที่เป็นประวิสรรชณีย์ |
| 11. | ThaiVowelPrefix | เลื่อนสระที่เป็นสระหน้าไปหน้าพยัญชนะ |
| 12. | ThaiAAInFist | เปลี่ยนลากข้าง(า) โดยผสมรูปพยัญชนะ อ กรณีอยู่ถัดรูปสระอื่น |
| 13. | convertAE | เลื่อนสระที่เป็นสระ เ ไปหน้าพยัญชนะหากตัวอักษรก่อนหน้าถูกแปลงเป็นไม้หันอากาศแล้วต้องแปลงกลับเป็นวิสรรชณีย์(ะ) |
| 14. | convertAO | เลื่อนสระที่เป็นสระ โ ไปหน้าพยัญชนะหากตัวอักษรก่อนหน้าถูกแปลงเป็นไม้หันอากาศแล้วต้องแปลงกลับเป็นวิสรรชณีย์(ะ) |

* + 1. **ขั้นตอนวิธีของกฎการปริวรรตอักษรไทยแบบคงรูป**

ขั้นตอนวิธีของกฎการปริวรรตอักษรไทยแบบคงรูป จะทำการตามลำดับกฎดังต่อไปนี้

**กฎที่ 1. RomanChandrabinduToSingle**

เปลี่ยนเครื่องหมายจันทรพินทุ จากการเขียนด้วยตัวอักษร m และเครื่องหมาย Dot Moonที่อยู่ด้านบน(m̐)ให้เป็นเครื่องหมายจันทรพินแบบตัวอักษรเดี่ยวคือ ṁ ออกแบบไว้เพื่อรองรับไว้สำหรับการใช้เครื่องหมายจันทรพินทุโรมันทั้งสองแบบ และเตรียมไว้สำหรับการแปลงต่อไปโดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RS**(‘m̐’ , ‘ṁ', input)
4. RETURN output
5. END

**กฎที่ 2. RomanNumberAndPunctuate**

เปลี่ยนตัวเลขและเครื่องหมายวรรคตอนอักษรโรมันให้เป็นตัวเลขและเครื่องหมายวรรคตอนอักษรไทย โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RNP** , **TNP**, input)
4. RETURN output
5. END

**กฎที่ 3. RomanMixConsonant**

เปลี่ยนชุดพยัญชนะอักษรโรมันที่มีลักษณะอักษรที่ประกอบกันมากกว่า 1 ตัว ให้เป็นชุดพยัญชนะอักษรไทย โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RMC**, **TMC**, input)
4. RETURN output
5. END

**กฎที่ 4. RomanMixVowel**

เปลี่ยนชุดสระอักษรโรมันที่มีลักษณะอักษรที่ประกอบกันมากกว่า 1 ตัว ให้เป็นชุดสระอักษรไทย โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RMV**, **RMV**, input)
4. RETURN output
5. END

**กฎที่ 5. RomanSingleConsonant**

เปลี่ยนชุดพยัญชนะอักษรโรมันที่เป็นอักษรเดี่ยว ให้เป็นชุดพยัญชนะอักษรไทย ในชุด **TSC** มีกรณีพิเศษคือ เครื่องหมายจันทรพินแบบตัวอักษรเดี่ยวคือ ṁ จะถูแปลงเป็น Moon dot (ँ )ก่อนเพื่อเตรียมประมวลผลในขั้นตอนต่อไป โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RSC** , **TSC**, input)
4. RETURN output
5. END

**กฎที่ 6. RomanSingleVowel**

เปลี่ยนชุดสระอักษรโรมันที่เป็นอักษรเดี่ยว ให้เป็นชุดสระอักษรไทย โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RSV**, **TSV**, input)
4. RETURN output
5. END

**กฎที่ 7. Bindu**

เติมเครื่องหมายพินทุในพยัญชนะอักษรไทยที่ไม่มีสระกำกับ โดยมีขั้นตอนวิธีดังนี้

1. # anusvara เครืองหมายนิคหิต(ํ)
2. # chandrabindu เครืองหมาย Moon dot( ँ )
3. # avagaraha เครืองหมาย apostrophe( ‘ )
4. # bindu เครืองหมายพินทุ( ฺ )
5. # ThaiVisarga เครืองหมายวิสรรชนีย์(ะ)
6. START
7. READ input
8. FOR each character of the input
9. current character = character
10. next character = input[current index +1]
11. condition = [
12. current character NOT IN **TC** And
13. current character ≠ anusvara And
14. current character ≠ chandrabindu And
15. current character ≠ ThaiVisarga And
16. current character ≠ Avagaraha And
17. next character ≠ "a" And
18. next character NOT IN **TV**
19. ]
20. IF condition THEN
21. character = bindu
22. ENDIF
23. output = output Concatenated character
24. ENDFOR
25. RETURN output
26. END

**กฎที่ 8. ThaiVowelInFist**

เพิ่มรูปพยัญชนะ อ เข้าไปในกรณีสระอักษรไทยอยู่ต้นคำ แต่จะละเว้นไว้รูปสระอา ไว้ก่อน เพื่อประกอบรูปสระ เอา ก่อนแล้วจะพิจารณารูปสระ อา ในกฎสุดท้าย โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. ADD 1 space before input #Prevent Index Out of Bound
4. ADD 1 space after input #Prevent Index Out of Bound
5. FOR each character of the input
6. current character = character
7. previous character = input[current index -1]
8. next character = input[current index +1]
9. main condition = [
10. Previous character NOT IN **TV** And
11. current character IN **TIF**
12. ]
13. condition 1 = [
14. current character NOT IN **{ เ ,า}**
15. ]
16. condition 2 = [
17. current character ≠ ‘**เ**’
18. next character ≠ ‘**า**’
19. ]
20. condition 3 = [
21. current character ≠ ‘**า**’
22. previous character ≠ ‘**เ**’
23. ]
24. IF main condition AND condition 1 THEN
25. character = **F-RSA**(**TIF**, **TIFR**, input)
26. ELSE IF main condition AND condition 2 THEN
27. character = **F-RSA**(**TIF**, **TIFR**, input)
28. ELSE IF main condition AND condition 3 THEN
29. character = **F-RSA**(**TIF**, **TIFR**, input)
30. ENDIF
31. output = output Concatenated character
32. ENDFOR
33. RETURN output
34. END

**กฎที่ 9. ThaiVowelPrefix**

ทำการสลับที่สระอักษรไทยที่เป็นสระโดยการเลื่อนจากหลังพยัญชนะ ไปไว้หน้าพยัญชนะหากพยัญชนะตัวหน้าเป็นรูปอักษร ร ให้ตรวจสอบดูว่าพยัญชนะที่ถัดจาก อักษร ร นั้นเป็นพยัญชนะหรือสระหากเป็นพยัญชนะ ให้เลื่อนสลักอักษรไทยที่เป็นสระหน้าไปไว้หน้าพยัญชนะตัวนั้น โดยมีขั้นตอนวิธีดังนี้

1. # bindu เครืองหมายพินทุ( ฺ )
2. START
3. READ input
4. ADD 4 space before input #Prevent Index Out of Bound
5. FOR each character of the input
6. current character = character
7. IF current character IN {'เ','โ', 'ไ'} THEN
8. previous character 1 = input[current index -1]
9. previous character 2= input[current index -2]
10. previous character 3= input[current index -3]
11. previous character 4 = input[current index-4 ]
12. conjunct r condition = [
13. previous character 1 = "ร" AND
14. previous character 2 = bindu AND
15. previous character 3 IN **TC** AND
16. previous character 4 ≠ "เ" AND
17. previous character 4 ≠ "โ" AND
18. previous character 4 ≠ "ไ"
19. ]
20. normal condition = [
21. previous character 1 IN **TC**
22. ]
23. IF conjunct r condition THEN
24. SWAP value in current index TO current index -1
25. SWAP value in current index TO current index -2
26. SWAP value in current index TO current index -3
27. ELSE IF normal condition THEN
28. SWAP value in current index TO current index -1
29. ENDIF
30. output = output Merge character
31. ENDFOR
32. RETURN output
33. END

**กฎที่ 10. ARemove**

ทำการลบสระอักษรโรมัน A ออกจากข้อความทั้งหมด

1. START
2. READ input
3. output = **F-RS**(‘A’ , empty value, input)
4. RETURN output
5. END

**กฎที่ 11. SwapAnusvaraAndChandrabindu**

เปลี่ยนเครื่องหมายอนุสาวระอักษรไทยและ Moon Dot ไปข้างหน้ากรณีอยู่หลัง ลากข้าง(า)

1. #anusvara เครืองหมายนิคหิต(ํ)
2. #chandrabindu เครืองหมาย Moon dot( ँ )
3. START
4. READ input
5. output = **F-RS(‘าํ’ , ‘ํา’ , input)**
6. output = **F-RS(‘าँ’ , ‘ँา’, input)**
7. RETURN output
8. END

**กฎที่ 12. Chandrabindu**

แปลงเครื่องหมาย Moon Dot โดยใช้ไม้หันอากาศและนิคหิตแทน

1. START
2. READ input
3. output = **F-RS**(**‘ँ’** , **‘ัํ’**, input)
4. RETURN output
5. END

**กฎที่ 13. ThaiAAInFist**

เปลี่ยนลากข้าง(า) โดยผสมรูปพยัญชนะ อ กรณีอยู่ถัดรูปสระอื่น ซึ่งเป็นสระตัวสุดท้ายที่จะแปลง เนื่องจากต้องแปลงรูปสระเอา ก่อน

1. START
2. READ input
3. FOR each character of the input
4. current character = character
5. previous character = input[current index -1]
6. IF previous character NOT IN **TV** AND character = "า" THEN
7. character = "อา"
8. ENDIF
9. output = output Concatenated character
10. ENDFOR
11. RETURN output
    * 1. **ขั้นตอนวิธีของกฎการปริวรรตอักษรไทยแบบปรับรูป**

**กฎที่ 1. AvagarahaRemove**

โดยมีขั้นตอนวิธีดังนี้

1. #avagaraha เครื่องหมาย ' (apostrophe)
2. START
3. READ input
4. output = **F-RSA**(avagaraha , empty value, input)
5. RETURN output
6. END

**กฎที่ 2. RomanChandrabinduToSingle**

เปลี่ยนเครื่องหมายจันทรพินทุ จากการเขียนด้วยตัวอักษร m และเครื่องหมาย Dot Moonที่อยู่ด้านบน(m̐)ให้เป็นเครื่องหมายจันทรพินแบบตัวอักษรเดี่ยวคือ ṁ ออกแบบไว้เพื่อรองรับไว้สำหรับการใช้เครื่องหมายจันทรพินทุโรมันทั้งสองแบบ และเตรียมไว้สำหรับการแปลงต่อไปโดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RS**(‘m̐’ , ‘ṁ', input)
4. RETURN output
5. END

**กฎที่ 3. RomanNumberAndPunctuate**

เปลี่ยนตัวเลขและเครื่องหมายวรรคตอนอักษรโรมันให้เป็นตัวเลขและเครื่องหมายวรรคตอนอักษรไทย โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RNP** , **TNP**, input)
4. RETURN output
5. END

**กฎที่ 4. RomanMixConsonant**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RMC** , **TMC**, input)
4. RETURN output
5. END

**กฎที่ 5. RomanMixVowel**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RMV** , **TMV**, input)
4. RETURN output
5. END

**กฎที่ 6. RomanSingleConsonant**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RSC** , **TSC2**, input)
4. RETURN output
5. END

**กฎที่ 7. RomanSingleVowel**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RSV** , **TSV2**, input)
4. RETURN output
5. END

**กฎที่ 8. AnusvaraAndChandrabindu**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RSC** , **TSC**, input)
4. RETURN output
5. END

**กฎที่ 9. ThaiVowelInFist**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. ADD 1 space before input #Prevent Index Out of Bound
4. ADD 1 space after input #Prevent Index Out of Bound
5. FOR each character of the input
6. current character = character
7. previous character = input[current index -1]
8. next character = input[current index +1]
9. main condition = [
10. Previous character NOT IN **TV** And
11. current character IN **TIF2**
12. ]
13. condition 1 = [
14. current character NOT IN **{ เ ,า}**
15. ]
16. condition 2 = [
17. current character ≠ ‘**เ**’
18. next character ≠ ‘**า**’
19. ]
20. condition 3 = [
21. current character ≠ ‘**า**’
22. previous character ≠ ‘**เ**’
23. ]
24. IF main condition AND condition 1 THEN
25. character = **F-RSA**(**TIF2**, **TIFR2**, input)
26. ELSE IF main condition AND condition 2 THEN
27. character = **F-RSA**(**TIF2**, **TIFR2**, input)
28. ELSE IF main condition AND condition 3 THEN
29. character = **F-RSA**(**TIF2**, **TIFR2**, input)
30. ENDIF
31. output = output Concatenated character
32. ENDFOR
33. RETURN output
34. END

**กฎที่ 10. ThaiVisarga**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RSA**(**RSC** , **TSC**, input)
4. RETURN output
5. END

**กฎที่ 11. ThaiVowelPrefix**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. ADD 3 space before input #Prevent Index Out of Bound
4. FOR each character of the input
5. current character = character
6. IF current character IN {'เ','โ', 'ไ'} THEN
7. previous character 1 = input[current index -1]
8. previous character 2= input[current index -2]
9. previous character 3= input[current index -3]
10. conjunct r condition = [
11. previous character 1 = "ร" AND
12. previous character 2 IN **TC** AND
13. previous character 3 ≠ "เ" AND
14. previous character 3 ≠ "โ" AND
15. previous character 3 ≠ "ไ"
16. ]
17. normal condition = [
18. previous character 1 IN **TC**
19. ]
20. IF conjunct r condition THEN
21. SWAP value in current index TO current index -1
22. SWAP value in current index TO current index -2
23. ELSE IF normal condition THEN
24. SWAP value in current index TO current index -1
25. ENDIF
26. output = output Merge character
27. ENDFOR
28. RETURN output
29. END

**กฎที่ 12. ThaiAAInFist**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. FOR each character of the input
4. current character = character
5. previous character = input[current index -1]
6. IF previous character NOT IN **TV** AND character = "า" THEN
7. character = "อา"
8. ENDIF
9. output = output Concatenated character
10. ENDFOR
11. RETURN output

**กฎที่ 13. convertAE**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RS**(**‘**ัเ**’** , ‘ะเ’, input)
4. RETURN output
5. END

**กฎที่ 14. convertAO**

โดยมีขั้นตอนวิธีดังนี้

1. START
2. READ input
3. output = **F-RS**(‘ัโ’,, ‘ะโ’,, input)
4. RETURN output
5. END
   * 1. **ขั้นตอนวิธี (Algorithm) ในการปริวรรตอักษร**

เนื่องจากการทำงานปริวรรตนั้นต้องทำกับข้อความจำนวนมากและโปรแกรมทำงานแบบ Server-Client หากการทำงานใช้เวลามากเกินไปจะเกิดปัญหา Time Out ใน Web Browser ได้ เพื่อให้การปริวรรตได้ราบรื่น และให้ได้ลักษณะการส่งข้อความไปประมวลผลให้เหมาะกับกฎที่ออกแบบไว้ ที่มี Loop อยู่เป็นจำนวนมาก ผู้จัดทำจึงได้ออกแบบ ขั้นตอนวิธีในการปริวรรตอักษรไว้ 3 วิธี

วิธีที่ 1. รับตัวอักษรต้นฉบับอักษรโรมันทำการปริวรรตทั้งข้อความ

1. START
2. READ roman text
3. thai text = TRANSLITERATION (roman text)
4. RETURN thai text
5. END

วิธีที่ 2. รับตัวอักษรต้นฉบับอักษรโรมันทำการปริวรรตที่ละบรรทัด

1. START
2. READ roman text
3. FOR each line of the roman text
4. thai text = thai text + TRANSLITERATION (line)
5. ENDFOR
6. RETURN thai text
7. END

วิธีที่ 3. รับตัวอักษรต้นฉบับอักษรโรมันทำการปริวรรตที่ละคำในแต่ละบรรทัด

1. START
2. READ roman text
3. FOR each line of the roman text
4. FOR each word of the line
5. thai text = thai text + TRANSLITERATION (word)
6. ENDFOR
7. ENDFOR
8. RETURN thai text
9. END
   1. **การพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)**

ผู้จัดทำได้ทำการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype) เพื่อตรวจสอบความเป็นไปได้และการพิสูจน์แนวคิดของการพัฒนาส่วนตัวปริวรรตอักษร(POC: Proof-of-Concept) ว่าระบบจะสามารถทำงานได้อย่างถูกต้องและตรงความต้องการหรือไม่ โดยผู้จัดทำได้ทำการพัฒนาโดยใช้ภาษา JAVA ทำการโปรแกรมให้ผลลัพธ์ออกมาในรูปแบบ Console Command โดยเลือกทำเฉพาะส่วนที่ประเมินว่ามีความยากในการพัฒนามากที่สุดคือ การปริวรรตอักษรแบบปรับรูป หรือ แบบทั่วไป โดยทดสอบใช้กฎ 7 กฎ อีกทั้งการทำตัวต้นแบบนี้ยังเพื่อพิสูจน์ขั้นตอนวิธี (Algorithm) ในการปริวรรตอักษร 3 วิธีที่ออกแบบไว้ก่อนหน้าว่า ขั้นตอนวิธีใดมีประสิทธิภาพมากที่สุด เพื่อเลือกเอาไปพัฒนาต่อไป

รูปที่ 2.5 ผลลัพธ์หลังจากการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)



รูปที่ 2.5 ผลลัพธ์หลังจากการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)

หลังจากพิสูจน์ขั้นตอนวิธี (Algorithm) ในการปริวรรตอักษร 3 วิธีที่ออกแบบไว้ โดยทดสอบกับตัวอักษรโรมันต้นฉบับ จำนวน 34,769 ตัวอักษร ขั้นตอนวิธีมีประสิทธิภาพมากที่สุด คือ วิธีที่ 2. การปริวรรตที่ละบรรทัด มีเวลาที่ใช้โดยเฉลี่ยจากการ execute 5 ครั้งที่น้อยที่สุด ผลลัพธ์ตามตารางที่

ตารางที่ 2.2 การเปรียบเทียบรูปพยัญชนะวรรค

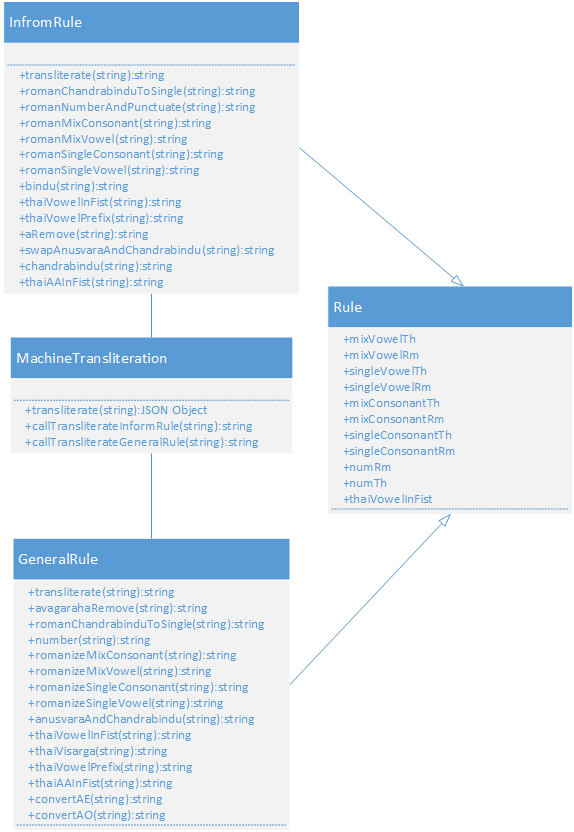
|  |  |
| --- | --- |
| **ขั้นตอนวิธีในการปริวรรตอักษร** | **เวลาที่ใช้โดยเฉลี่ย(วินาที)** |
| วิธีที่ 1. ทำการปริวรรตทั้งข้อความ | 9.86 |
| วิธีที่ 2. ทำการปริวรรตที่ละบรรทัด | 0.61 |
| วิธีที่ 3. ทำการปริวรรตที่ละคำในแต่ละบรรทัด | 0.76 |

* 1. **การพัฒนาส่วนตัวปริวรรตอักษร(Transliteration Machine)**

ผู้จัดทำได้ทำการพัฒนาส่วนตัวปริวรรตอักษร(Transliteration Machine) ให้สามารถทำงานได้ตามกฎ (Rule-based) และขั้นตอนวิธี (Algorithm) ที่ได้ออกแบบไว้ อีกทั้งทำการทดสอบ ตรวจสอบผลลัพธ์ ภายใต้คำแนะนำของผู้เชี่ยวชาญภาษาสันสกฤตโดยการพัฒนานั้นพัฒนาโดยใช้ภาษา PHP พัฒนาเป็นแบบ Object Oriented Programming ประกอบไปด้วยคลาส 3 Class ได้แก่ MachineTransliteration Class InfromRule Class และ GeneralRule Class

MachineTransliteration Class จะทำหน้าที่เป็น API สำหรับเชื่อมต่อกับระบบอื่นภายนอก

InfromRule Class จะทำหน้าที่ปริวรรตข้อความให้เป็นอักษรไทยแบบคงรูปหรือแบบแผน และGeneralRule Class จะทำหน้าที่ การปริวรรตข้อความให้เป็นอักษรไทยแบบปรับรูปหรือแบบทั่วไป ส่วนการทำงานนั้นหากมีการส่งข้อความเข้ามา MachineTransliteration Class จะรับเอาข้อความเข้ามา แล้วนำไปประมวลผลโดยจะแบ่งข้อความแบบทีละบรรทัด ตามที่ได้ออกแบบไว้ และส่งบรรทัดข้อความนั้น ๆ ไปให้ทั้งสอง Class โดยเรียกใช้ฟังก์ชัน transliterate() จากคลาส InfromRule และ GeneralRule ทำการประมวลผล ซึ่งจะได้ผลลัพธ์ทั้งแบบคงรูปและแบบปรับรูปเมื่อเสร็จจะส่งผลลัพธ์กลับไปยังที่ MachineTransliteration Class ทำการสะสมค่าผลลัพธ์เอาไว้ เมื่อเสร็จสิ้นกระบวนการทุกบรรทัดในข้อความแล้ว MachineTransliteration Class จะนำผลลัพธ์ที่ได้มาประมวลผลแล้วส่งออกเป็น JSON Object ให้กับระบบอื่นภายนอก ดังมีโครงสร้างใน class diagram ดังต่อไปนี้



* 1. **การพัฒนาส่วนต่อผู้ใช้งาน(User Interface)**

ผู้จัดทำได้ทำการพัฒนาส่วนต่อผู้ใช้งาน(User Interface) ที่ไว้สำหรับสำหรับในการป้อนอักษรต้นฉบับจากผู้ใช้ เพื่อส่งไปยังส่วนตัวปริวรรตอักษร(Transliteration Machine) เพื่อทำประมวลให้ได้ผลลัพธ์ออกมาแสดงและส่วนการแสดงการเปรียบเทียบตรวจสอบจากอักษรต้นฉบับกับอักษรผลลัพธ์จากการปริวรรต การพัฒนาส่วนต่อผู้ใช้งาน(User Interface) โดยมีหน้าจอไว้สำหรับรับข้อมูลจากผู้ใช้ โดยลักษณะหน้าจอจะมี select box สำหรับเลือกตัวอักษรต้นฉบับ select box ตัวที่ 2 ใช้สำหรับเลือกอักษรชนิดอื่น ๆ ในการเปรียบเทียบหากผู้ใช้เลือกอักษรต้นฉบับและอักษรเปรียบเทียบตัวที่ 2 แล้ว จะปรากฏตารางเปรียบเทียบตัวอักษรที่ด้านล่างสุด ถัดลงมาด้านล่างจะเป็น text area ช่องแรกไว้สำหรับข้อความต้นฉบับช่องที่ 2 เป็นผลลัพธ์จากการปริวรรตอักษรอื่น ๆ ด้านล่างจะมีปุ่มไว้สำหรับทำการปริวรรตอักษรเป็นอักษรไทย

หากผู้ใช้กรอกตัวอักษรต้นฉบับและกดปุ่ม การปริวรรตเป็นอักษรไทยแล้ว จะมีผลลัพธ์ผลลัพธ์การปริวรรตเป็นอักษรไทยแบบคงรูป และแบบปรับรูป ออกแสดงที่ตารางเปรียบเทียบผลการปริวรรตโดยสามารถเลือกการเปรียบเทียบได้มากที่สุดถึง 4 ชนิดตัวอักษรและยังสามารถซ่อนหรือแสดงตารางเปรียบเทียบตัวใดตัวหนึ่งได้ อีกทั้งยังสามารถทำการเปรียบเทียบโดยการไฮไลท์ที่คำในแต่ละบรรทัดระบบจะทำการไฮไลท์ข้อความจากทุกๆตัวอักษร



รูปที่ 2.5 ผลลัพธ์หลังจากการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)



รูปที่ 2.5 ผลลัพธ์หลังจากการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype)

* 1. **การทดสอบและประเมินผลการใช้งานจากผู้เชี่ยวชาญภาษาสันสกฤต**

เพื่อให้ผลการทดสอบสามารถประเมินผลการดำเนินโครงการ ให้ความถูกต้องตามวัตถุประสงค์ของการศึกษาโครงการเฉพาะเรื่อง จึงมีการกำหนดขอบเขตและออกแบบวิธีการทดสอบและประเมินผล ดังต่อไปนี้

* + 1. **การทดสอบการตัวปริวรรตอักษร**

การทดสอบการตัวปริวรรตอักษรเป็นขั้นตอนการทดสอบความถูกต้องของส่วนตัวปริวรรตอักษร ประกอบด้วย ชุดข้อมูลของการทดสอบ และวิธีการทดสอบ มีรายละเอียดดังต่อไปนี้

1. **ชุดข้อมูลของการทดสอบ**
   1. กรณีที่ใช้ในการทดสอบ(Test Case) โดยเป็นชุดข้อความผลลัพธ์ที่ถูกต้องและสามารถครอบคลุมกฎทั้งกรณีการปริวรรตแบบคงรูปและปรับรูป จำนวนอย่างละ 59 กรณี รวมเป็น 118 กรณี
   2. ตัวอย่างตัวอย่างการปริวรรต ผู้จัดทำได้เลือกหนึ่งในตัวอย่างการปริวรรตที่กล่าวมาก่อนแล้ว โดยเลือกที่มีเนื้อหาไม่มากไม่น้อยจนเกินไป และมีต้นฉบับที่เก็บอยู่ในรูปแบบ TEXT โดยตัวอย่างการปริวรรต ผู้จัดทำได้เลือกคือ
      1. ตัวอย่างการปริวรรตแบบคงรูป ใช้ คัมภีร์วัชรเฉทิกปรัชญาปารมิตาสูตร ฉบับอักษรไทย ขนาด 31,436 ตัวอักษร
      2. ตัวอย่างการปริวรรตแบบปรับรูป ใช้ คัมภีร์วัชรเฉทิกปรัชญาปารมิตาสูตร ฉบับคำอ่านอักษรไทย ขนาด 34,117 ตัวอักษร
   3. ตัวอย่างต้นฉบับ คัมภีร์วัชรเฉทิกปรัชญาปารมิตา อักษรโรมัน จากโครงการ DSBC (www.dsbcproject.org )
2. **วิธีการทดสอบความถูกต้อง**

วิธีการทดสอบความถูกต้องของผลลัพธ์ที่ได้มาจาก ส่วนตัวปริวรรตอักษร มี 2 วิธีดังนี้

* 1. **การทดสอบแบบแยกส่วน (Unit Testing)**

การทดสอบแบบแยกส่วนนั้นจะใช้ PHPUnit (PHP Testing Framework) เป็นเครื่องมือในการตรวจสอบและทดสอบกฎ โดยใช้ชุดคำเป็นกรณีที่ใช้ทดสอบ(Test Case) 118 กรณี เพื่อทดสอบการทำงานของส่วนการปริวรรตว่า ผลลัพธ์ที่ออกมาตรงกับชุดคำดังกล่าวหรือไม่

* 1. **การทดสอบโดยการเปรียบเทียบกับตัวอย่างการปริวรรต**

การทดสอบโดยการเปรียบเทียบกับตัวอย่างการปริวรรต มีขั้นตอนดังต่อไปนี้

* + 1. ใช้ตัวอย่างต้นฉบับ คัมภีร์วัชรเฉทิกปรัชญาปารมิตา อักษรโรมัน ทำการปริวรรตเป็นอักษรไทยแบบคงรูปและปรับรูป
    2. นำผลลัพธ์การปริวรรตอักษรไทยแบบคงรูป ไปเปรียบเทียบกับ คัมภีร์วัชรเฉทิกปรัชญาปารมิตาสูตร ฉบับอักษรไทย และ ผลลัพธ์การปริวรรตอักษรไทยแบบปรับรูป ไปเปรียบเทียบกับ คัมภีร์วัชรเฉทิกปรัชญาปารมิตาสูตร ฉบับคำอ่านอักษรไทย โดยใช้โปรแกรม Pretty Diff v2.0.0 เป็นโปรแกรมเปรียบเทียบเพื่อคำนวณหาอัตราความถูกต้องของส่วนตัวปริวรรตอักษร
    3. **ประเมินผลการใช้งานจากผู้เชี่ยวชาญภาษาสันสกฤต**

การประเมินผลการใช้งานประเมินผลโดยผู้เชี่ยวชาญด้านภาษาสันสกฤตโดยมี ข้อคิดเห็นจากผู้ประเมินพร้อมหัวข้อแบบประเมิน แบ่งหัวข้อได้ดังต่อไปนี้

1. **แบบประเมินผลการใช้งานระบบ ด้านการทำงานได้ตามฟังก์ชันงานของระบบ (Function)**

โดยจะประเมิน ความถูกต้องของการผลลัพธ์ ความรวดเร็วในการประมวลผล ความครอบคลุมของงาน ความสามารถป้องกันข้อผิดพลาดที่เกิดขึ้น และความน่าเชื่อถือได้ของระบบ โดยมีงานที่จะประเมินดังนี้

* 1. การปริวรรตจากอักษรไทยแบบคงรูปเป็นอักษรโรมัน
  2. การปริวรรตจากอักษรโรมันเป็นอักษรไทยแบบคงรูป
  3. การปริวรรตจากอักษรโรมันเป็นอักษรไทยแบบปรับรูป
  4. ตารางการเปรียบเทียบผลการปริวรรต
  5. ตารางการเปรียบเทียบอักษร

1. **แบบประเมินผลการใช้งานระบบ ด้านความง่ายต่อการใช้งานระบบ (Usability)**

โดยจะประเมินจากส่วนต่อผู้ใช้งาน(User Interface)แบ่งหัวข้อได้ดังต่อไปนี้

1. ความง่ายต่อการใช้งานของระบบ
2. ความเหมาะสมในการเลือกใช้ชนิดตัวอักษรบนจอภาพ
3. ความเหมาะสมในการใช้สีของตัวอักษรและรูปภาพ
4. ความเหมาะสมในการใช้ข้อความเพื่ออธิบายสื่อความหมาย
5. ความเหมาะสมในการใช้สัญลักษณ์หรือรูปภาพในการสื่อความหมาย
6. ความเป็นมาตรฐานเดียวกันในการออกแบบหน้าจอภาพ
7. ความเหมาะสมในการปฏิสัมพันธ์โต้ตอบกับผู้ใช้
8. ความเหมาะสมในการวางตำแหน่งของส่วนประกอบบนจอภาพ
9. คำศัพท์ที่ประกอบใช้ผู้ใช้มีความคุ้นเคยและสามารถปฏิบัติตามได้โดยง่าย