หัวข้อการศึกษาโครงการเฉพาะเรื่อง การปริวรรตอักษรด้วยเครื่องสำหรับภาษาสันสกฤตแบบอักษรโรมันเป็นอักษรไทย

หน่วยกิต 6

ผู้เขียน นายธนกฤต พรหมศิริ

อาจารย์ที่ปรึกษา ดร.โอฬาร โรจนพรพันธ์

หลักสูตร วิทยาศาสตรมหาบัณฑิต

สาขาวิชา เทคโนโลยีสารสนเทศ

คณะ คณะเทคโนโลยีสารสนเทศ

ปีการศึกษา 2558

บทคัดย่อ

ในงานค้นคว้าทางวิชาการด้านภาษาสันสกฤตในประเทศไทยนั้น การปริวรรตอักษรเอกสารภาษาสันสกฤตถือว่าเป็นงานที่ใช้ความวิริยะอุตสาหะในการปริวรรตอักษรและพิสูจน์อักษรเป็นอย่างยิ่งเนื่องจากเอกสารต่างๆนั้นมีเนื้อหาจำนวนมาก เพื่อหาทางแก้ไขปัญหา การศึกษาโครงการเฉพาะเรื่องนี้จึงได้เสนอการพัฒนากฎ ขั้นตอนวิธี การปริวรรตอักษรด้วยเครื่องสำหรับภาษาสันสกฤตแบบอักษรโรมันมาตรฐาน IAST เป็นอักษรไทย ใน 2 รูปแบบ คือ อักษรไทยในรูปแบบแบบแผน และรูปแบบทั่วไป จากผลการศึกษาพบว่า ผลการปริวรรตรูปแบบแบบแผน มีความถูกต้อง 98.16 % จากการเปรียบเทียบกับต้นฉบับจำนวน 27,440 ตัวอักษรและผลการปริวรรตรูปแบบแบบทั่วไป มีความถูกต้อง 95.82 % จากการเปรียบเทียบกับตันฉบับจำนวน 30,049 ตัวอักษร และการประเมินผลจากผู้เชี่ยวชาญด้านภาษาสันสกฤต ให้ความเห็นว่า ระบบนี้สามารถทำงานปริวรรตอักษรภาษาสันสกฤตได้เป็นอย่างดี ช่วยให้ประหยัดเวลาในการปริวรรตได้อย่างมาก

คำสำคัญ : การปริวรรตอักษร/ การถอดอักษร/ภาษาสันสกฤต

Special Project Study Title Romanized Sanskrit – Thai Machine Transliteration

Special Project Study Credits 6

Candidate Mr. Thanakrit Promsiri

Special Project Study Advisor Dr. Olarn Rojanapornpun

Program Master of Science

Field of Study Information Technology

Faculty School of Information Technology

Academic Year 2015

Abstract

Sanskrit study and research in Thailand. Transliteration from Sanskrit documents and proofreading is a difficult and time-consuming task. Due to Sanskrit documents have a lot of content. To find solutions. This paper presents Romanized Sanskrit – Thai Machine Transliteration. For the development rules and algorithm of transliteration Romanized Sanskrit based on the IAST standard to Thai alphabet. By there are two patterns. Transliteration patterns are composed of stable pattern and common pattern. The experimental results show that the proposed stable pattern transliteration can perform accurately up to 98.16 % from 21,399 samples. The common pattern can perform accurately up to 95.82 % from 30,049 samples. The evaluation of an expert in Sanskrit. Comments this system can work transliterated Sanskrit as well. Saves a lot of time in transliteration.

Keywords: Transliteration / Transliteration Machine/ Sanskrit

**กิตติกรรมประกาศ**

การจัดทำโครงการศึกษาเฉพาะเรื่องในครั้งนี้สำเร็จไปด้วยดีด้วยความอนุเคราะห์จากบุคคลหลายท่านที่คอยให้คำปรึกษา คำชี้แนะต่าง ๆ ที่เป็นประโยชน์ต่อโครงการศึกษาในครั้งนี้ ขอกราบขอบพระคุณ ดร.โอฬาร โรจนพรพันธุ์ อาจารย์ที่ปรึกษาโครงการ ที่คอยให้ความช่วยเหลือในด้านแนวคิดและคำแนะนำที่เป็นประโยชน์ต่อการจัดทำโครงการศึกษาเฉพาะเรื่องในครั้งนี้ ขอกราบขอบพระคุณ คณาจารย์ บุคลากรในคณะเทคโนโลยีสารสนเทศ ที่คอยให้แนวคิด ตรวจแก้ไขโครงการนี้ นอกจากนี้ ขอขอบคุณท่านอาจารย์ธวัชชัย ดุลยสุจริต ผู้ประเมินผลการทำงาน ที่กรุณาให้ความช่วยเหลือให้คำปรึกษาด้านภาษาสันสกฤตและตรวจสอบระบบ ให้โครงการนี้สำเร็จลุล่วงไปตามวัตถุประสงค์ และขอนอบน้อมกราบขอบพระคุณพระพุทธพจน์บทต่อไปนี้ ที่ทำให้ข้าพเจ้ามีกำลังใจในการทำการศึกษาจนสำเร็จสมบูรณ์

“*วายเมเถว ปุริโส ยาว อตฺถสฺส นิปฺปทา*”

บุรุษควรพยายามจนกว่าจะประสบความสำเร็จ

ท้ายสุดนี้ ความสำเร็จและประโยชน์อันใดที่เกิดขึ้นจากการศึกษาในครั้งนี้ ข้าพเจ้าขอมอบให้กับมารดาบิดาและบรรดาบุพการีทุกท่านที่ได้อุปการะเลี้ยงดูข้าพเจ้ามา ครูบาอาจารย์ทุกท่านที่ได้อบรมพร่ำสอนศิลปวิทยาการทั้งในอดีตและปัจจุบัน และทุก ๆ ท่านที่ให้การให้ความช่วยเหลือและเป็นกำลังใจในการจัดทำโครงการศึกษาเฉพาะเรื่อง จนสำเร็จสมบูรณ์ ข้าพเจ้าจึงใคร่ขอขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

นายธนกฤต พรหมศิริ

**สารบัญ**

|  |  |
| --- | --- |
|  | **หน้า** |
| บทคัดย่อภาษาไทย | ข |
| บทคัดย่อภาษาอังกฤษ | ค |
| กิตติกรรมประกาศ | ง |
| สารบัญ | จ |
| รายการตาราง | ซ |
| รายการรูปประกอบ | ญ |
|  |  |
| **บทที่** |  |
| 1. **บทนำ** | **1** |
| * 1. ความสำคัญและที่มาของงาน | 1 |
| * 1. วัตถุประสงค์ของการศึกษา | 4 |
| * 1. ประโยชน์ที่คาดว่าจะได้รับ | 4 |
| * 1. ขอบเขตของงาน | 4 |
| * 1. กระบวนการศึกษา | 5 |
|  |  |
| 1. **ทฤษฎีและงานวิจัยที่เกี่ยวข้อง** | **7** |
| * 1. ทฤษฎีที่เกี่ยวข้อง | 7 |
| * + 1. การปริวรรตอักษรหรือการเขียนคำทับศัพท์ (transliteration) | 7 |
| * + 1. ภาษาสันสกฤต | 8 |
| * + 1. การปริวรรตอักษรภาษาสันสกฤตสากล (IAST) | 11 |
| * 1. ทฤษฎีการปริวรรตอักษรด้วยเครื่อง | 32 |
| * + 1. การปริวรรตอักษรด้วยเครื่อง | 32 |
| * + 1. ประเภทของโมเดลปริวรรตอักษรด้วยเครื่อง | 33 |
| * 1. งานวิจัยที่เกี่ยวข้อง | 34 |
|  |  |
| 1. **ขั้นตอนการดำเนินโครงการ** | **36** |
| * 1. 3.1.การศึกษาหลักการทฤษฎีและงานวิจัยรวมถึงเทคโนโลยีที่เกี่ยวข้อง | 36 |
| * + 1. การศึกษาด้านทฤษฎีและงานวิจัยที่เกี่ยวข้อง | 36 |
|  | **หน้า** |
|  |  |
| * + 1. การศึกษาด้านเทคโนโลยีที่เกี่ยวข้อง | 38 |
| * 1. การจัดเตรียมแหล่งข้อมูลไว้สำหรับการศึกษาและการทดสอบระบบ | 38 |
| * + 1. แหล่งข้อมูลตัวอย่างสำเนาเอกสารต้นฉบับภาษาสันสกฤต | 38 |
| * + 1. ตัวอย่างการปริวรรตอักษรภาษาสันสกฤตเป็นไทย | 39 |
| * 1. การวิเคราะห์ข้อมูล กำหนดภาพรวมของงาน เลือกรูปแบบการพัฒนา | 41 |
| * + 1. สรุปวิธีการปริวรรตอักษรที่จะนำมาพัฒนา | 41 |
| * + 1. สรุปแนวทางและโมเดลที่จะนำมาพัฒนา | 43 |
| * + 1. สรุปเทคโนโลยีที่ช่วยพัฒนาส่วนเสริมการปริวรรตและการทดสอบ | 43 |
| * + 1. สรุปภาพรวมของงาน | 44 |
| * 1. การออกแบบและกำหนด กฎ (Rule-based) ขั้นตอนวิธี (Algorithm) | 46 |
| * + 1. สัญลักษณ์ประกอบขั้นตอนวิธีของกฎการปริวรรต | 46 |
| * + 1. กำหนดกฎการปริวรรตอักษรไทยแบบคงรูป | 48 |
| * + 1. กำหนดกฎการปริวรรตอักษรไทยแบบปรับรูป | 49 |
| * + 1. ขั้นตอนวิธีของกฎการปริวรรตอักษรไทยที่ใช้ร่วมกัน(กฎกลาง) | 50 |
| * + 1. ขั้นตอนวิธีของกฎการปริวรรตอักษรไทยแบบคงรูป | 54 |
| * + 1. ขั้นตอนวิธีของกฎการปริวรรตอักษรไทยแบบปรับรูป | 57 |
| * + 1. ขั้นตอนวิธี (Algorithm) ในการปริวรรตอักษร | 61 |
| * 1. การพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype) | 62 |
| * 1. การพัฒนาส่วนตัวปริวรรตอักษร(Transliteration Machine) | 64 |
| * 1. การพัฒนาส่วนต่อผู้ใช้งาน(User Interface) | 65 |
| * 1. การทดสอบและประเมินผลการใช้งานจากผู้เชี่ยวชาญภาษาสันสกฤต | 67 |
|  |  |
| 1. **ผลการดำเนินงาน** | **71** |
| * 1. 7.1. ผลการดำเนินงาน | 71 |
| * 1. 7.2.ผลการทดสอบการตัวปริวรรตอักษร | 74 |
| * + 1. ผลการทดสอบแบบแยกส่วน (Unit Testing) | 74 |
| * + 1. ผลการทดสอบโดยการเปรียบเทียบกับตัวอย่างการปริวรรต | 75 |
| * + 1. ผลการทดสอบประสิทธิภาพของระบบ | 77 |
| * 1. ผลการประเมินผลการใช้งานจากผู้เชี่ยวชาญภาษาสันสกฤต | 77 |
| * + 1. ผลการประเมิน ด้านการทำงานได้ตามฟังก์ชันงานของระบบ (Function) | 78 |
|  | **หน้า** |
|  |  |
| * + 1. แบบประเมินผล ด้านความง่ายต่อการใช้งานระบบ (Usability) | 80 |
|  |  |
| 1. **สรุปผลและข้อเสนอแนะ** | **81** |
| * 1. สรุปผลการทดลอง | 82 |
| * 1. ข้อเสนอแนะ | 82 |
|  |  |
| **เอกสารอ้างอิง** | **83** |
| **ภาคผนวก** | **86** |
| 1. หนังสือขอความอนุเคราะห์ในการประเมินผลการใช้งาน   แบบประเมินผลการใช้งานจากผู้เชี่ยวชาญด้านภาษาสันสกฤต | 86 |
| 1. ประวัติผู้ประเมินผลและผู้เชี่ยวชาญภาษาสันสกฤต | 92 |
| 1. ตัวอย่างผลการปริวรรตจากโปรแกรม | 95 |
| 1. ผลการทดสอบ Unit Testing โดยใช้ PHPUnit | 96 |
|  |  |
| **ประวัติผู้วิจัย** | **103** |

**รายการตาราง**

|  |  |  |
| --- | --- | --- |
| **ตาราง** |  | **หน้า** |
| 2.1 | ตัวชี้วัดสำหรับการวัดทีมพัฒนาซอฟต์แวร์ | 9 |
| 2.1 | การเปรียบเทียบรูปสระ | 15 |
| 2.2 | การเปรียบเทียบรูปพยัญชนะวรรค | 16 |
| 2.3 | การเปรียบเทียบรูปพยัญชนะอวรรค | 16 |
| 2.4 | การเปรียบเทียบรูปเครื่องหมายและสัญลักษณ์พิเศษ | 17 |
| 2.5 | การเปรียบเทียบรูปตัวเลข | 18 |
| 2.6 | ตัวอย่างการถอดรูปสระa | 19 |
| 2.7 | ตัวอย่างการถอดรูปสระeaio au ที่ประสมอยู่พยัญชนะต้นอักษรโรมัน | 20 |
| 2.8 | ตัวอย่างการถอดรูปสระอักษรโรมันขึ้นต้นคำ | 20 |
| 2.9 | ตัวอย่างการถอดรูปพยัญชนะอักษรโรมันที่ไม่มีสระอักษรโรมันกำกับ | 20 |
| 2.1 | ตัวอย่างการถอดรูปพยัญชนะอักษรโรมันที่ประสมด้วยสระ ṛ ṝ ḷ ḹ | 21 |
| 2.11 | ตัวอย่างการถอดรูปเครื่องหมายอนุสวาระ | 21 |
| 2.12 | ตัวอย่างการถอดรูปรูปเครื่องหมายจันทรพินทุหรืออนุนาสิกะ | 21 |
| 2.13 | ตัวอย่างการถอดรูปเครื่องหมายวิสรรคะ | 22 |
| 2.14 | ตัวอย่างการถอดรูปชุดพยัญชนะอักษรโรมันที่เป็นพยัญชนะสังโยค | 23 |
| 2.15 | ตัวอย่างการถอดรูปพยัญชนะสังโยคที่ซ้อนเฉพาะ y r l v ṣ | 23 |
| 2.16 | ตัวอย่างการถอดรูปพยัญชนะสังโยคที่ซ้อนเฉพาะ r | 24 |
| 2.17 | ตัวอย่างการถอดรูปพยัญชนะหรือสระที่ไม่คำนึงถึงพยัญชนะสังโยค | 24 |
| 2.18 | สรุปตัวอย่างการตัวอย่างการถอดรูปพยัญชนะสังโยคแบบต่างๆ | 24 |
| 2.19 | ตัวอย่างการถอดรูปสระ a | 25 |
| 2.2 | ตัวอย่างการถอดรูปสระ a ขึ้นต้นคำ | 26 |
| 2.21 | ตัวอย่างการถอดรูปพยัญชนะที่ประสมสระ a ตามด้วยพยัญชนะ | 26 |
| 2.22 | ตัวอย่างการถอดรูปพยัญชนะที่ประสมสระ a ตามด้วย r | 26 |
| 2.23 | ตัวอย่างการถอดรูปพยัญชนะที่ประสมสระ a ตามด้วย v | 27 |
| 2.24 | ตัวอย่างการถอดรูปพยัญชนะที่ประสมสระ a ตามด้วย h | 27 |
| 2.25 | ตัวอย่างการถอดรูปอนุสวาระที่ไม่มีพยัญชนะโรมันอื่นตามติดมา | 28 |
| 2.26 | ตัวอย่างการถอดรูปอนุสวาระแบบสนธิมีพยัญชนะวรรคตาม | 28 |
| 2.27 | ตัวอย่างการถอดรูปอนุสวาระแบบสนธิมีพยัญชนะตามแบบอูษมะ | 29 |
| 2.28 | ตัวอย่างการถอดรูปอนุสวาระแบบไม่สนธิมีพยัญชนะตามแบบอูษมะ | 29 |
|  |  | **หน้า** |
|  |  |  |
| 2.29 | ตัวอย่างการถอดรูปอนุสวาระและมีพยัญชนะตามแบบอันตัสถะ | 30 |
| 2.30 | ตัวอย่างการถอดรูปเครื่องหมายวิสรรคะ | 30 |
| 2.31 | ตัวอย่างการถอดรูปพยัญชนะสังโยคแบบคำที่ใช้ในภาษาไทย | 31 |
| 2.32 | ตัวอย่างการถอดรูปพยัญชนะสังโยคแบบคำควบกล้ำไทย | 31 |
| 3.1 | สัญลักษณ์ประกอบขั้นตอนวิธีของกฎการปริวรรต | 46 |
| 3.2 | กำหนดกฎการปริวรรตอักษรไทยแบบคงรูป | 48 |
| 3.3 | กำหนดกฎการปริวรรตอักษรไทยแบบปรับรูป | 49 |
| 3.4 | การพิสูจน์ขั้นตอนวิธี (Algorithm) ในการปริวรรตอักษร 3 วิธี | 63 |
| 4.1 | ผลการทดสอบแบบแยกส่วน (Unit Testing) | 75 |
| 4.3 | ผลการทดสอบโดยการเปรียบเทียบกับตัวอย่างการปริวรรตแบบปรับรูป | 76 |
| 4.4 | ผลการทดสอบประสิทธิภาพของระบบ | 77 |
| 4.5 | ผลการประเมินการปริวรรตจากอักษรโรมันเป็นอักษรไทยแบบคงรูป | 78 |
| 4.6 | ผลการประเมินการปริวรรตจากอักษรโรมันเป็นอักษรไทยแบบปรับรูป | 78 |
| 4.7 | ผลการประเมินตารางการเปรียบเทียบผลการปริวรรต | 79 |
| 4.8 | ผลการประเมินตารางการเปรียบเทียบอักษร | 79 |
| 4.9 | ผลการประเมินด้านความง่ายต่อการใช้งานระบบ (Usability) | 80 |

**รายการรูปประกอบ**

|  |  |  |
| --- | --- | --- |
| **รูป** | | **หน้า** |
|  |  |  |
| 2.1 | ตัวอย่างการเขียนภาษาสันสกฤตด้วยอักษรชนิดต่างๆ | 9 |
| 2.2 | ตัวอย่างการเขียนภาษาสันสกฤตด้วยอักษรโรมันแบบต่างๆ | 10 |
| 2.3 | ตัวอย่างจารึกภาษาสันสกฤตในประเทศไทย | 10 |
| 2.4 | การปริวรรตอักษรภาษาสันสกฤตเป็นอักษรโรมันตามข้อเสนอของบุคคลต่างๆ | 11 |
| 2.5 | บางส่วนในรายงานมติที่ประชุม เรื่องการปริวรรตอักษรภาษาสันสกฤตและบาลี | 13 |
| 2.6 | การปริวรรตอักษรภาษาสันสกฤตแบบโรมัน IAST เป็นไทยรูปแบบต่างๆ | 14 |
| 2.7 | ตัวอย่างลักษณะพยัญชนะสังโยค อักษรเทวนาครีเทียบกับอักษรโรมัน | 18 |
| 3.1 | สรุปภาพรวมของงาน | 45 |
| 3.2 | ผลลัพธ์หลังจากการพัฒนาตัวต้นแบบการปริวรรตอักษร (Prototype) | 63 |
| 3.3 | โครงสร้างใน class diagram | 65 |
| 4.1 | หน้าแรกของตัวระบบ | 71 |
| 4.2 | หน้าปริวรรตอักษรส่วนฟอร์มรับข้อมูล | 72 |
| 4.3 | หน้าปริวรรตอักษรส่วนตารางเปรียบเทียบผลการปริวรรต | 72 |
| 4.4 | หน้าปริวรรตอักษรส่วนตารางเปรียบเทียบตัวอักษร | 73 |
| 4.5 | หน้าหลักตัวอย่างการปริวรรตอักษร | 73 |
| 4.6 | หน้าตัวอย่างการปริวรรต | 74 |