

การประยุกต์ใช้เทคนิคสารสนเทศเพื่อสกัดชื่อกษัตริย์ผู้สร้างวัด ในสมัยล้านนา จากเอกสารทางประวัติศาสตร์

Applied Information Extraction technique for extracting the king name who build a temple in Lanna Kingdom from historical documents

พงศ์กร จันทราช (Pongkorn Chantaraj)* และ จรัสศรี รุ่งรัตนอุบล (Jaratsri Rungrattanaubol)*

Received : June 1, 2018
Revised : August 31, 2018
Accepted : September 21, 2018

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ที่จะประยุกต์ใช้เทคนิคสารสนเทศเพื่อสกัดชื่อกษัตริย์ผู้สร้างวัดในสมัยล้านนาจากเอกสารประวัติศาสตร์ในรูปแบบไฟล์ข้อความ โดยกระบวนการวิจัยเริ่มจากรวบรวมเอกสารข้อมูลประวัติวัด ในเขตอำเภอเมืองเชียงใหม่ ซึ่งในอดีตได้ชื่อว่าเป็นศูนย์กลางการปกครองของอาณาจักรล้านนา โดยนำเข้าข้อมูลเอกสารที่อยู่ในรูปของเอกสารอิเล็กทรอนิกส์ ประเภทไฟล์ข้อความ (Text File) สู่กระบวนการสกัดชื่อกษัตริย์ผู้สร้างวัด โดยเริ่มจากการตัดคำภาษาไทยที่เน้นการแยกคำแต่ละคำออกจากกันโดยที่ยังคงความหมายเดิม ด้วยโปรแกรมเลกซ์โต (LexTo) โดยผู้วิจัยได้ออกแบบขั้นตอนวิธีและพัฒนาโปรแกรมด้วยภาษาจาวา เพื่อสกัดชื่อกษัตริย์ผู้สร้างวัด โดยประยุกต์ใช้การรู้จำชื่อเฉพาะด้วยการกำหนดชื่อกษัตริย์และคำสำคัญที่บ่งบอกถึงการสร้างวัดไว้ในพจนานุกรมข้อมูลเพื่อใช้ในการอ้างอิง จากนั้นไปทดสอบกับเอกสารประวัติวัดจำนวน 100 วัด และวัดประสิทธิภาพความถูกต้องของการสกัดสารสนเทศ โดยเปรียบเทียบผลลัพธ์ที่ได้กับข้อเท็จจริงที่ปรากฏในเอกสารต้นฉบับ พบว่าความถูกต้องของการทำงานเท่ากับร้อยละ 97 ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าขั้นตอนวิธีเทคนิคสารสนเทศที่นำเสนอในงานวิจัยนี้ สามารถนำไปประยุกต์ใช้ในการสกัดชื่อกษัตริย์ผู้สร้างวัดในสมัยล้านนาจากเอกสารทางประวัติศาสตร์ได้อย่างมีประสิทธิภาพ

คำสำคัญ: การรู้จำชื่อเฉพาะ การสกัดนิพจน์ระบุนาม การสกัดสารสนเทศ การตัดคำ ประวัติศาสตร์ล้านนา

Abstract

The objective of this study is to propose a technique to extract the king name in Lanna kingdom, who built the temple, from historical documents. The research processes began with collecting the data of temples in Amphor Muang Chiangmai, which is the capital and administrative center of the Lanna kingdom. The data transformed into electronic document or a text file are passed to the LexTo program for word segmentation with the nearly correct meaning. After that designing algorithm and the development of Java program to extract the king name who built the temple was proceeded. The algorithm is designed based on Named Entity Recognition (NER) by defining king names and some keywords, that indicates the temple building, in form of dictionary to be used as a reference. Then the experiment followed by the evaluation of the performance is set on 100 temples, by comparing the extracted results to the correct answers from the original documents. The accuracy is 97 percent. From the result indicates that the proposed technique can be efficiently applied to extract the king name of the temple builder in the Lanna kingdom from the historical documents.

Keywords: Named Entity Recognition, Name Expression Extraction, Information Extraction, Word Segmentation, History of Lanna Kingdom.

* ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

* Department of Computer Science and Information Technology, Faculty of Science, Naresuan University.

1. บทนำ

ประเทศไทยได้ชื่อว่าเป็นเมืองพุทธศาสนาอย่างช้านานเพราะประชากรส่วนใหญ่นับถือศาสนาพุทธ โดยในประเทศไทยมีผู้นับถือศาสนาพุทธ ร้อยละ 94.6 [1] ทำให้วัดเป็นศูนย์กลางทางพระพุทธศาสนาที่สำคัญของผู้คน ชุมชน หมู่บ้าน หรือเมืองมาอย่างช้านาน จากอดีตจนถึงปัจจุบันวัดทำหน้าที่เป็นศูนย์รวมความสามัคคีและการดำเนินชีวิตของสังคมไทยเป็นสถานที่ที่ประชาชนสามารถมาพบปะพูดคุย แลกเปลี่ยนความคิดเห็นซึ่งกันและกัน ชีวิตคนไทยผูกพันอยู่กับพระพุทธศาสนา ไม่ว่าจะเป็นชนชั้นประชาชน ชุนนาง จนถึงพระมหากษัตริย์ จากหลักฐานทางประวัติศาสตร์ กษัตริย์ทรงสร้างวัด ออกผนวช และทรงบำเพ็ญบุญกุศลต่างๆ โดยมีการจดบันทึก จัดเก็บหลักฐานในรูปแบบต่างๆ เช่น จารึก ตำรา จดหมายเหตุ ตลอดจนหลักฐานทางโบราณคดี เช่น ซากเจดีย์ วิหาร อุโบสถ พระพุทธรูป เป็นต้น ซึ่งพบว่าการสร้างวัดเป็นภาระหน้าที่หนึ่งของกษัตริย์เพื่อให้ชาวเมืองใช้เป็นสถานที่ประกอบพิธีทางศาสนาและยึดเหนี่ยวจิตใจ วัดจึงถูกสร้างพร้อมกับการสร้างเมืองเสมอ [2]

ในอดีตประเทศไทยหรือสยามมีด้วยกันหลายแคว้น โดยอาณาจักรล้านนาเป็นดินแดนแห่งหนึ่งที่มีความสำคัญทางประวัติศาสตร์มายาวนาน ประชาชนนับถือพุทธศาสนามาเป็นเวลานานนับพันปี พุทธศาสนาได้เข้ามาเผยแผ่และเจริญรุ่งเรืองตั้งแต่พุทธศตวรรษ ที่ 17 เป็นต้นมา [2] เกิดการสร้างวัดและศาสนวัตถุเป็นจำนวนมาก โดยข้อมูลที่เกี่ยวข้องกับประวัติวัด ความเป็นมา กษัตริย์ผู้สร้าง และปีที่สร้าง ฯลฯ ของแต่ละวัด พบว่ามีการจัดเก็บและเผยแพร่ในรูปแบบหนังสือ ตำรา และเอกสารอิเล็กทรอนิกส์ ส่วนใหญ่จะเป็นเอกสารเวิร์ด (Word) หรือพีดีเอฟ (PDF) ซึ่งจัดเป็นเอกสารที่ไม่มีโครงสร้าง (Unstructured Data) และไม่มี การสกัดสารสนเทศที่ระบุได้ว่าวัดแห่งนั้นสร้างโดยกษัตริย์พระองค์ใดหรือสร้างช่วงเวลาไหนในสมัยล้านนา ทำให้การค้นหาข้อมูลเหล่านี้ ต้องอ่านจากเอกสารต้นฉบับทำให้ไม่สะดวกและเกิดความล่าช้า จากปัญหาที่กล่าวมาหากนำเอกสารดังกล่าวมาประมวลผลด้วยการสกัดสารสนเทศอย่างเป็นระบบ จะก่อให้เกิดองค์ความรู้ทางด้านประวัติศาสตร์ที่คนในสังคมสามารถรับรู้และเข้าใจร่วมกันได้ง่ายและสะดวกขึ้น ซึ่งจะแสดงให้เห็นถึงความมั่นคง เจริญรุ่งเรือง การอยู่ร่วมกันอย่างสงบสุข โดยมีสถาบันพระมหากษัตริย์เป็น

สถาบันหลักสืบต่อมาจนถึงปัจจุบัน

เทคโนโลยีสารสนเทศ ซึ่งประกอบด้วยเทคโนโลยีคอมพิวเตอร์และเทคโนโลยีการสื่อสาร (Information and Communication Technology) มีการพัฒนาไปอย่างรวดเร็ว โดยภาครัฐได้ให้ความสำคัญในการผลักดันนโยบายเศรษฐกิจและสังคมดิจิทัล (Digital Economy) ที่มุ่งเน้นการนำเทคโนโลยีดิจิทัลมาสนับสนุนและขับเคลื่อนภาคเศรษฐกิจและสังคมของประเทศ เพื่อให้ไทยก้าวทันนานาอารยประเทศ และการเข้าสู่เศรษฐกิจดิจิทัลอย่างยั่งยืน ดังนั้นเพื่อให้เกิดประโยชน์ต่อสังคมและผู้สนใจด้านประวัติศาสตร์ที่เกี่ยวข้องกับวัด ผู้วิจัยจึงได้ประยุกต์ใช้หลักการสกัดสารสนเทศเพื่อค้นหาและวิเคราะห์คำที่ต้องการออกจากเอกสารประวัติวัด โดยใช้เทคนิคการรู้จำชื่อเฉพาะสำหรับดึงข้อมูลเพื่อระบุถึงชื่อกษัตริย์ผู้สร้างวัด ปีที่สร้างวัด ภายในข้อความ โดยใช้ภาษาจาวาในการพัฒนาระบบและใช้โปรแกรมเล็กซ์โตในการตัดคำภาษาไทย จึงได้ทำงานวิจัยเรื่อง “การประยุกต์ใช้เทคนิคสารสนเทศเพื่อสกัดข้อมูลประวัติผู้สร้างวัด ในสมัยล้านนา จากเอกสารทางประวัติศาสตร์” มีจุดมุ่งหมายเพื่อสกัดข้อมูลประวัติผู้สร้างวัด หรือช่วงเวลาในการสร้างวัดจากเอกสารไฟล์ข้อความ โดยสิ่งที่สกัดได้สามารถนำไปพัฒนาต่อยอดเพื่อการสืบค้นข้อมูลเชิงเวลาที่เกี่ยวข้องกับประวัติศาสตร์ล้านนาได้ และนำไปประยุกต์ใช้เพื่อแสดงความสัมพันธ์ของสถานที่ เวลา และกษัตริย์ผู้สร้างวัดได้ ซึ่งอาจก่อให้เกิดองค์ความรู้ด้านประวัติศาสตร์ที่สามารถนำมาใช้งานและสร้างความภาคภูมิใจในเอกลักษณ์ของชาติสืบต่อไป

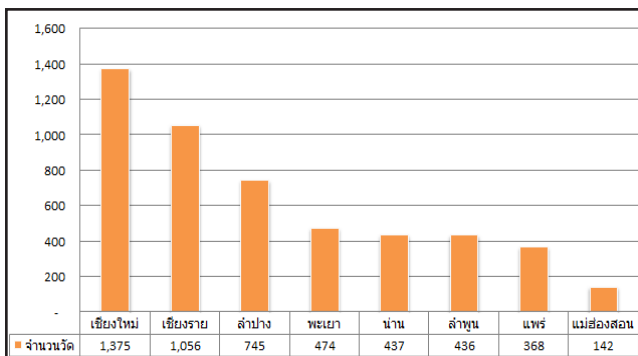
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 อาณาจักรล้านนา

อาณาจักรล้านนาได้ถูกสถาปนาขึ้นเมื่อปี พ.ศ. 1839 โดยพญามังราย พระองค์ได้ยึดครองแคว้นหริภุญไชยและสร้างเมืองเชียงใหม่ขึ้นเป็นราชธานี ทรงยอมรับสืบทอดอารยธรรมทางศาสนาจากเมืองหริภุญไชย ทรงยกพระตำหนักที่ประทับถวายเป็นพระอารามหลวงชื่อ “วัดเชียงใหม่” ซึ่งจัดเป็นวัดแห่งแรกในเขตกำแพงเมืองและเป็นวัดที่เก่าแก่ที่สุดของเมืองเชียงใหม่ และเคยเป็นที่ประดิษฐานของพระพุทธรูปสำคัญคือพระเสด็จมณีหรือพระแก้วขาว ซึ่งเป็นที่เคารพสักการะของประชาชน ในสมัยพญามังราย พระพุทธศาสนาเจริญรุ่งเรืองในดินแดนล้านนาเป็นอย่างมาก โดยมีเมือง

เชียงใหม่เป็นศูนย์กลางการศึกษาพระพุทธศาสนาและเผยแผ่พระพุทธศาสนาไปตามเมืองต่าง ๆ ในเขตทางตอนเหนือของอาณาจักรทำให้ล้านนามีชื่อเสียงในฐานะรัฐพุทธศาสนาที่ยิ่งใหญ่ [3]

พัฒนาการของประวัติศาสตร์อาณาจักรล้านนาสามารถแบ่งออกเป็น 3 สมัยคือ สมัยอาณาจักรล้านนารุ่งเรือง สมัยภายใต้การปกครองของพม่า และสมัยฐานประเทศราชของสยาม โดยอาณาจักรล้านนาในสมัยแรกมีกษัตริย์ปกครองทั้งหมด 17 พระองค์ ได้แก่ พญามังราย พญาไชยสงคราม พญาแสนพู พญาคำฟู พญาผายู พญาเกือนา พญาแสนเมืองมา พญาสามฝั่งแกน พญาติโลกราช พญาอดเชียงราย พญาแก้ว พญาเกษมราช ท้าวชาย มหาเทวีเจ้า จิรประภา พญาอุปฮาด พญาแม่กุ และพระนางวิสุทธิเทวี [4] ปัจจุบันดินแดนอาณาจักรล้านนาในประเทศไทยประกอบด้วย 8 จังหวัดทางภาคเหนือ ได้แก่ เชียงใหม่ เชียงราย ลำพูน ลำปาง แม่ฮ่องสอน พะเยา แพร่ และน่าน โดยในพื้นที่นี้มีจำนวนวัดทั้งหมด 5,033 วัด [5] กระจายอยู่ในจังหวัดต่างๆ โดยข้อมูลประวัติวัดเหล่านี้มีความสำคัญทางด้านประวัติศาสตร์ ศิลปะและวัฒนธรรมล้านนามาอย่างยาวนาน

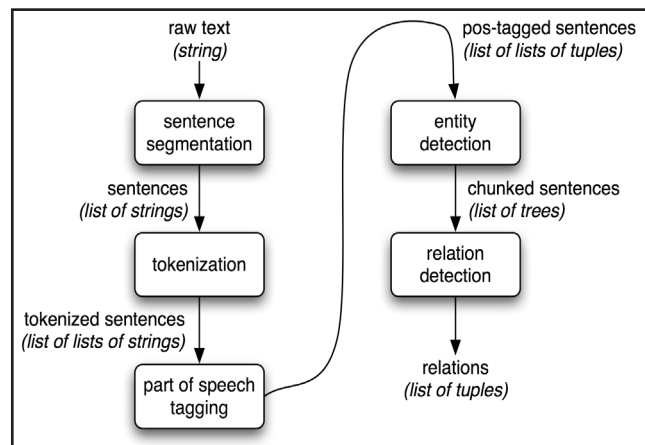


ภาพที่ 1 แผนภูมิแท่งจำนวนวัด 8 จังหวัดภาคเหนือ

2.2 การสกัดสารสนเทศและการรู้จำชื่อเฉพาะ

ปัจจุบันข้อมูล สารสนเทศมีอัตราการเติบโตเพิ่มขึ้น การจัดการข้อมูลให้เป็นระบบมีความสำคัญต่อการสร้างฐานความรู้ที่มีประสิทธิภาพ ทำให้การรวบรวมข้อมูล และการสกัดสารสนเทศมีความสำคัญอย่างยิ่งที่จะช่วยจัดการกับปริมาณข้อมูลที่เพิ่มมากขึ้น เพราะการสกัดสารสนเทศ (Information Extraction) คือการสกัดโครงสร้างข้อมูลหรือองค์ความรู้จากข้อความที่ไม่มีโครงสร้างโดยการระบุการอ้างอิงไปยังชื่อเฉพาะ (Name Entities) และความสัมพันธ์ระหว่างชื่อเฉพาะนั้น ซึ่งสามารถใช้ในการสกัดความรู้จาก

คลังข้อความหรือเพื่อดึงข้อมูลจากชุดเอกสารได้อย่างถูกต้อง [6] โดยกระบวนการรู้จำชื่อเฉพาะที่เป็นวิธีการหนึ่งในการสกัดสารสนเทศได้ถูกนำมาใช้อย่างแพร่หลาย โดยการรู้จำชื่อเฉพาะ (Named Entity Recognition) คืองานที่เกี่ยวข้องกับการดึงข้อมูลเพื่อระบุและจัดประเภท ที่การกล่าวถึงบุคคล องค์กร สถานที่ และชื่อเฉพาะอื่นๆ ภายในข้อความ เป็นส่วนประกอบหลักในการประมวลผลภาษาธรรมชาติ รวมถึง การถาม-ตอบ การสรุปความ และการแปลภาษาของเครื่อง [7] โดยมีส่วนประกอบของสถาปัตยกรรมการสกัดสารสนเทศ ดังภาพที่ 2



ภาพที่ 2 สถาปัตยกรรมการสกัดสารสนเทศ [8]

จากภาพที่ 2 แสดงให้เห็นถึงสถาปัตยกรรมของการสกัดสารสนเทศ โดยมีรายละเอียดการทำงานเริ่มตั้งแต่นำเข้าเอกสารประเภทข้อความ ทำการแบ่งข้อความออกเป็นประโยค นำแต่ละประโยคแบ่งย่อยเป็นคำโดยใช้ Tokenizer โดยจะมีการระบุชนิดของคำให้เหมาะสม (Part of speech) แต่ละประโยคจะถูกจำแนกหน้าที่ของคำ ตรวจสอบชื่อเฉพาะ (Name Entity) ซึ่งในขั้นตอนนี้เป็นขั้นตอนการค้นหาชื่อเฉพาะที่ให้ความสนใจในแต่ละประโยค ขั้นตอนสุดท้ายตรวจสอบความสัมพันธ์ เพื่อค้นหาความสัมพันธ์ที่เป็นไปได้ระหว่างชื่อเฉพาะต่างๆ ในข้อความ

2.3 การประยุกต์ใช้หลักการสกัดสารสนเทศ

ปัจจุบันมีงานวิจัยที่นำหลักการสกัดสารสนเทศและการรู้จำชื่อเฉพาะ ไปประยุกต์ใช้อย่างกว้างขวางในศาสตร์หลายแขนง และหลายภาษา อาทิเช่น การวิจัยเกี่ยวกับการสกัดค่านามจากเอกสารภาษามลายูโบราณ โดยใช้หลักการรู้จำชื่อเฉพาะร่วมกับคำศัพท์ภาษามลายูโบราณ และผู้เชี่ยวชาญด้านภาษาในการพิจารณาความถูกต้องร่วมกัน [9] การสกัด

ข้อมูลจากข้อความภาษาอาหรับที่เป็นทางการและภาษาอาหรับที่ไม่เป็นทางการ (ภาษาพูด) โดยการสกัดชื่อจากข้อความแบบอ้างอิงพจนานุกรมข้อมูล สามารถนำไปสู่การแยกชื่อของบุคคลได้อย่างถูกต้อง [10] การสกัดความสัมพันธ์ของนิพจน์ระบุนามชื่อภาษาจีนที่เรียกว่า LaSE (Location and Semantic Extraction) เป็นแนวทางการรวมข้อมูลที่ได้จากตำแหน่งของคำและการใช้ความหมายตามพจนานุกรม HowNet พบว่าผลลัพธ์ที่ได้มีประสิทธิภาพดีกว่าวิธีการสกัดแบบเดิมที่ใช้คุณลักษณะตำแหน่งหรือคุณลักษณะทางความหมายเพียงอย่างเดียว [11] การรู้จำชื่อในข้อความภาษามลายู จากบทความหนังสือพิมพ์ภาษามลายู โดยใช้แอปพลิเคชันสำหรับจัดทำดัชนีชื่อ จากชุดข้อความมลายู และจัดทำเครื่องมือโดยใช้ภาษา Perl ร่วมกับ Regular Expression (RE) เพื่อกำหนดรูปแบบหรือกลุ่มคำสำหรับใช้ค้นหาข้อความตามที่ต้องการ พบว่าได้ผลลัพธ์ที่แม่นยำ [12] และงานวิจัยที่เกี่ยวกับการแยกชื่อสถานที่ออกจากบทความเกี่ยวกับการท่องเที่ยวในบล็อกแบบอัตโนมัติ โดยการวิเคราะห์ข้อจำกัด บริบทของการท่องเที่ยว ร่วมกับพจนานุกรมชื่อสถานที่โดยใช้วิธีการ Classification network ทำให้เกิดความชัดเจนในการจำแนกประเภทของชื่อสถานที่ [13] นอกจากนี้ยังมีงานวิจัยที่นำหลักการรู้จำชื่อเฉพาะมาประยุกต์ใช้ร่วมกับภาษาไทย อาทิเช่น การวิเคราะห์เนื้อหาข่าวภาษาไทยเพื่อติดแท็กอัตโนมัติ โดยใช้เทคนิคการสกัดคำ ซึ่งทำให้สามารถวิเคราะห์หาคำสำคัญ (Keyword) จากหัวข้อข่าวและข้อมูลรายละเอียดของข่าว โดยพิจารณาจากค่าความถี่ของคำแต่ละคำในบทความ แล้วเปรียบเทียบประสิทธิภาพความถูกต้องระหว่างการหาคำสำคัญด้วยคอมพิวเตอร์แบบอัตโนมัติกับข้อมูลที่มีการหาคำสำคัญโดยมนุษย์ ซึ่งสามารถช่วยให้ความรวดเร็วในการค้นคืนสารสนเทศได้ [14] ส่วนปัญหาของการรู้จำชื่อเฉพาะในเอกสารภาษาไทย การเขียนติดกันทั้งประโยคและการไม่มีขอบเขตที่แน่นอน ไม่มีตัวอักษรพิมพ์ใหญ่กับพิมพ์เล็กในการใช้นิพจน์ระบุนามเหมือนกับภาษาอังกฤษ และสามารถเกิดขึ้นใหม่ได้ตลอดเวลา ทำให้ยากในการจดจำนิพจน์ระบุนามในภาษาไทย [15] และการรู้จำชื่อเฉพาะในภาษาไทยมีความแตกต่างที่ทำให้การสกัดชื่อเฉพาะยากกว่าภาษาอังกฤษ คือ การหาขอบเขตของคำ การที่ไม่มีข้อมูลบ่งบอกถึงชื่อเฉพาะ การสร้างคำในภาษาไทยไม่มีหลักการที่แน่นอน [16]

2.4 เครื่องมือสำหรับการตัดคำ

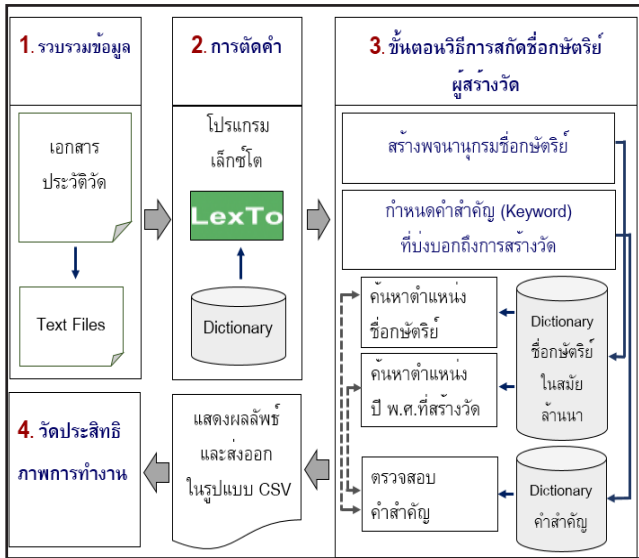
การสกัดสารสนเทศให้มีความถูกต้องนั้นในขั้นตอนแรกจะต้องมีการตัดคำ (Word Segmentation) เพื่อแยกคำแต่ละคำในประโยคของเอกสารก่อนที่จะนำไปวิเคราะห์ในขั้นตอนต่อไป ขั้นตอนการตัดคำมีความสำคัญอย่างยิ่งโดยเฉพาะการตัดคำภาษาไทยต้องอาศัยเครื่องมือที่มีประสิทธิภาพในการประมวลผล ซึ่ง ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ [17] ได้พัฒนาโปรแกรมเล็กซ์โต (A Thai Lexeme Tokenization and Normalization Tool: LexTo) เป็นระบบแบ่งคำสำหรับภาษาไทยที่ให้ความแม่นยำสูงแบบอิงพจนานุกรม (Dictionary Based) โดยใช้เทคนิคการแบ่งคำแบบ Longest Matching ผู้ใช้สามารถเพิ่มรายการคำศัพท์ได้ตามต้องการ เพื่อให้การตัดคำเหมาะสมกับงานที่นำไปใช้ปัจจุบันถูกนำไปใช้ในงานวิจัยอย่างแพร่หลาย อาทิเช่น การสกัดคุณลักษณะต่างๆ จากข้อความออนไลน์ โดยใช้โปรแกรมที่พัฒนาขึ้นเอง ร่วมกับโมดูลการตัดคำภาษาไทยอาศัยโค้ดจากโปรแกรมเล็กซ์โต [18] และการสกัดคุณลักษณะเด่นของหมวดหมู่สถานที่ท่องเที่ยว โดยรวบรวมชื่อและคำอธิบายสถานที่ท่องเที่ยวในจังหวัดภูเก็ตจากอินเทอร์เน็ต แล้วนำมาผ่านโปรแกรมเล็กซ์โตตัดคำภาษาไทยเพื่อนับจำนวนคำที่ถูกใช้ในคำอธิบายสถานที่ท่องเที่ยวมากที่สุดของแต่ละหมวดหมู่ [19] จากที่กล่าวมาพบว่าการโปรแกรมเล็กซ์โต สามารถใช้เป็นเครื่องมือช่วยตัดคำภาษาไทย เพื่อช่วยให้การสกัดสารสนเทศมีความถูกต้องและมีประสิทธิภาพ

3. วิธีดำเนินการวิจัย

งานวิจัยนี้ ผู้วิจัยได้นำเสนอขั้นตอนการประยุกต์ใช้เทคนิคสารสนเทศเพื่อสกัดข้อมูลผู้สร้างวัดในสมัยล้านนาแบ่งออกเป็น 4 ขั้นตอนแสดงได้ดังภาพที่ 3

กระบวนการดำเนินงานของงานวิจัยนี้สามารถแบ่งได้เป็น 4 ขั้นตอนคือ

- (1) การรวบรวมข้อมูล ซึ่งในที่นี้คือเอกสารข้อมูลประวัติวัดในเขตอำเภอเมืองเชียงใหม่
- (2) การตัดคำ จากเอกสารข้อมูลในรูปแบบไฟล์ข้อความ (Text file) ที่รวบรวมมาจะถูกส่งเข้าโปรแกรมตัดคำภาษาไทยด้วยโปรแกรมเล็กซ์โต (LexTo)
- (3) ขั้นตอนวิธีการสกัดข้อมูลผู้สร้างวัด เป็นขั้นตอน



ภาพที่ 3 กรอบแนวคิดของงานวิจัย

ที่ผู้วิจัยได้ออกแบบขั้นตอนวิธีเองเพื่อที่จะได้ผลลัพธ์ของการสกัด คือชื่อกษัตริย์ผู้สร้างวัด หรือ ปี พ.ศ. ที่สร้างวัด

(4) วัดประสิทธิภาพการทำงาน โดยในขั้นนี้เป็นการประเมินความถูกต้องของการสกัดข้อมูล ด้วยการเปรียบเทียบผลลัพธ์ที่ได้กับข้อเท็จจริงในเอกสารข้อมูลประวัติวัด

โดยรายละเอียดของแต่ละขั้นตอนได้ถูกอธิบายเพิ่มเติมดังต่อไปนี้

3.1 การรวบรวมข้อมูล

ในอดีตเชียงใหม่เคยเป็นศูนย์กลางการปกครองและเป็นเมืองหลวงของอาณาจักรล้านนา โดยพื้นที่ส่วนนี้ปัจจุบันคืออำเภอเมืองเชียงใหม่ ทำให้การศึกษานี้ผู้วิจัยจะใช้ข้อมูลวัดในเขตอำเภอเมืองเชียงใหม่ซึ่งมีจำนวนทั้งหมด 123 วัด [20] โดยได้เลือกวัดที่มีการจัดเก็บหรือเผยแพร่ข้อมูลในรูปแบบเอกสารอิเล็กทรอนิกส์ที่มีเนื้อหาทางด้านประวัติวัด เช่น ความเป็นมา สถานที่ตั้ง ผู้สร้างวัด ปีที่สร้าง โบราณสถานหรือโบราณวัตถุที่สำคัญภายในวัด เป็นต้น จำนวน 120 วัดจากทั้งหมด 123 วัด ในเขตอำเภอเมืองเชียงใหม่ ในขั้นนี้เป็นการนำข้อมูลจากการเผยแพร่โดยสำนักงานวัฒนธรรมจังหวัดเชียงใหม่ จำนวน 58 วัด [21] และจากระบบฐานข้อมูลวัดจำนวน 62 วัดจาก [22] รวบรวมเอกสารและจัดการเอกสารเหล่านี้ในรูปแบบของไฟล์ข้อความ (Text file) ซึ่งมีลักษณะของเอกสารดังตัวอย่างที่แสดงในภาพที่ 4

โดยในขั้นนี้จะศึกษาเฉพาะวัดที่สร้างโดยกษัตริย์ในยุคแรกของอาณาจักรล้านนาที่มีกษัตริย์ปกครองอาณาจักรทั้งหมด 17 พระองค์ [4] โดยแต่ละพระองค์มีช่วงเวลาการครองราชย์

ในพงศาวดารโยนก กล่าวว่า วัดนันทารามได้รับการสถาปนาขึ้นเมื่อ พ.ศ.2060 ในสมัยของพญาแก้ว ฦ ที่ถวายพระเพลิง พญาอดเชียงราย ผู้เป็นพระราชบิดา แล้วจัดให้มีการสมโภชซึ่งวัดนี้คงมีมาก่อนแล้ว เพราะในตำนานมูลศาสนากล่าวถึงวัดนันทารามว่าเป็นวัดใหญ่และมี ความสำคัญ เมื่อ พ.ศ.1987 พระนันทปัญญาวัดนันทารามได้เป็นพระสังฆราชและเป็นอาจารย์ของพระมหาญาณคัมภีระผู้ซึ่งต่อมาเป็นเจ้าอาวาสวัดป่าแดง ซึ่งเป็นผู้นำพุทธศาสนาแบบสิงหลิกายหรือพุทธศาสนาแบบลังกาวงศ์มาสู่เมืองเชียงใหม่

ภาพที่ 4 ตัวอย่างไฟล์ข้อความประวัติวัด

ตารางที่ 1 ลำดับกษัตริย์ในสมัยล้านนาเรียง

ลำดับ	ชื่อกษัตริย์	ช่วงปี พ.ศ. การครองราชย์
1	พญามังราย	1804 - 1854
2	พญาไชยสงคราม	1854 - 1868
3	พญาแสนพู	1868 - 1877
4	พญาคำฟู	1877 - 1879
5	พญาผายู	1879 - 1898
6	พญาเกือนา	1898 - 1928
7	พญาแสนเมืองมา	1928 - 1944
8	พญาสามฝั่งแกน	1945 - 1984
9	พญาติโลกราช	1984 - 2030
10	พญาอดเชียงราย	2031 - 2038
11	พญาแก้ว	2038 - 2068
12	พญาเกสเชษฐราช	ครั้งที่ 1 2069 - 2081 ครั้งที่ 2 2086 - 2088
13	ท้าวชาย	2081 - 2086
14	มหาเทวีเจ้า จิระประภา	2088 - 2089
15	พญาอุปฮาด	2089 - 2090
16	พญาแม่กุ	2094 - 2107
17	พระนางวิสุทธิเทวี	2107 - 2121

ดังแสดงในตารางที่ 1

จากตารางที่ 1 ถ้าเราทราบชื่อกษัตริย์ผู้สร้างวัดเราก็จะสามารถรู้ช่วงเวลาการสร้างวัดแห่งนั้นได้

3.2 การตัดคำ

ไฟล์ข้อความประวัติวัดที่รวบรวมไว้จะถูกส่งเข้าโปรแกรมตัดคำ เพื่อแยกคำแต่ละคำในไฟล์ออกจากกัน โดยที่ยังคงความหมายเดิม ในงานวิจัยนี้ใช้โปรแกรมตัดคำภาษาไทยชื่อโปรแกรมเลกซ์โตที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ โดยผลลัพธ์การตัดคำจะมีลักษณะดังตัวอย่างที่แสดงดังภาพที่ 5

ในพงศาวดารโยนก |กล่าวว่า| วัดนันทาราม|ได้รับการสถาปนา|
ขึ้นเมื่อ|พ.ศ.2060| ในสมัย|ของ|พญาแก้ว| ณ|ที่|ถวายเป็น|พระ
เพลิง|พญาออตเชียงราย| ผู้เป็น|พระ|ราช|บิดา|แล้ว|จัด|ให้มีการ
สมโภช|ซึ่ง|วัด|นี้|มี|มา|ก่อน|แล้ว|เพราะ|ใน|ตำนาน|มูล|ศาสนา|
กล่าวถึง|วัดนันทาราม|ว่าเป็น|วัด|ใหญ่|และ|มีความ|สำคัญ|เมื่อ|
พ.ศ.1987| พระ|นันท|ปัญญา|วัดนันทาราม|ได้|เป็น|พระ|สังฆราช|
และ|เป็น|อาจารย์|ของ|พระ|มหา|ญาณคัมภีระ|ผู้|ซึ่ง|ต่อมา|เป็น|เจ้า
อาวาส|วัด|ป่าแดง| ซึ่งเป็น|ผู้นำ|พุทธศาสนา|แบบ|สิงหล|นิกาย|
หรือ|พุทธศาสนา|แบบ|ลังกา|วงศ์|มา|สู่|เมือง|เชียงใหม่|

ภาพที่ 5 ตัวอย่างผลลัพธ์ที่ได้จากการตัดคำ

ไฟล์ที่ผ่านกระบวนการตัดคำจะถูกนำไปใช้ในขั้นตอนการสกัดสารสนเทศต่อไป

3.3 ขั้นตอนวิธีการสกัดข้อมูลผู้สร้างวัด

งานวิจัยนี้ได้ออกแบบขั้นตอนวิธีและพัฒนาโปรแกรมเพื่อสกัดข้อมูลผู้สร้างวัดด้วยภาษาจาวา โดยได้ประยุกต์ใช้หลักการสกัดสารสนเทศแบบอ้างอิงพจนานุกรมและการรู้จำชื่อเฉพาะ โดยต้องการดึงข้อมูลชื่อผู้สร้างวัดหรือปีที่สร้างวัดภายในไฟล์ข้อความดังขั้นตอนต่อไป

ขั้นตอนที่ 1: สร้างพจนานุกรมชื่อผู้สร้างวัดในสมัยล้านนา ด้วยข้อมูลจากตารางที่ 1 โดยเริ่มต้นจากการเลือกไฟล์ข้อความประวัติวัดแบบไม่เจาะจงจำนวน 20 วัดมาศึกษา โดยได้พบว่าชื่อผู้สร้างวัดที่ปรากฏในเอกสารมีการเขียนหรือสะกดแตกต่างกันได้ เช่น พระเจ้ามังราย พระยามังราย พญามังราย พ่อขุนเม็งราย เป็นต้น ดังนั้นจึงทำการจัดกลุ่มชื่อผู้สร้างวัดที่มีความหมายเดียวกันและปรับปรุงพจนานุกรมชื่อผู้สร้างวัดในสมัยล้านนาให้สามารถใช้งานได้ถูกต้องยิ่งขึ้น ดังภาพที่ 6

kinglanna.txt	
1	พระยามังราย,พระเจ้ามังราย,พญามังราย,พ่อขุนเม็งราย
2	พระยาไชยสงคราม,พญาไชยสงคราม,
3	พระเจ้าแสนภู, พญาแสน,
4	พระเจ้าคำฟู, พญาคำฟู,
5	พระเจ้าผาย, พญาผาย,
6	พระเจ้ากือนา,พญากือนา,
7	พระเจ้าแสนเมืองมา, พญาแสนเมืองมา,
8	พระเจ้าสามฝั่งแกน, พญาสามฝั่งแกน,
9	พระเจ้าติโลกราช, พญาติโลกราช,

ภาพที่ 6 ตัวอย่างชื่อผู้สร้างวัดในพจนานุกรม

ขั้นตอนที่ 2: กำหนดคำสำคัญที่บ่งบอกถึงการสร้างวัด โดยศึกษาจากไฟล์ข้อความประวัติวัดทั้ง 20 ไฟล์เพื่อหากลุ่มคำและรูปแบบการจัดวางตำแหน่งของคำที่ระบุถึงการสร้าง โดยสังเกตได้ว่าคำเหล่านี้จะอยู่ในตำแหน่งด้านหน้าและหลังของชื่อผู้สร้างวัดดังแสดงในภาพที่ 7

สุเทพ,สร้าง,ขึ้น,ใน,รัชสมัย,พญา,กือนา,กษัตริย์,ลำดับ,ที่,2,แห่ง
วัดพระสิงห์,เดิม,ชื่อ,วัดพระเชียง,พญาผาย,ทรง,สถาปนา,เมื่อ,ปี,พ.ศ.
เมื่อ,พ.ศ.,2020,ผู้สถาปนา,วัดเจดีย์ยอด,พญาติโลกราช,กษัตริย์,องค์,ที่,9,แห่ง
พ.ศ.,2088,ว่า,เริ่ม,ตั้งแต่,พญา,เกษไชยราช,เสด็จ,มาจาก,ล้านช้าง,เพื่อ,ขึ้น
สร้าง,วัด,เมื่อ,ประมาณ,พ.ศ.,2387,ใน,สมัย,รัตนโกสินทร์,เดิม,บริเวณ
สร้าง,ขึ้น,เมื่อ,ปี,พ.ศ.,2040,โดย,ขุนนาง,ยศ,พัน,ชื่อ,ว่า,แหวน

ภาพที่ 7 ตัวอย่างคำสำคัญที่บ่งบอกถึงการสร้างวัดและตำแหน่งการวาง

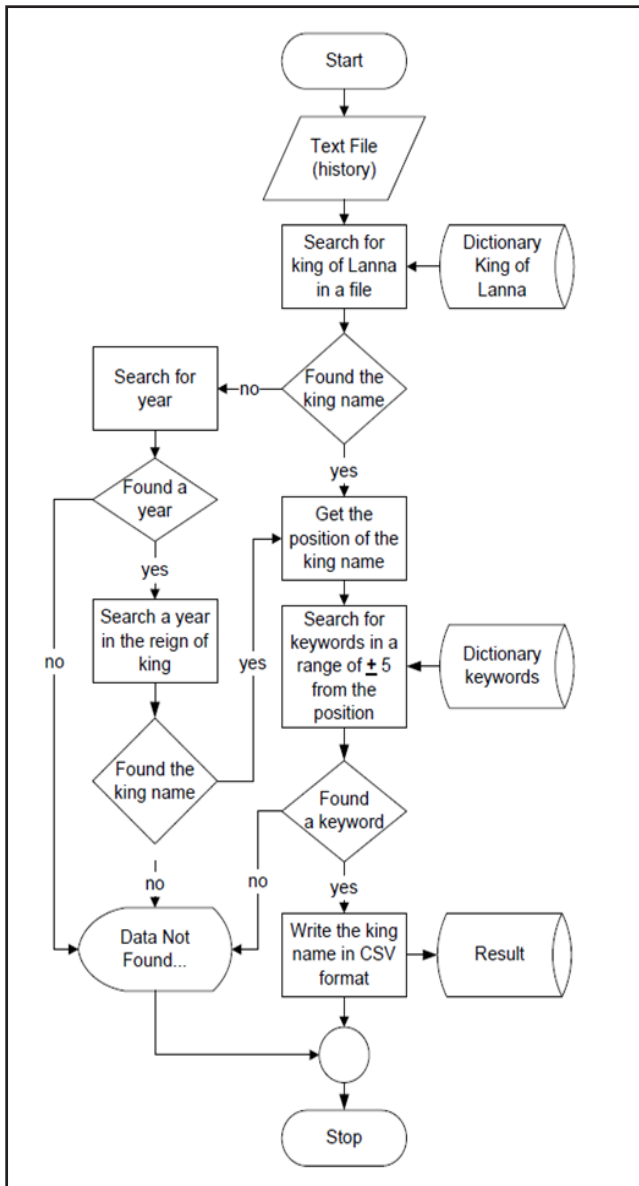
จากนั้นได้รวบรวมคำสำคัญที่บ่งบอกถึงการสร้างวัดมาสร้างเป็นไฟล์พจนานุกรมโดยมีตัวอย่างของคำดังภาพที่ 8

keyword.txt	
1	สร้าง,ผู้สร้าง,
2	สถาปนา,ผู้สถาปนา,ทรงสถาปนา,
3	โดย,
4	ตั้งแต่,
5	ปรากฏ

ภาพที่ 8 ตัวอย่างคำสำคัญบ่งบอกถึงการสร้างวัด

จากการศึกษาพบว่าคำสำคัญเหล่านี้จะปรากฏอยู่ในตำแหน่งที่ใกล้กับตำแหน่งที่พบชื่อผู้สร้างวัดนั้นๆ คือปรากฏด้านหน้าหรือด้านหลังในช่วงระยะ 5 คำ นับจากตำแหน่งที่พบชื่อผู้สร้างวัด โดยในกรณีที่ไม่มีพบชื่อผู้สร้างวัดจะทำการค้นหาปี พ.ศ. ที่สร้างวัดและนำไปเทียบเคียงกับปี พ.ศ. ของการครองราชย์เพื่อให้ได้ชื่อผู้สร้างวัดในที่สุด

ขั้นตอนที่ 3: พัฒนาโปรแกรมเพื่อสกัดชื่อกษัตริย์ผู้สร้างวัด โดยขั้นตอนการทำงานเริ่มตั้งแต่การนำเข้าไฟล์ข้อความ ประวัติวัดที่ผ่านการตัดคำแล้วเข้าสู่ระบบ โดยเริ่มค้นหาชื่อ กษัตริย์ที่ปรากฏในข้อความด้วยการอ้างอิงจากพจนานุกรม ชื่อกษัตริย์ในสมัยล้านนาที่สร้างขึ้น หากพบชื่อกษัตริย์ก็จะ รายงานคำคือชื่อกษัตริย์ออกมา หากไม่พบชื่อกษัตริย์ก็จะ ค้นหาปี พ.ศ. ที่สร้างวัดและนำปีที่ได้ไปเทียบกับช่วงปี พ.ศ. ช่วงของการครองราชย์ของกษัตริย์ โดยถ้าพบข้อมูลตาม เงื่อนไขจะคืนค่าชื่อกษัตริย์ออกมา โดยถ้าไม่พบจะคืนค่า ปีที่สร้างวัด



ภาพที่ 9 ผังงานของโปรแกรมการสกัดชื่อกษัตริย์ผู้สร้างวัดหรือปี พ.ศ. ที่สร้างวัด

โดยในการระบุความเป็นผู้สร้างหรือปีที่สร้างนั้นจะทำได้ โดยการตรวจสอบว่ามีคำสำคัญที่บ่งบอกถึงการสร้างวัดอยู่ในตำแหน่งด้านหน้าและด้านหลังของชื่อกษัตริย์หรือปี พ.ศ. ในระยะ 5 ตำแหน่งหรือไม่ถ้าพบก็จะส่งค่าชื่อกษัตริย์ผู้สร้างวัดแห่งนี้เขียนลงไฟล์ประเภท CSV แต่หากไม่พบก็จะแจ้งข้อความไม่พบข้อมูลดังกล่าว ดังผังงานในภาพที่ 9

โดยรายละเอียดโปรแกรมแสดงเป็นรหัสเทียม (Pseudo Code) ดังแสดงในภาพที่ 10

```

Begin "king name extraction algorithm"
Set arrayKing ← kingname.txt
Set arrayKeyword ← keyword.txt
Read history.txt
Get arrayTemple ← history.txt
for i=0 to length(arrayKing)
    for j=0 to length(arrayTemple)
        if arrayTemple[j]==arrayKing[i]
            search(j)
        else
            result=null
    if result==null
        for m=0 to length(arrayTemple)
            if arrayTemple[m]==matches("\\d{4}")
                search(m)
            else
                result="Data not found"
        End "algorithm"

function search(num)
    for k=num-5 to num+5
        for l=0 to length(arrayKeyword)
            if arrayTemple[k]==arrayKeyword[l]
                result=arrayTemple[k]
            if arrayTemple[k] is a number and in the year on throne of the king
                result=king name
        Write result to CSV file
    
```

ภาพที่ 10 รหัสเทียมวิธีการสกัดชื่อกษัตริย์

3.4 การวัดประสิทธิภาพการทำงานของระบบฯ

ผลลัพธ์ที่ได้จากการทำงานของระบบจะเป็นไฟล์ประเภท CSV ดังตัวอย่างไฟล์คำตอบที่แสดงในภาพที่ 11 โดยถ้าพบว่าการระบุชื่อกษัตริย์ผู้สร้างในเอกสารอย่างชัดเจนจะได้ชื่อกษัตริย์ดังคำตอบในลำดับ 01 - 04 โดยถ้าไม่ระบุชื่อกษัตริย์ผู้สร้างแต่ระบุปีที่สร้างในเอกสารจะได้คำตอบในลำดับ 05, 99, 100 กรณีที่ไม่ระบุชื่อกษัตริย์ผู้สร้างและปีที่สร้างในเอกสารจะได้คำตอบในลำดับ 98

การตรวจสอบความถูกต้องของผลลัพธ์ที่สกัดได้จากการประมวลผลของระบบที่จัดทำขึ้น โดยเปรียบเทียบกับคำตอบที่ถูกต้องตามที่ปรากฏในเอกสารต้นฉบับ



01,วัดคำคำ,พญาออตเชียงราย
02,วัดหม้อคำตาง,พญาดีโลกราช
03,วัดลังกา,พญาอ้อม
04,วัดพันอัน,พญาแก้ว
05,วัดเมืองล้ง,ไม่พบกษัตริย์สมัยล้านนาที่ครองราชย์ปี 2300
...
...
98,วัดธาตุคำ,ไม่ระบุชื่อกษัตริย์ผู้สร้าง และปีที่สร้าง
99,วัดพากช้าง,สร้างปี 2026 สมัยพญาดีโลกราช
100,วัดตาลดิ่ง,สร้างปี 2040 สมัยพญาแก้ว

ภาพที่ 11 ผลลัพธ์ในรูปแบบไฟล์ประเภท CSV

4. ผลการดำเนินงาน

งานวิจัยนี้ได้นำเสนอขั้นตอนวิธีการสกัดข้อมูลผู้สร้างวัดในสมัยล้านนาจากเอกสารประวัติวัด โดยเน้นใช้เทคนิคสารสนเทศและเครื่องมือที่มีอยู่โดยนำสิ่งเหล่านี้มาประยุกต์ใช้ร่วมกัน ระบบที่พัฒนาขึ้นมาโดยใช้โปรแกรมตัดคำภาษาไทยเล็กซ์โตและการเขียนโปรแกรมด้วยภาษาจาวา เพื่อสกัดข้อมูลผู้สร้างวัด ได้นำไปใช้กับไฟล์ข้อความประวัติวัดจำนวน 100 วัด โดยจากผลลัพธ์การทำงาน ผู้วิจัยแบ่งเอกสารตามลักษณะของผลลัพธ์เป็น 3 แบบ ดังตารางที่ 2

ตารางที่ 2 ลักษณะของเอกสารประวัติวัด

เอกสาร	รายการ
แบบที่ 1	มีการระบุชื่อกษัตริย์ผู้สร้างในเอกสารอย่างชัดเจน
แบบที่ 2	ไม่ระบุชื่อกษัตริย์ผู้สร้างแต่ระบุปีที่สร้างในเอกสาร
แบบที่ 3	ไม่ระบุชื่อกษัตริย์ผู้สร้างและปีที่สร้าง

ผลลัพธ์ที่ได้จากการประมวลผลมีรายละเอียด ดังตารางที่ 3 โดยการคำนวณความถูกต้องของผลลัพธ์เป็นดังสมการที่ 1

$$\text{ร้อยละความถูกต้อง} = \frac{\text{จำนวนเอกสารที่ถูกต้อง}}{\text{จำนวนเอกสารทั้งหมด}} \times 100 \quad (1)$$

ตารางที่ 3 ผลลัพธ์จากการประมวลผล

เอกสาร	จำนวนเอกสาร	ผลลัพธ์		ร้อยละความถูกต้อง
		ถูกต้อง	ไม่ถูกต้อง	
แบบที่ 1	38	35	3	92
แบบที่ 2	41	41	0	100
แบบที่ 3	21	21	0	100
รวม	100	97	3	97

จากตารางที่ 3 พบว่าการทำงานเพื่อสกัดข้อมูลผู้สร้างวัดในเอกสารแบบที่ 1 ยังมีข้อผิดพลาดอยู่ในขณะที่แบบที่เหลือได้ผลลัพธ์ถูกต้องทั้งหมด โดยความถูกต้องของการทำงานเท่ากับร้อยละ 97

5. สรุป

อภิปรายผล

จากการสกัดข้อมูลผู้สร้างวัดจากเอกสารข้อมูลประวัติวัดพบว่าข้อผิดพลาดในเอกสารแบบที่ 1 ทั้งหมด 3 รายการ ซึ่งเกิดจากเอกสารดังกล่าวมีการระบุชื่อกษัตริย์แต่ไม่ระบุคำสำคัญที่เกี่ยวกับการสร้างวัดทำให้ไม่สามารถสกัดได้ว่ากษัตริย์พระองค์ใดเป็นผู้สร้างวัดแห่งนั้น ซึ่งทางผู้วิจัยต้องหาแนวทางในการปรับปรุงต่อไป ในส่วนของเอกสารแบบอื่น ระบบสามารถสกัดคำตอบได้อย่างถูกต้อง

ปัญหาที่พบและแนวทางการปรับปรุง

ไฟล์ต้นฉบับประวัติวัดบางฉบับที่ใช้ในงานวิจัยนี้เป็นเอกสารที่ไม่มีการจัดรูปแบบที่ดีพอ เช่น การกำหนดวรรคตอนในการพิมพ์ที่ไม่ถูกต้อง การพิมพ์ปี พ.ศ. เป็นเลขไทย การพิมพ์ชื่อกษัตริย์ผู้สร้างวัดในเอกสารที่แตกต่างกัน เอกสารประวัติวัดไม่ได้ระบุชื่อผู้สร้างหรือปีที่สร้าง เป็นต้น ส่งผลให้การตัดคำในบางครั้งไม่ถูกต้อง หรือไม่สามารถสกัดคำตอบที่ต้องการได้ โดยการจะนำระบบการสกัดข้อมูลผู้สร้างวัดไปใช้กับข้อมูลชุดอื่นๆ ในอนาคต ผู้วิจัยจำเป็นต้องพิจารณาเทคนิคอื่นๆ เพื่อแก้ปัญหาในประเด็นเกี่ยวกับการเตรียมไฟล์ ส่วนในเรื่องของคำสำคัญเพื่อรองรับการสกัดข้อมูลผู้สร้างวัดให้ดีขึ้น ผู้วิจัยจำเป็นต้องปรับปรุงคำในพจนานุกรมให้ครอบคลุมยิ่งขึ้น

ในส่วนของพจนานุกรมถ้ามีการขยายชื่อกษัตริย์ในล้านนาและมีการเพิ่มคำศัพท์ที่เกี่ยวข้อง ผู้วิจัยจะนำฐานข้อมูลมาใช้แทนการจัดการด้วยไฟล์เพื่อให้เกิดความยืดหยุ่นและประสิทธิภาพการทำงานที่ดีขึ้น

สรุปผล

งานวิจัยที่จัดทำขึ้นนี้ทำให้ได้เทคนิคการสกัดสารสนเทศในการสกัดข้อมูลผู้สร้างวัดในสมัยล้านนาจากเอกสารประวัติวัดในรูปแบบของไฟล์ข้อความ โดยนำไปทดสอบกับเอกสารประวัติวัด ในเขตอำเภอเมืองเชียงใหม่ จำนวน 100 วัด ผลการทดสอบพบว่าความถูกต้องของการทำงานเท่ากับร้อยละ 97 ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าขั้นตอนวิธีเทคนิค

สารสนเทศที่นำเสนอในงานวิจัยนี้สามารถสกัดชื่อกษัตริย์ผู้สร้างวัด หรือช่วงเวลาสร้างวัด ในสมัยล้านนาจากเอกสารทางประวัติศาสตร์ได้อย่างมีประสิทธิภาพ สามารถนำไปพัฒนาต่อยอดในเรื่องการสืบค้นข้อมูลเชิงความหมายด้านประวัติศาสตร์สมัยล้านนาที่มีความเกี่ยวข้องกับข้อมูลเชิงเวลาและกษัตริย์ผู้สร้างวัดได้อย่างถูกต้องต่อไป

6. เอกสารอ้างอิง

- [1] National Statistical Office. *The survey on status of society and culture*. Available Online at <http://www.nso.go.th/>, accessed on 28 January 2018.
- [2] A. Wichiankhieo. *Study of the antiquities and ancient sites in Lanna temples of Northern Thailand*. Chaing mai: Sangsilp Printing. pp. 1 – 2, 2006.
- [3] S. Damrikun. “Decorative religious patterns illustrated in places in the Post-Lanna era in Chiang Mai.” *Arts and culture journal Chaingmai Rajabhat University*, Vol. 1, No. 1, pp. 1 – 27, 2016.
- [4] S. Ongsakul. *History of Lanna*. Silkworm Books, 2005.
- [5] National Office of Buddhism. *Summary report of the temple in Thailand 2017*. Available Online at <http://www.onab.go.th/>, accessed on 28 January 2018.
- [6] R. J. Mooney and R. Bunesco. “Mining Knowledge from Text Using Information Extraction.” *SIGKDD Explorations*, Vol. 7, No. 1, pp. 3 – 10, 2005.
- [7] J. Nothman, Nicky Ringland, Will Radford, Tara Murphy and James R. Curran. “Learning multilingual named entity recognition from Wikipedia.” *Artificial Intelligence*, Vol. 194, pp. 151 – 175, 2013.
- [8] S. Bird, E. Klein and E. Loper. *Information Extraction Architecture*. Available Online at <https://www.nltk.org/book/ch07.html>, accessed on 6 August 2018.
- [9] S. S. Sazali, N. A. Rahman and Z. A. Bakar. “Information Extraction: Evaluating Named Entity Recognition from Classical Malay Documents.” *Third International Conference on Information Retrieval and Knowledge Management*, pp. 48 – 53, 2016.
- [10] O. H. Zayed and S. R. EI-Beltagy. “Person Name Extraction from Modern Standard Arabic or Colloquial Text.” *The 8th International Conference on INFOrmatics and Systems (INFOS)*, pp. 44 – 48, 2012.
- [11] H. Li, X. Wu, Z. Li and G. Wu. “A relation extraction method of Chinese named entities based on location and semantic features.” *Springer Science Business Media*, pp. 1 – 15, 2013.
- [12] M. Y. Sharum, M. T. Abdullah, N. Sulaiman, M. A. Azmi Murad and Z. A. Zainon Hamzah. “Name Extraction for Unstructured Malay Text.” *IEEE Symposium on Computer & Information*, pp. 787 – 791, 2011.
- [13] T. Nakatoh, C. Yin, S. Hirokawa. “Extraction and Disambiguation of Name of Place from Tourism Blogs.” *First ACIS International Symposium on Software and Network Engineering*, pp. 73 – 78, 2011.
- [14] A. Aun-a-nan, Soontaree Kankaew, Naphong Wannapirun and Thaksina Khongsomlap. “The Thai News Analysis to Auto Tagging using Text Extraction.” *The 1st Rajamangala University of Technology Suvarnnabhumi National Conference*, pp. 427 – 434, 2016.
- [15] S. Tepdang. “A Comparative study of the use and non-use of prefixes and Suffix in named entity recognition.” *The 7th National Conference on Information Technology (NCIT)*, pp. 358 – 362, 2015.
- [16] P. Wikaha and P. Netisopakul. “A study of named entity recognition.” *KMITL Journal Information Technology*, Vol. 3, No 1, pp. 32 – 37, 2014.
- [17] National Electronics and Computer Technology Center (NECTEC). *A Thai Lexeme Tokenization and Normalization Tool*. Available Online at <http://www.sansarn.com/lexto/>, accessed on 26 January 2018.
- [18] Rangsiapan Marukatat. “Author Identification of Thai Online Messages Using Support Vector Machine and Decision Tree.” *The Journal of KMUTNB*, Vol. 25, No. 1, pp. 103 – 111, 2015.



- [19] C. Rattanapoka and M. Sutikun. "Automatic Tourist Attraction Categorization System using Thai Description via Web Service." *The Tenth National Conference on Computing and Information Technology (NCCIT2014)*, pp. 109 – 114, 2014.
- [20] Chiang Mai Provincial Office of Buddhism. *List of temples in Chiang Mai*. Available Online at <http://cmi.onab.go.th/>, accessed on 28 January 2018.
- [21] Chiang mai Culture Office. *Activities promoting art and cultural information through the use of QR-CODE*. Available online at <https://culture.chiangmai.in/>, accessed on 28 January 2018.
- [22] P. Chantaraj. *Database System Development of the Temple in Chiang Mai Province*. The Far Eastern University, 2011.