



วารสารบริหารธุรกิจเทคโนโลยีมหานคร

MUT Journal of Business Administration

ปีที่ 12 ฉบับที่ 1 (มกราคม - มิถุนายน 2558)

Volume 12 Number 1 (January – June 2015)

การวิเคราะห์ประสิทธิภาพของอัลกอริทึมเนมแมทช์สำหรับระบบสืบค้น ข้อมูลท่องเที่ยว โดยใช้ F-Measure

Performance Analysis of Name Matching Algorithm for Tourist Information Searching System Using F-Measure

จักรกฤษณ์ เสน่ห์ นมะหุต Chakkrit Snae Namahoot^{1,*}

¹ Ph.D., ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์
มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก ประเทศไทย

¹ Ph.D., Assistant Professor, Department of Computer Science and Information Technology, Faculty of Science
Naresuan University, Phitsanulok Province, Thailand

บทคัดย่อ

เมื่อนักท่องเที่ยวต้องการสืบค้นข้อมูลท่องเที่ยวส่วนมากจะทำการสืบค้นผ่านเสิร์ชเอนจินที่ได้รับคามนิยมอย่างสูง แต่ผลลัพธ์ที่ได้นั้นมีจำนวนมากอาจต้องเสียเวลาในการเข้าถึงข้อมูลจากหลายเว็บไซต์และผลลัพธ์ที่ได้อาจไม่ตรงกับความต้องการ อีกทั้งในบางเว็บไซต์นำเสนอข้อมูลบนเว็บไซต์ที่อาจมีการเขียนไม่ถูกต้องตามหลักภาษาอังกฤษก็จะทำให้ได้ผลของการสืบค้นไม่พบเว็บไซต์ที่เขียนไม่ถูกต้อง ดังนั้นผู้วิจัยจึงได้ประยุกต์ใช้อัลกอริทึมเนมแมทช์ซึ่งมาใช้ในการสืบค้นข้อมูลท่องเที่ยวเพื่อแก้ปัญหาในการกรองข้อมูลชื่อที่สามารถเขียนได้หลายรูปแบบหรือมีความคล้ายคลึงกันทั้งคำพ้องรูป พ้องเสียง หรือสะกดผิดในภาษาอังกฤษ ทำให้ข้อมูลท่องเที่ยว เช่น ชื่อของสถานที่ หรือชื่อของจังหวัด ที่เขียนผิดถูกค้นพบ อีกทั้งยังได้ทำการวัดประสิทธิภาพของอัลกอริทึมเนมแมทช์ โดยใช้ F-Measure ซึ่งจากผลการวิจัยพบว่า อัลกอริทึม Soudex มีประสิทธิภาพดีที่สุดในการสืบค้นข้อมูลท่องเที่ยวที่สามารถช่วยในการแก้ไขปัญหาการพิมพ์ผิดหรือไม่ถูกหลักภาษาอังกฤษซึ่งจะทำให้ไม่พลาดข้อมูลที่ต้องการ

คำสำคัญ: การสืบค้นข้อมูล, ข้อมูลท่องเที่ยว, ระบบสืบค้นข้อมูล, อัลกอริทึมเนมแมทช์, การวัดประสิทธิภาพ, รหัสเสียง

* E-mail address: chakkrits@nu.ac.th

ABSTRACT

To extract tourism information, most travelers rely on web search engines such as Google, so they are faced with a huge number of web pages as results. Consequently, they apply a time consuming process accessing many websites but might still not get what they need. Moreover, if keywords or website contents are misspelled in English or other languages, the search results will consist of incomplete or incorrect information. To deal with these problems, name matching algorithms are used to make connections between several similar forms in English words, such as homographs and misspellings. This helps discover mistakes in tourist information, such as names of locations and administrative entities. In order to measure the performance of the name matching algorithms, F-Measure is used. The results show that the Soundex algorithm has the best performance in tourist information searching and can solve misspellings of information with names in English language effectively and will not miss any important tourist information.

Keywords: Information Retrieval, Tourist Information, Information Searching System, Name Matching Algorithm, Performance Measurement, Sound Code

บทนำ

การท่องเที่ยวถือเป็นอุตสาหกรรมบริการที่ทำให้ประเทศไทยรายได้เพิ่มขึ้นและยังมีบทบาทต่อการพัฒนาเศรษฐกิจของชาติอีกด้วย ทั้งนี้ในปัจจุบันพบว่าข้อมูลท่องเที่ยวเป็นจำนวนมากที่ถูกนำเสนอบนอินเทอร์เน็ตและมีรูปแบบการนำเสนอที่ต่างกันอย่างออกไป เช่น รูปแบบข้อความ รูปแบบมัลติมีเดีย เป็นต้น ดังนั้นในการค้นหาข้อมูลท่องเที่ยวเพื่อให้ได้ตรงกับความต้องการของนักท่องเที่ยวจึงมีความสำคัญต่อนักท่องเที่ยวเป็นอย่างมาก โดยเฉพาะการป้อนคำค้นหาที่บางครั้งมีการเขียนผิดโดยไม่ตั้งใจหรือการเขียนหรือสะกดคำไม่ถูกต้องตามหลักภาษาไทยหรือภาษาอังกฤษ (Snae *et al.*, 2007) หรือมีการใช้คำค้นหาที่เป็นชื่อเรียกอื่นๆ ของสถานที่นั้นๆ เช่น วัดใหญ่ หมายถึง วัดพระศรีรัตนมหาธาตุวรมหาวิหาร สองแคว หมายถึง จังหวัดพิษณุโลก เป็นต้น ซึ่งทำให้ได้ผลลัพธ์ไม่ตรงกับความต้องการปะปนกันอยู่

การค้นหาข้อมูลท่องเที่ยวอาจมาจากหลายๆ แหล่ง เช่น สื่อสิ่งพิมพ์ แผ่นพับ อินเทอร์เน็ต บางครั้งมีการใช้คำค้นหาที่เป็นชื่อเรียกอื่นๆ ที่ใช้เรียกกันมาตั้งแต่สมัยโบราณของสถานที่นั้นๆ แต่ชื่อเหล่านี้หมายถึงสถานที่เดียวกัน โดยมีชื่อเรียกที่หลากหลายตามวัฒนธรรม พื้นเพ ภูมิสำเนา ลักษณะภูมิประเทศ เช่น วัดใหญ่ หมายถึง วัดพระศรีรัตนมหาธาตุวรมหาวิหาร เมืองสองแคว เมืองออกแตก

เมืองพุทธรินราช หมายถึง จังหวัดพิษณุโลก บางครั้งการค้นหามักจำเป็นต้องใช้ส่วนประกอบอื่นๆ เช่น คำขวัญ ดอกไม้ ต้นไม้ สัญลักษณ์ ประจำจังหวัดนั้นๆ ทำให้เกิดการค้นหาข้อมูลผิดพลาดไป (นฤพนธ์ และ จักรกฤษณ์, 2554) หรือ บางครั้งในการค้นหาข้อมูลท่องเที่ยวเกี่ยวกับจังหวัดนั้นๆ สถานที่ท่องเที่ยว อาจมีการใช้คำค้นหาที่มีการเขียนผิดหรือไม่ถูกต้องตามหลักภาษาไทยและภาษาอังกฤษซึ่งคำเหล่านั้นปรากฏอยู่มากมายในเว็บไซต์ท่องเที่ยว เช่น ค้นหาข้อมูลท่องเที่ยวในจังหวัดพิษณุโลก (Phitsanulok) แต่พบว่าบางเว็บไซต์ใช้คำว่า “พิงสะนุโลก” หรือ “พิษณุโลก” หรือ “พิตโลก” หรือ “พิตสะนุโลก” เป็นการเขียนชื่อจังหวัดพิษณุโลกไม่ถูกต้องตามหลักภาษาไทย นอกจากนี้พบว่าบางเว็บไซต์ใช้คำว่า “Phisanulok” หรือ “pitsanulok” หรือ “pisanulok” หรือ “phitsanurok” มีการเขียนชื่อจังหวัดพิษณุโลกไม่ถูกต้องตามหลักภาษาอังกฤษ (Snae and Brückner, 2007) โดยชื่อที่มีการสะกดผิดหรือเขียนผิดนั้นก็ย่อมสามารถก่อให้เกิดปัญหาในการค้นหาสถานที่ท่องเที่ยวได้ ซึ่งถ้าการบันทึกชื่อที่สะกดผิดหรือการเขียนที่แตกต่างกัน แม้เพียงเล็กน้อยจะสามารถนำไปสู่ความผิดพลาดในการระบุสถานที่ท่องเที่ยวเนื่องจากชื่อที่เขียนแตกต่างกันสองชื่อนั้นอาจบ่งชี้ถึงสถานที่เดียวกันก็ได้ โดย Snae and Brückner (2006) ได้แบ่งความคลาดเคลื่อนที่อาจเกิดขึ้นจากการค้นหาชื่อที่สามารถออกเสียงและการสะกดที่แตกต่างกันได้ 3 ประเภท ดังต่อไปนี้

1. ความคลาดเคลื่อนจากรูปแบบตัวอักษร ปัญหานี้เกิดจากการใช้คำขยายความและการย่อ, เครื่องหมายวรรคตอน, การเว้นวรรค, ตัวพิมพ์ใหญ่

2. ความคลาดเคลื่อนจากรูปแบบการสะกดคำ ความคลาดเคลื่อนนี้จะขึ้นอยู่กับการสันนิษฐานที่มาและชื่อเป้าหมายที่มีความแตกต่างกันเนื่องจากความผิดพลาดในการคัดลอกเช่นการออกเสียงที่ต่างออกไปรวมถึงการสะกดคำผิดและการพิมพ์ผิด สิ่งคลาดเคลื่อนเหล่านี้แม้จะไม่ได้ส่งผลถึงการออกเสียงแต่ก็ก่อให้เกิดปัญหาในการจับคู่ของชื่อได้

3. ความคลาดเคลื่อนจากรูปแบบการออกเสียง ความคลาดเคลื่อนจากการออกเสียงเกิดจากการที่โครงสร้างของชื่อถูกเปลี่ยนแปลงไปด้วยการออกเสียงเช่นการสะกดคำว่า “Pooh” ในภาษาอังกฤษแต่ในภาษาเยอรมันจะสะกดเป็น “Phu” แทน ซึ่งการออกเสียงที่เปลี่ยนแปลงไปนี้อาจเกิดจากการได้ยินผิดจึงทำให้โครงสร้างของชื่อมีการเปลี่ยนแปลงไป เช่น คำว่า “อยุธยา” และ “อโยธยา” เป็นชื่อที่เกี่ยวข้องกันแต่มีการออกเสียงแตกต่างกันเป็นอย่างมาก ซึ่งจริงๆ แล้วรูปแบบการออกเสียงของชื่อนั้นสามารถแสดงได้หลายรูปแบบ เช่น อาจแสดงเป็น “พิษณุโลก” หรืออีกแบบที่ได้ถูกย่อให้สั้นเป็น “พิตโลก” หรือ “ราชบุรี” เป็น “ราตรี” เป็นต้น

ในงานวิจัยนี้ได้ทำการวิเคราะห์ประสิทธิภาพของอัลกอริทึมเนมแมทซึ่งที่อยู่ในกลุ่มที่มีการเปลี่ยนชื่อเป็นรหัสเสียง ได้แก่ Soundex, Soundex2, Phonex, Metaphone, Simplex และ NYSIIS มาทำการทดสอบกับฐานข้อมูลชื่อจังหวัดในประเทศไทยจากที่ได้จากการสืบค้นข้อมูลจากกูเกิ้ล เพื่อนำมาใช้ในการแก้ปัญหาชื่อที่มีการเขียนผิดหรือไม่ถูกต้องตามหลักภาษาอังกฤษ เช่น ความคล้ายคลึงกันของตัวสะกดที่สามารถเขียนแทนกันได้ และคำพ้องรูป พ้องเสียง ทำให้ข้อมูลท่องเที่ยวที่เขียนผิดสามารถถูกค้นพบได้ในกูเกิ้ล อีกทั้งยังสามารถนำเอาอัลกอริทึมเนมแมทซึ่งที่มีประสิทธิภาพดีที่สุดมา

ทำการแก้ปัญหาเกี่ยวกับข้อมูลท่องเที่ยวที่มีการเขียนผิดในภาษาอังกฤษที่มีอยู่มากมายในเว็บไซต์ท่องเที่ยวต่างๆ เพื่อที่จะสามารถเพิ่มประสิทธิภาพของ ระบบสืบค้นข้อมูลท่องเที่ยวได้อีกด้วย

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ประสิทธิภาพของอัลกอริทึมเหมแมชซึ่งต่างๆ ที่นำมาใช้ในระบบสืบค้นข้อมูลท่องเที่ยว เพื่อนำมาแก้ปัญหาคำที่มีความหลากหลายและเขียนผิดหรือไม่ถูกต้องตามหลักภาษาอังกฤษ ด้วยหลักการ F-Measure
2. เพื่อประยุกต์ใช้อัลกอริทึมที่มีประสิทธิภาพที่ดีที่สุดที่ได้จากข้อที่ 1 มาทำการออกแบบและพัฒนาระบบสืบค้นข้อมูลท่องเที่ยวที่สามารถตอบสนองความต้องการในการค้นหาอย่างมีประสิทธิภาพ

ประโยชน์ของการวิจัย

1. ได้อัลกอริทึมเหมแมชซึ่งที่มีประสิทธิภาพดีที่สุดในการสืบค้นข้อมูลที่สามารถช่วยในการแก้ไขปัญหาการพิมพ์ผิดหรือไม่ถูกหลักภาษาอังกฤษซึ่งจะทำให้ไม่พลาดข้อมูลที่ต้องการและสามารถทำงานผ่านเว็บแอปพลิเคชันได้
2. ได้ระบบสืบค้นข้อมูลท่องเที่ยวด้วยเหมแมชซึ่งที่มีประสิทธิภาพที่สามารถช่วยอำนวยความสะดวกต่อนักท่องเที่ยว อีกทั้งยังช่วยส่งเสริมและพัฒนาตลาดการท่องเที่ยวและการบริการให้เพิ่มมากขึ้น
3. ได้ระบบค้นหาข้อมูลท่องเที่ยวที่มีประสิทธิภาพที่สามารถช่วย
 - 3.1 เพิ่มการส่งเสริมและพัฒนาตลาดการท่องเที่ยวและการบริการ
 - 3.2 ลดค่าใช้จ่ายในการแสวงหาข้อมูลที่ต้องการ และครบถ้วนของผู้ใช้บริการ
 - 3.3 ช่วยกระตุ้นการท่องเที่ยวได้มากขึ้น อีกทั้งยังเป็นการเพิ่มยอดค่าบริการต่างๆ ด้วย

การทบทวนวรรณกรรม

การใช้อินเทอร์เน็ตเพื่อการสืบค้นข้อมูลได้รับความนิยมอย่างแพร่หลาย เนื่องจากถือได้ว่าเป็นคลังข้อมูลที่ใหญ่ โดยเฉพาะข้อมูลทางด้านการท่องเที่ยวที่มีความสำคัญต่อนักท่องเที่ยวเป็นอย่างมาก โดยมีงานวิจัยที่ได้นำเทคโนโลยีและแนวคิดต่างๆ มาช่วยนำเสนอข้อมูลด้านการท่องเที่ยวให้มีประสิทธิภาพ ซึ่งในปัจจุบันมีเว็บแนะนำการท่องเที่ยวและเว็บที่ให้ข้อมูลการท่องเที่ยวอย่างมากมาย เช่น Web Application สำหรับให้บริการข้อมูลแหล่งท่องเที่ยว ที่พัก ร้านค้า แผนที่ และวางแผนการท่องเที่ยวแก่นักท่องเที่ยวหรือผู้ที่สนใจจะท่องเที่ยวยังสถานที่ต่างๆ ซึ่งสำหรับการบริการสืบค้นข้อมูลท่องเที่ยว แผนที่ยังสถานที่ท่องเที่ยวถือว่าเป็นข้อมูลท่องเที่ยวที่มีความสำคัญมากในการนำไปใช้แสดงผลการสืบค้นของระบบการให้บริการข้อมูลท่องเที่ยว โดยส่วนใหญ่จะอยู่ในรูปแบบระบบ

สารสนเทศภูมิศาสตร์ (Geographic Information System: GIS) ซึ่งจะได้เห็นได้จากงานวิจัยของ สุรางค์รัตน์ และ จักรกฤษณ์ (2554) ได้นำเสนอระบบบริหารจัดการข้อมูลและเชื่อมโยงเส้นทางแหล่งท่องเที่ยวในรูปแบบอนิเมชัน (Animation) เพื่อดึงดูดนักท่องเที่ยว ซึ่งใช้แผนที่ระบบสารสนเทศภูมิศาสตร์ในการเชื่อมโยงเส้นทางเดินทางไปยังแหล่งท่องเที่ยวและสถานที่ให้บริการต่างๆ ภายในเขตภาคเหนือ ประกอบด้วย จังหวัดพิษณุโลก สุโขทัย เพชรบูรณ์ และอุตรดิตถ์ โดยใช้ Google Map ในการนำเสนอแผนที่เชื่อมโยงเส้นทางแหล่งท่องเที่ยวรวมทั้งเสนอแหล่งท่องเที่ยวที่ใกล้เคียงเพื่อดึงดูดความสนใจของนักท่องเที่ยว ผลการทำงานทำให้วางแผนการเดินทางได้หลายทางเลือกตามความต้องการ สะดวกรวดเร็ว และยังใช้เป็นต้นแบบสำหรับภูมิภาคอื่นๆ ได้อีกด้วย ข้อดีของระบบนี้คือมีความยืดหยุ่นในการค้นหาข้อมูลเนื่องจากการใช้แผนที่จากกูเกิ้ล แต่อย่างไรก็ตามงานวิจัยนี้ไม่รองรับการสืบค้นของข้อมูลท่องเที่ยวที่อาจมีการเขียนผิดหรือมีรูปแบบการเขียนหรือเรียกที่หลากหลาย ซึ่งต่อมา Brückner *et al.* (2008) ได้ออกแบบระบบ ONTO-GIS โดยนำ Canonical Name Matching Algorithm มาช่วยในการแก้ปัญหาสำหรับค้นหาชื่อสถานที่ในแผนที่ระบบสารสนเทศทางภูมิศาสตร์ที่สามารถเขียนได้หลากหลายรูปแบบหรือคำที่พ้องรูป พ้องเสียง และใช้หลักการออนโทโลยีเพื่อสนับสนุนการค้นหาในภาษาไทย ภาษาอังกฤษ และภาษาญี่ปุ่น แล้วใช้ Google Map มาช่วยแสดงแผนที่

Park *et al.* (2009) ได้นำหลักการออนโทโลยีมาใช้ในการออกแบบระบบข้อมูลนักท่องเที่ยวอย่างชาญฉลาด เนื่องจากความก้าวหน้าของระบบเครือข่ายอินเทอร์เน็ตทำให้ผู้ใช้ผ่านเว็บและมีสื่อเพิ่มมากขึ้น โดยโดเมนออนโทโลยีประกอบไปด้วยแนวคิดสำหรับศูนย์บริการนักท่องเที่ยว สถานที่ท่องเที่ยวและออนโทโลยีที่ใช้ในระบบ ซึ่งระบบประกอบไปด้วยระบบศูนย์บริการนักท่องเที่ยวที่เก็บออนโทโลยีและฐานข้อมูลแล้วใช้การเชื่อมต่อเชิงความหมายระหว่างออนโทโลยีกับฐานข้อมูล ระบบบริการแผนที่การจราจรแบบให้ข้อมูลแบบทันที ระบบจัดการข้อมูลนักท่องเที่ยว ซึ่งมีการติดต่อกับผู้ใช้แบบเมนูกราฟิกที่สามารถใช้ทั้งเครือข่ายมีสายและไร้สายได้ รวมถึงโทรศัพท์มือถือและ PDA อีกทั้งยังใช้ RFID ในการบริการข้อมูลแหล่งท่องเที่ยวให้กับนักท่องเที่ยวอีกด้วย แต่ยังคงพบว่าระบบนี้ไม่รองรับการค้นหาสถานที่ท่องเที่ยวต่างๆ และไม่มีระบบวางแผนและแนะนำการท่องเที่ยว ซึ่งเป็นเพียงศูนย์บริการข้อมูลนักท่องเที่ยวเท่านั้น ทำให้ไม่ครอบคลุมความต้องการของนักท่องเที่ยวได้ทั้งหมด ซึ่งระบบบริการวางแผนการท่องเที่ยวนั้นถือว่าเป็นเครื่องช่วยอำนวยความสะดวก ทำให้นักท่องเที่ยวสามารถลดเวลาการสืบค้นข้อมูลท่องเที่ยวและช่วยในการตัดสินใจท่องเที่ยวได้อีกด้วย ซึ่งระบบวางแผนการท่องเที่ยวนี้มีทั้งที่แสดงผลแบบ Static และแบบ Dynamic เช่น งานวิจัยของ Castillo *et al.* (2008) ได้นำเสนอระบบที่ช่วยวางแผนการท่องเที่ยวให้กับนักท่องเที่ยวแบบไดนามิก โดยสามารถนำพฤติกรรมการท่องเที่ยวในอดีตและลักษณะการเที่ยวที่คล้ายกันของนักท่องเที่ยวมาวิเคราะห์ด้วยหลักการ Case-Based Reasoning อีกทั้งระบบนี้ยังได้นำหลักการออนโทโลยีมาใช้ในการจัดเก็บข้อมูลนักท่องเที่ยว สถานที่ท่องเที่ยว พฤติกรรมการท่องเที่ยว กิจกรรมการท่องเที่ยว การเดินทาง รวมถึงการเดินทางจากที่หนึ่งไปยังอีกที่ด้วยการเดินทาง โดยระบบการวางแผนท่องเที่ยวนี้สามารถรองรับอุปกรณ์พกพาหรือโทรศัพท์มือถือพร้อมส่งออกแผนที่การเดินทางไปยังอุปกรณ์ดังกล่าวในรูปแบบไฟล์ JPG หรือ PDF โดยใช้ A* Algorithm ในการแนะนำเส้นทาง แต่พบว่ามีข้อเสียคือพบปัญหาในการท่องเที่ยวแบบระบุสถานที่เฉพาะอาจจะไม่สามารถค้นเจอถ้าพิมพ์หรือสะกดไม่ถูกต้อง ซึ่งในงานวิจัยของ

Kongthong *et al.* (2011) ได้พัฒนาระบบสืบค้นข้อมูลท่องเที่ยวที่มีความยืดหยุ่นกว่า เนื่องจากสามารถให้ผู้ใช้อินเตอร์เฟซที่เป็นประโยคเพื่อสอบถามข้อมูลตามความต้องการในการสืบค้นข้อมูลท่องเที่ยวในประเทศไทยได้ เช่น “รบกวนแนะนำที่พักที่เกาะกูดหน่อยครับไปกัน 10 คน” ซึ่งระบบจะมีกระบวนการตัดคำภาษาไทยโดยเทียบจากพจนานุกรมและจัดกลุ่มของคำที่ตัดได้ให้เข้ารูปแบบโครงสร้างของการจัดเก็บข้อมูลในออนโทโลยีท่องเที่ยวเพื่อส่งรูปแบบโครงสร้างคำเหล่านี้ไปสืบค้นข้อมูลท่องเที่ยวที่ถูกเก็บไว้ในออนโทโลยีท่องเที่ยวและแสดงผลลัพธ์ตามที่ใช้ได้ทำการสอบถาม แต่ระบบนี้รองรับแค่ภาษาไทยเท่านั้น เช่นเดียวกันกับงานวิจัยของ Cao *et al.* (2011) และ Cao and Nguyen (2012) ได้พัฒนาระบบ STAAR (Semantic Tourist Information Access and Recommending) สำหรับค้นหาและแนะนำข้อมูลท่องเที่ยวจากการป้อนคำถามแต่ใช้ได้เฉพาะภาษาอังกฤษเท่านั้น โดยใช้ออนโทโลยีในการเก็บข้อมูลท่องเที่ยวและสามารถประมวลผลผ่านโทรศัพท์มือถือด้วยเทคนิคการประมวลผลภาษาธรรมชาติโดยใช้กระบวนการตัดคำและกระบวนการเทียบจากพจนานุกรมหรือคลังคำศัพท์มาใช้ในการสืบค้นข้อมูลตามโครงสร้างคำถามที่ผู้ใช้ต้องการ

จากที่ได้กล่าวมาจะเห็นได้ว่าระบบการสืบค้นข้อมูลท่องเที่ยวโดยทั่วไปยังไม่รองรับการค้นหาในกรณีที่มีพจน์ชื่อผิดหรือสะกดไม่ถูกต้องตามหลักภาษาไทยหรือภาษาอังกฤษ ทำให้ได้ผลลัพธ์ของการค้นหาไม่ตรงกับความต้องการหรือไม่พบสถานที่ที่ต้องการค้นหา อีกทั้งผลลัพธ์ที่ได้ไม่ได้ถูกจัดหมวดหมู่ไว้อย่างเหมาะสม ทำให้เสียเวลาในการคัดเลือกข้อมูลท่องเที่ยวเพื่อให้ได้ข้อมูลท่องเที่ยวหรือกิจกรรมท่องเที่ยวที่น่าสนใจในช่วงเวลานั้นๆ ที่ตรงกับความต้องการทั้งหมดและยังทำให้เกิดความไม่สะดวกในการเข้าถึงข้อมูลท่องเที่ยวในเว็บไซต์เดียว (Panawong *et al.*, 2014) ซึ่งการที่นักท่องเที่ยวจะเที่ยวให้ครบทุกที่ในเวลาอันจำกัดย่อมเป็นไปได้ยาก เพราะแต่ละที่ใช้เวลาในการท่องเที่ยวซ้ำเร็วไม่เท่ากันขึ้นอยู่กับความสนใจของแต่ละคน และบางสถานที่ท่องเที่ยวยังมีเวลาเปิด ปิดแตกต่างกัน ทำให้นักท่องเที่ยวจะต้องค้นหาข้อมูลเพื่อที่จะวางแผนไว้ล่วงหน้าหรือเลือกเฉพาะสถานที่ท่องเที่ยวที่น่าสนใจเท่านั้น (Sylejmani and Dika, 2011) แม้เป็นการท่องเที่ยวในช่วงเวลาสั้นๆ แต่นักท่องเที่ยวจะค้นหาข้อมูลท่องเที่ยวจากแหล่งข้อมูลที่ต่างกันไปหรือโทรศัพท์สอบถามข้อมูลเส้นทางของสถานที่ท่องเที่ยวต่างๆ ภายในจังหวัดนั้นๆ ซึ่งจะใช้เวลามากในการสืบค้นข้อมูลในการตัดสินใจท่องเที่ยว (Maervoet *et al.*, 2008) ทำให้ นฤพนธ์ และ จักรกฤษณ์ (2553) ได้พัฒนาระบบสืบค้นข้อมูลออนโทโลยีท่องเที่ยวโดยใช้ Variation Name Matching Algorithm ในการแก้ปัญหาการเขียนชื่อสถานที่หรือคำที่มีการสะกดคล้ายคลึงกันในเว็บไซต์ต่างๆ หรือที่เกี่ยวข้องสามารถค้นหาเจอได้ครบถ้วนมากที่สุด ซึ่งข้อดีของการใช้ Variation Name Matching Algorithm นั้นจะทำให้คำที่เขียนหรือสะกดผิดในเว็บไซต์จะถูกค้นหาเจอในเว็บไซต์เดียว แต่อย่างไรก็ตาม วิธีการนี้ทำให้เสียเวลาอย่างมากในกระบวนการค้นหา เนื่องจากมีข้อมูลที่ไม่เกี่ยวข้องของคำที่มีการสะกดผิดพบอยู่เป็นจำนวนมากทำให้ (นฤพนธ์ และ จักรกฤษณ์ 2556) ได้ประยุกต์ใช้อัลกอริทึม ISG ซึ่งเป็น Name Matching ที่ใช้ในการหาความคล้ายคลึงของตัวอักษรของชื่อ โดย ISG จะทำการกรองข้อมูลที่ไม่เกี่ยวข้องเหล่านั้นก่อนนำไปทำการสืบค้นหาข้อมูลในกูเกิ้ล โดยใช้เทคนิค Web Crawler เพื่อช่วยลดเวลาในการสืบค้นข้อมูลท่องเที่ยวจากชื่อสถานที่ท่องเที่ยวต่างๆ ได้อย่างมีประสิทธิภาพซึ่งรองรับทั้งภาษาไทยและภาษาอังกฤษ อย่างไรก็ตามผลจากการวัดประสิทธิภาพของการสืบค้นข้อมูลท่องเที่ยวที่มีความหลากหลายนั้น พบว่า ระบบค้นหาชื่อแหล่ง

ท่องเที่ยวทั้งภาษาไทยและภาษาอังกฤษมีค่าประสิทธิภาพใกล้เคียงกันซึ่งอยู่ในเกณฑ์พอใช้ คิดเป็น 71.21 % และ 70.49 % ตามลำดับ และไม่ได้มีการนำเอา อัลกอริทึมเนมแมชชิงอื่นๆ มาใช้และเปรียบเทียบกันในระบบสืบค้นข้อมูลท่องเที่ยว

กรอบแนวคิด

กรอบแนวคิดในการวิเคราะห์ประสิทธิภาพของอัลกอริทึมเนมแมชชิงที่สามารถแบ่งออกเป็น 2 ประเด็นหลัก ดังต่อไปนี้

1. ขั้นตอนการดำเนินงาน
2. อัลกอริทึมเนมแมชชิง

ขั้นตอนการดำเนินงาน

การวิเคราะห์ประสิทธิภาพและอัลกอริทึมเนมแมชชิงมีขั้นตอนการดำเนินงานดังนี้

1. พัฒนาอัลกอริทึมเนมแมชชิงของกลุ่มที่มีการเปลี่ยนชื่อเป็นรหัสเสียง
2. ทดสอบอัลกอริทึมเนมแมชชิงและทำการวัดประสิทธิภาพของการสืบค้นของอัลกอริทึมเนมแมชชิง โดยใช้วิธีการ F-Measure โดยใช้ฐานข้อมูลชื่อจังหวัดที่มีการเขียนผิดหรือสะกดผิดในกูเกิ้ล ซึ่งมีสูตรการคำนวณประสิทธิภาพมีดังต่อไปนี้

$$f = 2 * \left(\frac{P * R}{P + R} \right) \quad (1)$$

$$P = \frac{|TP|}{(|TP| + |FP|)} \quad (2)$$

$$R = \frac{|TP|}{(|TP| + |FN|)} \quad (3)$$

โดยที่ f	คือ	ค่า f -score
P	คือ	ค่า $Precision$
R	คือ	ค่า $Recall$
TP	คือ	จำนวนชื่อที่ควรจะ Match และ Match (True Positive)
FP	คือ	จำนวนชื่อที่ควรจะไม่ Match แต่ Match (False Positive)
FN	คือ	จำนวนชื่อที่ควรจะ Match แต่ไม่ Match (False Negative)

อัลกอริทึมเนมแมชชิง

อัลกอริทึมเนมแมชชิงเป็นอัลกอริทึมที่ใช้ในการเปรียบเทียบชื่อสองชื่อว่ามีความคล้ายกันหรือไม่ (Match/No Match) ซึ่งได้มีการแบ่งชนิดของ Name Matching Algorithm ออกเป็น 4 ประเภท (Snae and Brückner, 2006) ได้แก่

1. **กฎพื้นฐานของตัวสะกด (Spelling/String Analysis based Algorithms)** ใช้ในการเปรียบเทียบความคล้ายคลึงกันระหว่างตัวอักษรของชื่อ 2 ชื่อ ว่ามีหลักการเขียนเหมือนกันอย่างไร และผลของการเปรียบเทียบนี้จะออกมาในรูปว่าคล้ายกันหรือไม่ (Match/No Match) ซึ่งอัลกอริทึมประเภทนี้ ได้แก่ Guth และ Levenshtein

2. **กฎพื้นฐานของการออกเสียง (Phonetic/Sound base Algorithms)** เป็นการใช้โครงสร้างของเสียงเป็นตัวเปรียบเทียบค่าความคล้ายคลึงกัน ซึ่งใช้ในการแก้ปัญหาชื่อที่มีเสียงเดียวกันแต่เขียนหรือสะกดต่างกัน โดยใช้วิธีการจัดกลุ่มตัวอักษรจัดให้เสียงเดียวกันอยู่ในกลุ่มตัวอักษรเดียวกัน ตัวอย่างเช่น ตัวอักษร Ph ใช้จัดอยู่ในกลุ่มตัวอักษร F หรือในภาษาไทย “จ ท ฐ ต ฎ” ให้จัดอยู่ในกลุ่มแมกด (ด) อาทิเช่น คำว่า กาจ กับ กาด จะออกเสียงเหมือนกัน ซึ่งอัลกอริทึมประเภทนี้ ได้แก่ Soundex, Soundex2, Metaphone, NYSIIS และ Phonex

3. **วิธีการรวมกัน (Composite Methods)** เป็นวิธีการรวมกันระหว่างกฎพื้นฐานของตัวสะกด (Spelling/String Analysis based Algorithms) และกฎพื้นฐานของการออกเสียง (Phonetic/Sound base Algorithms) ซึ่งการรวม 2 วิธีนี้เป็นการแก้ปัญหาของชื่อเกี่ยวกับคำพ้องรูปพ้องเสียงหรือคำที่เขียนเหมือนกันแต่ออกเสียงต่างกันและคำที่เขียนต่างกันแต่ออกเสียงเหมือนกัน ซึ่งอัลกอริทึมประเภทนี้ ได้แก่ Simplex และ ISG

4. **วิธีการไฮบริด (Hybrid Approaches)** เป็นการใช้ค่าความน่าจะเป็นในการวัดค่าความคล้ายคลึงของชื่อ (Snae and Diaz, 2002) ซึ่งอัลกอริทึมประเภทนี้ ได้แก่ LIG1, LIG2, LIG3 (ย่อมาจาก Levenshtein Index of Similarity and Guth)

ในงานวิจัยนี้ได้ทำการวัดประสิทธิภาพของอัลกอริทึมเนมแมชชิงในกลุ่มที่มีการเปลี่ยนชื่อเป็นรหัสเสียงเท่านั้นซึ่งเนมแมชชิงอัลกอริทึมในกลุ่มนี้เป็นการเปรียบเทียบระหว่างรหัสเสียงของชื่อ โดยการนำรหัสเสียงของชื่อ 2 ชื่อมาเปรียบเทียบกันโดยตรง ถ้าเหมือนกันแสดงว่าชื่อทั้ง 2 ชื่อ Match กัน แต่ถ้าไม่เหมือนกันแสดงว่าชื่อทั้ง 2 ชื่อ No Match กัน ซึ่งเนมแมชชิงอัลกอริทึมในกลุ่มนี้มี 6 อัลกอริทึม ได้แก่

1. **Soundex** Soundex Algorithm ถูกออกแบบมาใช้สำหรับชื่อภาษาอังกฤษ (Winchester, 1970) Soundex Algorithm ได้ถูกนำไปใช้ เพื่อลดจำนวนระเบียบของการแต่งงานโดยการเข้ารหัสทั้งนามสกุลของสามีและนามสกุลเดิมของภรรยา และยังมีการนำอัลกอริทึมนี้ ไปใช้ในการเชื่อมโยงประชากรอีกด้วย แต่อย่างไรก็ตาม จะพบว่าการละเว้นการออกเสียงอาจจะทำให้ข้อมูลที่สมควร Match กลับได้ผลเป็นไม่ Match และข้อมูลที่ ไม่ Match กลับให้ผลที่ Match กัน (Davey and Jarvis, 1990)

ซึ่งข้อเสียที่พบจากการใช้ Soundex Algorithm คือ ไม่เหมาะกับชื่อที่สั้น หรือชื่อที่มีจำนวนสระมาก (Gill *et al.*, 1993)

Soundex Algorithm จะทำการเปลี่ยนชื่อเป็นรหัสของ Soundex ซึ่งประกอบด้วย 1 ตัวอักษร และ 3 ตัวเลข โดยมีขั้นตอนการเปลี่ยนดังต่อไปนี้

- 1) เก็บตัวอักษรตัวแรกของ Input ชื่อไว้
- 2) ถ้าตัวอักษร A, E, H, I, O, U, W, Y ไม่ใช่ตัวอักษรตัวแรกให้ตัดออกไป
- 3) จากข้อที่ 1 เปลี่ยนอักษรที่เหลือเป็นรหัส Soundex ดังแสดงในตารางที่ 1
- 4) ผลลัพธ์ของรหัส Soundex ต้องอยู่ในรูปของตัวอักษร 1 ตัวและตามด้วยตัวเลข 3 ตัว
- 5) จากข้อ 4 ถ้าผลลัพธ์ของรหัส Soundex น้อยกว่า 4 ตัวรหัส Soundex ให้ใส่ศูนย์แทนจนครบรหัส แต่ถ้ารหัส Soundex มากกว่า 4 รหัส ให้ตัดออก

ตารางที่ 1 : การเปลี่ยนอักษรเป็นรหัส Soundex

Letter	Number is coded to
B, P, F and V	1
C, G, J, K, Q, S, X and Z	2
D and T	3
L	4
M and N	5
R	6

2. **Soundex 2** ได้ปรับปรุงมาจากอัลกอริทึม Soundex ซึ่งจะขั้นตอนเหมือนกับ Soundex Algorithm แต่จะมีการเปลี่ยนเพียงขั้นตอนที่ 3 โดยรหัสของ Soundex 2 สามารถแสดงในตารางที่ 2

ตารางที่ 2 : การเปลี่ยนอักษรเป็นรหัส Soundex 2

Letter	Number is coded to
B and P	1
F and V	2
C, K and S	3
G and J	4
Q, X and Z	5
D and T	6
L	7
M and N	8
R	9

3. Metaphone ส่วนใหญ่จะใช้แก้ปัญหาเกี่ยวกับชื่ออังกฤษ โดยอัลกอริทึมมีขั้นตอนดังต่อไปนี้

- 1) ถ้ามีตัวเลขให้ตัดตัวเลขออก แล้วเปลี่ยนอักษรทุกตัวให้เป็นตัวใหญ่
- 2) ถ้า KN หรือ GN หรือ PM หรือ AE หรือ WR เป็นตัวแรกให้วางไว้ตำแหน่งแรก
- 3) ถ้า X เป็นตัวอักษรแรกให้เปลี่ยนเป็น S
- 4) ถ้า WH เป็นตัวอักษรแรกให้เปลี่ยนเป็น W
- 5) ถ้ามีอักษรที่เหมือนกันไม่ต้องเพิ่มรหัส
- 6) ถ้าตัวอักษร A, E, I, O, U และ Y ไม่ใช่ตัวอักษรตัวแรกให้ตัดออกไป
- 7) เปลี่ยนตัวอักษรให้เป็นรหัสของ Metaphone ดังแสดงในตารางที่ 3

ตารางที่ 3 : การเปลี่ยนตัวอักษรเป็นรหัส Metaphone

ตัวอักษร	รหัส	เงื่อนไข
B	B	ยกเว้นถ้าอยู่ตำแหน่งสุดท้ายของคำหลังจากตัวอักษร "m"
C	X	(เสียง 'sh') ถ้า "-cia-" หรือ "-ch-"
	S	เป็น "-ci-", "-ce-" หรือ "-cy-"
		ถ้าไม่ออกเสียง "-sci-", "-sce-" หรือ "-scy-"
	K	มีฉะนั้น, รวมทั้ง "-sch-"
D	J	ถ้าอยู่ใน "-dge-", "-dgy-", "-dgi"
	T	นอกนั้นเป็น
F	F	
G		ไม่ออกเสียงถ้ามันอยู่ใน "-gh-" และไม่อยู่ในตำแหน่งสุดท้าย หรือก่อนหน้าสระ
		ใน "-gn" หรือ "-gned"
		ใน "-dge-", อื่นๆ., ที่เหนือกว่ากฎ
	J	ถ้า "i", หรือ "e", หรือ "y" ก่อน, ถ้าไม่อยู่เป็นคู่ติดกัน "gg"
	K	นอกนั้นเป็น
H		ไม่ออกเสียงถ้าอยู่หลังสระและไม่มีสระตามมา
	H	นอกนั้นเป็น
J	J	
K		ถ้าไม่ออกเสียงใน "c"
	K	นอกนั้นเป็น
L	L	
M	M	
N	N	

ตารางที่ 3 : (ต่อ)

ตัวอักษร	รหัส	เงื่อนไข
P	F	ถ้า "h" ก่อน
	P	นอกนั้นเป็น
Q	K	
R	R	
S	X	(sh) ถ้า "h" ก่อน หรือใน "-sio-" หรือ "-sia-"
	S	นอกนั้นเป็น
T	X	(sh) ถ้า "h" ก่อน หรือใน "-sio-" หรือ "-sia-"
	0 (Zero)	(th) ถ้า "h" ก่อน
		ถ้าไม่ออกเสียงใน "tch-"
	T	นอกนั้นเป็น
V	F	
W		ไม่ออกเสียงถ้าไม่ถูกตามโดยสระ
	W	ถ้าถูกตามโดยสระ
X	KS	
Y		ไม่ออกเสียงถ้าไม่ถูกตามโดยสระ
	Y	ถ้าถูกตามโดยสระ
Z	S	

4. Phonex เกิดจากการรวมกันของสองกระบวนการ คือ Soundex และ Metaphone ซึ่งกระบวนการนี้ พบว่า สามารถทำงานได้ดีเมื่อนำไปใช้กับชื่อในภาษาอังกฤษ (Lait and Randell, 1998) โดยขั้นตอนของ Phonex ในการแปลงชื่อให้เป็นรหัส 4 ตัวอักษร มีดังต่อไปนี้

- 1) ลบตัวอักษร S ที่ต่อท้ายออก
- 2) ถ้ามี KN หรือ WR ขึ้นต้นให้ตัดตัวอักษรแรกทิ้งไป
- 3) ถ้ามีตัวอักษรดังต่อไปนี้ขึ้นต้นให้เปลี่ยนดังนี้
 - 3.1) ตัวอักษร H ให้ตัดทิ้ง
 - 3.2) ตัวอักษร E I O U Y ให้เปลี่ยนเป็น A
 - 3.3) ตัวอักษร K และ Q ให้เปลี่ยนเป็น C
 - 3.4) ตัวอักษร P ให้เปลี่ยนเป็น B
 - 3.5) ตัวอักษร J ให้เปลี่ยนเป็น G
 - 3.6) ตัว PH และ V ให้เปลี่ยนเป็น F
 - 3.7) ตัว Z ให้เปลี่ยนเป็น S
- 4) ให้เก็บตัวอักษรตัวแรกไว้แล้วลบตัวอักษร A, E, H, I, O, U, W, Y ในตำแหน่งอื่นออก
- 5) กำหนดหมายเลขดังต่อไปนี้ให้ตัวอักษรที่เหลือหลังตัวแรก

- 6) B, F, P, V เปลี่ยนเป็น 1
- 7) C, G, J, K, Q, S, X, Z เปลี่ยนเป็น 2
- 8) D, T เปลี่ยนเป็น 3 ถ้าไม่ตามด้วย C
- 9) L เปลี่ยนเป็น 4 ถ้าไม่ตามด้วยสระและเป็นตัวอักษรสุดท้าย
- 10) M, N เปลี่ยนเป็น 5 ข้ามไปพิจารณาตัวอักษรถัดไปถ้าตัวอักษรถัดไปเป็น D หรือ G
- 11) R เปลี่ยนเป็น 6 ถ้าไม่ตามด้วยสระและเป็นตัวอักษรสุดท้าย
- 12) ข้ามตัวอักษรปัจจุบันถ้ามีหมายเลขเดียวกับตัวอักษรก่อนหน้า
- 13) รหัสของ Phonex ต้องมีลักษณะ คือ ตัวอักษรตามด้วยตัวเลข 3 ตัว
- 14) ถ้ารหัสมีตัวเลขไม่ถึง 3 ตัวให้เติมด้วย 0 จนครบ และถ้ามีตัวเลขครบ 3 ตัวแล้วยังมีตัวอักษรที่ยังไม่ได้เปลี่ยนให้เปลี่ยนที่ตัวเลขสุดท้าย

5. NYSIIS เป็นวิธีการทางตัวอักษรที่ง่ายและให้ผลลัพธ์อัลกอริทึมเหมือนกับ Soundex อย่างไรก็ตาม NYSIIS ต่างจาก Soundex ตรงที่มันเก็บข้อมูลเกี่ยวกับตำแหน่งของสระในการเข้ารหัสของคำโดยแปลงสระทั้งหมดให้เป็นตัวอักษร A (Gill *et al.*, 1993) โดยวิธีการเข้ารหัสเบื้องต้นของ NYSIIS มีดังนี้

- 1) ใช้ตัวอักษรตัวแรกของชื่อมาเป็นรหัสตัวแรกของ NYSIIS
- 2) ลบสระที่ซ้ำกันให้เหลือเพียงหนึ่งตัว
- 3) ถ้าชื่อประกอบด้วยสระครบทุกตัว (AEIOU) ให้ตัดสระทิ้ง
- 4) แทนค่าสระด้วยตัว A
- 5) ถ้าตัวอักษรตัวสุดท้ายเป็นสระให้ตัดทิ้ง
- 6) ตัวอักษรที่เหลือให้เปลี่ยนเป็นรหัส NYSIIS

6. Simplex เป็นวิธีที่มีการปรับปรุงจากวิธีของ NYSIIS และ Soundex Algorithm ซึ่งจะเพิ่มจำนวนการ Match มากขึ้นจากเดิม โดยการแก้ไขส่วนใหญ่เน้นไปที่การแก้ไขตัวอักษร การเข้ารหัสของ Simplex มีพื้นฐานมาจาก NYSIIS โดยการเปลี่ยนสระ "AIEOUY" เป็น A ทั้งหมด

- 1) ทำตามขั้นตอนข้อ 1-6 ของ NYSIIS Algorithm
- 2) เข้ารหัสตัวอักษรตัวสุดท้าย ดังต่อไปนี้
 - 2.1) เปลี่ยน X เป็น C
 - 2.2) เปลี่ยน V เป็น F
 - 2.3) เปลี่ยน SE, ES, AS, CE เป็น S
 - 2.4) เปลี่ยน YE, EE, IE, EY, EI เป็น Y
 - 2.5) เปลี่ยน NT, ND เป็น N
- 3) เข้ารหัสตัวอักษรตัวอื่นๆ ดังต่อไปนี้
 - 3.1) เปลี่ยน PH เป็น F
 - 3.2) เปลี่ยน JH เป็น J

- 3.3) เปลี่ยน HN เป็น N
- 3.4) เปลี่ยน TH เป็น T
- 3.5) เปลี่ยน RH, RM เป็น R
- 3.6) เปลี่ยน WR เป็น R
- 3.7) เปลี่ยน WM, WN, M เป็น N
- 4) ถ้ามี K ตามด้วย N ให้เปลี่ยนเป็น N แต่ถ้า K ตามด้วยตัวอักษรอื่นที่ไม่ใช่ N ให้เปลี่ยนเป็นตัว C
- 5) ถ้ามี Z เป็นตัวอักษรตัวสุดท้ายให้เปลี่ยน Z เป็น S แต่ถ้าไม่ใช่ Z เป็นตัวสุดท้าย ให้คงตัวอักษรเดิมไว้
- 6) ถ้ามี Y เป็นทั้งตัวอักษรตัวแรกและตัวสุดท้าย ให้คงเดิม แต่ถ้าเป็นตัวอักษรอื่นให้เปลี่ยนเป็นตัว A

ระเบียบวิธีการวิจัย

ขั้นตอนการวิเคราะห์ประสิทธิภาพของอัลกอริทึมเนมแมชซึ่งแบ่งออกเป็น 3 ขั้นตอนหลักดังต่อไปนี้

1. การสร้างคำค้นหาหรือชื่อของจังหวัดที่สามารถเขียนได้หลากหลายรูปแบบในภาษาอังกฤษ
2. คัดกรองและจัดเก็บข้อมูลชื่อเพื่อใช้ในการวัดประสิทธิภาพของอัลกอริทึมเนมแมชซึ่ง
3. วัดประสิทธิภาพเนมแมชซึ่งอัลกอริทึมด้วยวิธีการ F-Measure

การสร้างคำค้นหาหรือชื่อของจังหวัด

ทำการป้อนชื่อจังหวัดที่เขียนถูกต้องเป็นภาษาอังกฤษ จากนั้นทำการสร้างชื่อที่สามารถเขียนได้หลากหลายหรือคล้ายคลึงกันกับชื่อที่เขียนถูกต้องโดยใช้อัลกอริทึม Name Variation Finding ซึ่งจะทำการสร้างชื่อจากความคล้ายคลึงของพยัญชนะและตัวสะกด (Snae *et al.*, 2007; Snae and Brückner, 2009) ดังแสดงตัวอย่างตามตารางที่ 4

ตารางที่ 4 : การทำงานของอัลกอริทึมของ Name Variation Finding สำหรับการสร้างคำคล้ายคลึงของ phitsanulok

คำที่ใช้	ตำแหน่งตัวอักษร	ตัวอักษรที่เปลี่ยน	เปลี่ยนเป็น	คำที่ได้
Phitsanulok	4	t	ตัดออก	Phisanulok
Phitsanulok	2	h	ตัดออก	Pitsanulok
Phitsanulok	2,4	h,t	ตัดออก	Pisanulok
Phitsanulok	9	l	r	Phitsanurok
Phitsanulok	4	t	ตัด t ออก เพิ่ม k	Phisanuklok
Phitsanulok	4,5	t,s	ตัด t ออก เปลี่ยน s-> z	Phizanulok
Phitsanulok	8	u	oo	Phitsanoolok
Phitsanulok	8	u	ok	phitsanoklok

จากตารางที่ 4 จะเห็นได้ว่าคำว่า Phitsanulok สามารถเขียนได้หลากหลายรูปแบบ เช่น Phisanulok (t ถูกตัดออกในตำแหน่งที่ 4) หรือ Pitsanulok (h ตำแหน่งที่ 2 ถูกตัดออก) เป็นต้น ซึ่งในงานวิจัยนี้จะใช้ตัวอย่างชื่อที่เขียนเป็นคำภาษาอังกฤษของจังหวัดในประเทศไทย 20 จังหวัด แต่ละจังหวัดประกอบด้วยชื่อที่สามารถเขียนได้หลากหลายที่ค้นเจอในกูเกิลและมีข้อมูลท่องเที่ยวปรากฏอยู่จริง จำนวน 100 ชื่อ (2,000 ชื่อ) มาใช้ในการวัดประสิทธิภาพของอัลกอริทึมที่แนะนำ

คัดกรองและจัดเก็บข้อมูลชื่อ

นำชื่อที่เขียนได้หลายรูปแบบทั้ง 20 จังหวัด ที่ได้จาก Name Variation Finding ไปทำการค้นหาข้อมูลท่องเที่ยวในกูเกิลและทำการคัดกรองเฉพาะชื่อแต่ละจังหวัดที่มีข้อมูลท่องเที่ยวปรากฏอยู่จริง (ตารางที่ 5) ที่ค้นเจอในกูเกิล จำนวน 100 ชื่อ (2,000 ชื่อ) และจัดเก็บชื่อจังหวัดและชื่อที่มีความคล้ายคลึงกันเหล่านี้นี้ลงฐานข้อมูล

ตารางที่ 5 : ตัวอย่างจำนวนรายการที่ค้นพบโดยกูเกิลของคำค้นหา พิษณุโลก (Phitsanulok)

Name Variation of		
No.	Name Variation	Records were from Google
1	phitsanulok	5,110,000 records
2	phisanulok	186,000 records
3	pitsanulok	32,800 records
4	pisanulok	5,510 records
5	phitsanurok	1,400 records
6	phisanuklok	134 records
7	phizanulok	5 records
8	phitsanoolok	5 records
9	phitsanoklok	4 records

จากนั้นนำรายชื่อที่ได้มาทั้งหมดไปเข้ากระบวนการอัลกอริทึมใหม่ซึ่งแต่ละวิธี จากนั้นทำการวิเคราะห์ประสิทธิภาพด้วยการคำนวณหาประสิทธิภาพของแต่ละอัลกอริทึมของอัลกอริทึมใหม่แมชชีนด้วย F-Measure

วัดประสิทธิภาพแมชชีนอัลกอริทึมด้วยวิธีการ F-Measure

ขั้นตอนการวัดประสิทธิภาพแมชชีนอัลกอริทึมด้วยวิธีการ F-Measure สำหรับกลุ่มอัลกอริทึมแมชชีนซึ่งมีการเปลี่ยนชื่อเป็นรหัสเสียงนั้น สามารถแสดงวิธีการคำนวณค่าประสิทธิภาพโดยใช้ ตัวอย่างของอัลกอริทึม Soundex ดังตารางที่ 6

ตารางที่ 6 : ตัวอย่างการวัดประสิทธิภาพอัลกอริทึมแมชชีนด้วยวิธีการ F-Measure ของ อัลกอริทึม Soundex

ชื่อจังหวัดที่ถูกต้อง และ Soundex Code	ฐานข้อมูลชื่อ ที่มีความ คล้ายคลึงกัน	Soundex Code	Result (Match/No Match)	Match Type (TP, FP, FN)
Phitsanulok	Phisanulok	P254	No match	FN
P325	Pitsanulok	P325	Match	TP
	Pisanulok	P254	No match	FN
	Phitsanurok	P325	Match	TP
	Phisanuklok	P252	No match	FN
	Phizanulok	P254	No match	FN
	Phitsanoolok	P325	Match	TP
	phitsanoklok	P325	Match	TP
	...	...	...	
	Phitjint	P325	Match	FP

จากตารางที่ 6 สามารถคำนวณค่าต่างๆ ได้ดังต่อไปนี้

■ หาค่า TP ได้จากการนับชื่อที่ Match ในการเปรียบเทียบ Soundex Code ของชื่อจังหวัดที่เขียนถูกต้อง Phitsanulok (P325) กับชื่อจังหวัดในฐานข้อมูล โดยในตัวอย่างนี้ TP เท่ากับ 4

■ หาค่า FN ได้จากการนับชื่อที่ No Match ในการเปรียบเทียบ Soundex Code ของชื่อจังหวัดที่เขียนถูกต้อง Phitsanulok (P325) กับชื่อจังหวัดในฐานข้อมูล โดยในตัวอย่างนี้ FN เท่ากับ 4

■ หาค่า FP ได้จากการนับชื่อที่ Match ในการเปรียบเทียบ Soundex Code ของชื่อจังหวัดที่เขียนถูกต้อง Phitsanulok (P325) กับชื่อจังหวัดในฐานข้อมูล โดยในตัวอย่างนี้ FP เท่ากับ 1

■ คำนวณหาค่า Precision, ค่า Recall และค่า F-Score ของจังหวัด Phitsanulok สามารถหาค่าต่างๆ ได้ดังนี้

■ Precision : $4/(4+1) \times 100 = 80 \%$

■ Recall : $4/(4+4) \times 100 = 50 \%$

$$\blacksquare \text{ F-Score : } 2 \cdot (80 \cdot 50) / (80 + 50) = 61.54 \%$$

ดังนั้นสรุปได้ว่าค่าประสิทธิภาพของอัลกอริทึม Soundex สำหรับจังหวัด phitsanulok มีค่าเท่ากับ 61.54 % จากนั้นทำการหาค่า F-Score จนครบทุก 20 จังหวัดในฐานข้อมูล และครบทุกอัลกอริทึมเนมแมทซ์ซึ่งพร้อมทั้งสรุปผลการวิเคราะห์ประสิทธิภาพ

ผลการวิจัย

ผลจากการวิเคราะห์ประสิทธิภาพของอัลกอริทึมเนมแมทซ์ซึ่ง 6 วิธีการโดยใช้ ข้อมูลคำของจังหวัด 20 จังหวัด และคำที่คล้ายคลึงที่ปรากฏอยู่จริงจากการสืบค้นข้อมูลท่องเที่ยวจากภูเก็ลนั้นสามารถแสดงค่าประสิทธิภาพดัง ตารางที่ 7 และ 8

ตารางที่ 7 : ตัวอย่างผลการวัดประสิทธิภาพของอัลกอริทึม Phonex (ซ้าย) และ Soundex (ขวา) ด้วย F-Measure

Phonex	P	R	F-Score	Soundex	P	R	F-Score
buriram	1.000	0.8	0.8889	buriram	1.000	0.831	0.9077
kalasin	1.000	0.82	0.9011	kalasin	1.000	0.842	0.9142
nakhonratchasima	1.000	0.39	0.5612	nakhonratchasima	0.9	0.42	0.5727
maharakham	0.523	0.575	0.5478	maharakham	0.673	0.521	0.5873
ubonratchathani	0.816	0.488	0.6107	ubonratchathani	0.823	0.501	0.6228
phichit	0.929	0.898	0.9132	phichit	0.942	0.901	0.9210
phitsanulok	0.942	0.857	0.8975	phitsanulok	0.953	0.845	0.8958
pathumthani	0.652	0.545	0.5937	pathumthani	0.677	0.585	0.6276
phetchabun	0.763	0.738	0.7503	phetchabun	0.794	0.731	0.7612
maehongson	0.629	0.667	0.6474	maehongson	0.702	0.653	0.6766
nakhonsawan	0.568	0.917	0.7015	nakhonsawan	0.784	0.923	0.8478
chiangmai	0.965	0.902	0.9324	chiangmai	0.972	0.917	0.9437
sukhothai	0.885	0.963	0.9224	sukhothai	0.921	0.958	0.9391
chiangrai	0.307	0.964	0.4657	chiangrai	0.67	0.879	0.7604
kamphaengphet	1.000	0.7	0.8235	kamphaengphet	0.914	0.645	0.7563
udonthani	1.000	0.722	0.8386	udonthani	1.000	0.794	0.8852
nakhonsithammarat	0.477	0.7	0.5674	nakhonsithammarat	0.509	0.673	0.5796
ayutthaya	0.750	0.98	0.8497	ayutthaya	0.822	0.874	0.8472
suphanburi	1.000	0.32	0.4848	suphanburi	0.965	0.442	0.6063
sakonnakhon	0.649	0.6	0.6235	sakonnakhon	0.812	0.694	0.7484
Average	0.7928	0.7273	0.7261	Average	0.8417	0.7315	0.7827

จากการวัดประสิทธิภาพของอัลกอริทึมเนมแมทซ์ซึ่งด้วย F-Measure นั้นพบว่า ค่า F-Score หรือประสิทธิภาพในการจับคู่ชื่อที่ให้ค่าที่สูงที่สุดสำหรับระบบสืบค้นข้อมูลท่องเที่ยวคือ Soundex (78.27 %) รองลงมาคือ Soundex2 (75.68 %) และ Phonex (72.61 %) ตามลำดับ ซึ่งจะเห็นว่ามีความใกล้เคียงกันและถือว่ามีประสิทธิภาพของระบบอยู่ในเกณฑ์ดี โดยผลการวัดประสิทธิภาพของอัลกอริทึมทั้งหมดแสดงในตารางที่ 8 และเวลาในการประมวลผลของเนมแมทซ์อัลกอริทึมแสดงในตารางที่ 9

ตารางที่ 8 : ผลการวัดประสิทธิภาพเนมแมทซ์อัลกอริทึมด้วยวิธีการ F-Measure

Algorithm	P	R	F- Score (%)
Soundex	84.17	73.23	78.31
Soundex2	80.50	71.45	75.68
Phonex	79.28	72.73	72.61
Metaphone	67.5	58.71	62.32
Simplex	63.11	55.89	59.55
NYSIIS	56.51	49.71	52.09

ตารางที่ 9 : แสดงเวลาในการประมวลผลเนมแมทซ์อัลกอริทึม

Algorithm	เวลา (วินาที)
Soundex	0.01275
Metaphone	0.04309
Soundex2	0.20771
Phonex	0.24502
NYSIIS	0.27016
Simplex	0.29341

อภิปรายผลการวิจัย

จากผลการวัดประสิทธิภาพของอัลกอริทึมจะเห็นว่า ชื่อจังหวัดที่นำมาทดสอบนั้นจะประกอบไปด้วยจำนวนตัวอักษรตั้งแต่ 7 จนถึง 17 ตัวอักษร เพื่อให้มีความหลากหลายของชื่อ อาทิเช่น phitchit มีจำนวนตัวอักษรเท่ากับ 7 ตัวอักษร และ nakhonsithammarat มีจำนวนตัวอักษรเท่ากับ 17 ตัวอักษร เป็นต้น และจะเห็นว่าจำนวนตัวอักษรนั้นมีผลต่อประสิทธิภาพของอัลกอริทึมด้วย เช่น จำนวนตัวอักษรของชื่อจังหวัดตั้งแต่ 12 ตัวอักษร ขึ้นไป ส่วนใหญ่ จะให้ค่า F-Score ที่ต่ำ อาทิเช่น ประสิทธิภาพของอัลกอริทึม Soundex (ตารางที่ 7) จังหวัด nakhonsithammarat (17 ตัวอักษร) nakhonratchasima (16 ตัวอักษร) ubonratchathani (15 ตัวอักษร) mahasarakham (12 ตัวอักษร) ให้ค่า F-Score ที่ 57.96% 57.27% 62.28% และ 58.73% ตามลำดับ ในทางตรงกันข้ามจะเห็นว่า

จำนวนตัวอักษรของชื่อยังมีจำนวนน้อย ค่า F-Score จะให้ค่าที่สูง (มากกว่า 0.9 หรือ 90%) อาทิเช่น ประสิทธิภาพของอัลกอริทึม Soundex (ตารางที่ 7) จำนวนตัวอักษรของชื่อส่วนใหญ่ที่ไม่เกิน 9 ตัวอักษร ได้แก่ จังหวัด kalasin (7 ตัวอักษร) phichit (7 ตัวอักษร) Chiangmai (9 ตัวอักษร) และ Sukhothai (9 ตัวอักษร) ให้ค่า F-Score ที่ 91.42% 92.10% 94.37% และ 93.91% ตามลำดับ ซึ่งผลลัพธ์ของการวัดประสิทธิภาพของอัลกอริทึม Soundex จะสอดคล้องกับอัลกอริทึมของ Phonex และอัลกอริทึมอื่นด้วยเช่นกัน

จากตารางที่ 9 การเปรียบเทียบเวลาในการประมวลผลอัลกอริทึมเนมแมทซ์ซึ่งพบว่าอัลกอริทึมที่มีเวลาในการประมวลผลน้อยที่สุดตามลำดับ ดังนี้ Soundex, Metaphone, Soundex2, Phonex, NYSIS และ Simplex จะเห็นได้ว่า Soundex และ Metaphone ที่มีเวลาในการประมวลผลน้อย เนื่องจาก Soundex และ Metaphone มีฟังก์ชันในการทำงานในภาษา PHP อยู่แล้ว จึงทำให้ฟังก์ชันมีการเรียกใช้ได้ทันทีและสามารถประมวลผลได้เร็วมากขึ้น

ในงานวิจัยนี้ผู้วิจัยได้นำเสนอผลของการวิเคราะห์ประสิทธิภาพของระบบสืบค้นข้อมูลท่องเที่ยวโดยใช้อัลกอริทึม Variation Name Finding มาช่วยในการคัดเลือกชื่อจังหวัดที่มีความคล้ายคลึงหรือใกล้เคียงกับชื่อจังหวัดที่เขียนถูกต้องเพื่อนำไปใช้เป็นข้อมูลในการเปรียบเทียบอัลกอริทึมเนมแมทซ์ซึ่งแต่ละวิธี ซึ่งการวัดประสิทธิภาพและเปรียบเทียบเนมแมทซ์อัลกอริทึมเพื่อให้ทราบว่าอัลกอริทึมใด ให้ประสิทธิภาพดีที่สุดสำหรับการสืบค้นชื่อหรือข้อมูลท่องเที่ยวภาษาอังกฤษ และจากการทดสอบพบว่า อัลกอริทึมเนมแมทซ์ซึ่งมีความสามารถในการสืบค้นข้อมูลท่องเที่ยวได้ค่อนข้างดีที่สุด ในจำนวนอัลกอริทึมเนมแมทซ์ของกลุ่มที่มีการเปลี่ยนชื่อเป็นรหัสเสียง ได้แก่ Soundex, Soundex2 และ Phonex ตามลำดับ และ Soundex นั้นยังใช้เวลาในการประมวลผลน้อยที่สุด อีกด้วย ซึ่งระบบสืบค้นข้อมูลท่องเที่ยวนี้สามารถนำอัลกอริทึม Soundex ไปประยุกต์ใช้หรือปรับปรุงให้ดียิ่งขึ้น เพื่อที่จะช่วยให้ข้อมูลท่องเที่ยวที่มีการเขียนผิดรูปผิดเสียง สามารถค้นหาเจอในอินเทอร์เน็ตได้อย่างรวดเร็ว อย่างไรก็ตามควรมีการเพิ่มจำนวนข้อมูลชื่อสถานที่ท่องเที่ยวที่นำมาทดสอบหาประสิทธิภาพให้มีความหลากหลายรูปแบบให้ครอบคลุมข้อมูลท่องเที่ยวมากขึ้นเพื่อผลลัพธ์ที่ได้จะมีความน่าเชื่อถือยิ่งขึ้นไปด้วย

ข้อเสนอแนะ

จากการวัดประสิทธิภาพและเปรียบเทียบเนมแมทซ์อัลกอริทึมของกลุ่มที่มีการเปลี่ยนชื่อเป็นรหัสเสียงทั้ง 6 อัลกอริทึมจะเห็นได้ว่าผลการทดสอบที่ได้มีความน่าเชื่อถือในระดับค่อนข้างดี แต่อย่างไรก็ตามขอเสนอแนะที่ควรมีการปรับปรุงให้ดียิ่งขึ้น มีดังต่อไปนี้

1. ควรทำการวิเคราะห์ประสิทธิภาพของอัลกอริทึมเนมแมทซ์ของกลุ่มที่มีการเปรียบเทียบกับค่า Threshold เพิ่มอีกด้วย ซึ่งอัลกอริทึมเนมแมทซ์ในกลุ่มนี้มี 6 อัลกอริทึมได้แก่ Levenshtein, Guth, ISG, LIG1, LIG2 และ LIG3

2. ควรมีการวัดประสิทธิภาพในภาษาอื่น ๆ นอกเหนือจากภาษาอังกฤษเพิ่มมากขึ้นเช่นภาษาไทย ซึ่งการประยุกต์ใช้อัลกอริทึมเนมแมทซ์สำหรับภาษานั้น Snae and Brückner (2009) จักรกฤษณ์ เสน่ห์ นมะหุต

ได้พัฒนาอัลกอริทึมเนมแมทซึ่งสำหรับภาษาไทย คือ Metasound ซึ่งนำไปใช้กับการตั้งชื่อและเปลี่ยนชื่อไทยให้สอดคล้องกับชื่อบิดาและมารดาหรือเปลี่ยนชื่อให้มีความคล้ายคลึงกับชื่อเดิม

3. ควรมีการเพิ่มการวิเคราะห์และการวัดประสิทธิภาพด้วยวิธีการอื่นเพิ่มมากขึ้นเช่น หลักการของ Lait and Randell (1998) และ หลักการจัดกลุ่ม เป็นต้น

4. ควรมีการทดลองเพื่อแสดงผลการเปรียบเทียบเวลาในการประมวลผลอัลกอริทึมเนมแมท ซึ่งเพื่อเป็นส่วนหนึ่งในการเลือกอัลกอริทึมเนมแมทซึ่งในการนำไปประยุกต์ในการสืบค้นและเพื่อเพิ่มความรวดเร็วของระบบสืบค้นข้อมูลท่องเที่ยวอีกด้วย

5. ควรเพิ่มจำนวนฐานข้อมูลของชื่อ ให้ครบทุกจังหวัดที่มีข้อมูลท่องเที่ยวปรากฏจะทำให้ค่าประสิทธิภาพมีความเชื่อถือได้มากยิ่งขึ้น

รายการอ้างอิง

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์ นมะหุต. 2556. การวิเคราะห์ประสิทธิภาพของระบบสืบค้นข้อมูลออนโทโลยีท่องเที่ยวด้วยอัลกอริทึม ISG และ Name Variation Matching. วารสารวิทยาศาสตร์ มหาวิทยาลัยนเรศวร. ปีที่ 9. ฉบับที่ 2. 47-64.

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์. 2553. ระบบค้นหาสถานที่ท่องเที่ยวในประเทศไทยด้วยการออนโทโลยีและเนมแมทซึ่ง. **Journal of Information Science and Technology**. ปีที่ 1. ฉบับที่ 2. 60-69.

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์. 2554. การพัฒนาระบบสืบค้นข้อมูลออนโทโลยีท่องเที่ยวด้วยภาษา SPARQL. เอกสารการประชุมทางวิชาการนเรศวรวิจัย ครั้งที่ 7. 195-204.

สุรางค์รัตน์ ชาวโคกสูง และ จักรกฤษณ์ เสน่ห์. 2554. การพัฒนาระบบบริการข้อมูลการท่องเที่ยวและแนะนำกำหนดการท่องเที่ยวด้วยการแสดงผลในลักษณะระบบสารสนเทศภูมิศาสตร์. วารสารมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยอุบลราชธานี. ปีที่ 2. ฉบับที่ 1. 31-45.

Brückner, Michael.; Snae, Chakkrit and Payakpate, Janjira. 2008. Ontology-Based Name Matching of Toponyms for Geographical Information Systems (ONTO-GIS). In **Proceedings of IADIS International Conference ICT, Society, and Human Beings**. 108-114.

Cao, Tuan-Dung and Nguyen, Quang-Minh. 2012. Semantic approach to travel information search and Itinerary recommendation. **International Journal of Web Information Systems**. Vol. 8. Issue. 3. 256-277.

Cao, Tuan-Dung; Nguyen, Quang-Minh; Nguyen, Anh-Duc and Le, Tan-Hung. 2011. Integrating Open Data and Generating Travel Itinerary in Semantic-Aware Tourist Information System. In **Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services**. 214-221.

- Castillo, Luis; Armengol, Eva; Onaindía, Eva; Sebastiá, Laura; González-boticario, Jesús; Rodríguez, Antonio; Fernández, Susana; D. Arias, Juan and Borrajo, Daniel. 2008. SAMAP: An User-Oriented Adaptive System for Planning Tourist Visits. **Expert Systems with Applications**. Vol. 34. Issue 2. 1318-1332.
- Davey, C. and Jarvis, A.S. 1990. Microcomputers for Microhistory: a Database Approach to the Reconstruction of Small English Populations. **History and Computing**. Vol. 2. No.3. 187-193.
- Gill, Leicester; Goldacre, Michael; Simmons, Hugh; Bettley, Glenys and Griffith, Myfanwy. 1993. Computerised Linking of Medical Records: Methodological Guidelines. **Journal of Epidemiology and Community Health**. Vol .47. 316-319.
- Kongthon, Alisa; Kongyoung, Sarawoot; Haruechaiyasak, Choochart and Palingoon, Pompimon. 2011. A Semantic Based Question Answering System for Thailand Tourism Information. **In Proceedings of the KRAQ11 Workshop**. 38-42.
- Lait, Alan. J. and Randell, Brian. 1998. An Assessment of Name Matching Algorithm. Society of Indexers Genealogical Group, Newsletter Contents. SIGGNL issues 17. 1-32.
- Maervoet, Joris X.; Souffriau, Wouter; Vansteenwegen, Pieter; Vanden, Berghe Greet and Van, Oudheusden Dirk. 2008. Tourist Decision Support for Mobile Navigation Systems: a Demonstration. **Applied Artificial Intelligence**. Vol. 22. No. 10. 964-985.
- Panawong, Naruepon; Snae Namahoot, Chakkrit and Brückner, Michael. 2014. Classification of Tourism Web with Modified Naïve Bayes Algorithm. **Advanced Materials Research**. Vol. 931-932. 1360-1364.
- Park, Heum; Kwon, Soonho and Kwon, Hyuk-Chul. 2009. Ontology-based Approach to Intelligent Ubiquitous Tourist Information System. **In Proceedings of the 4th International Conference on Ubiquitous Information Technologies & Applications**. 1-6.
- Snae, Chakkrit and Brückner, Michael. 2006. Concept and Rule Based Naming System. **Issues in Informing Science and Information Technology**. Vol. 3. 619-634.
- Snae, Chakkrit and Brückner, Michael. 2009. Novel Phonetic Name Matching Algorithm with a Statistical Ontology for Analysing Names Given in Accordance with Thai Astrology. **Issues in Informing Science and Information Technology**. Vol. 6. 497-515.
- Snae, Chakkrit and Brückner, Michael. 2007. LOWCOST: Local Organization Search with Consolidated Ontologies for Name, Space and Time. **In Proceedings of the 25th Conference on IASTED International Multi-Conference: Software Engineering**. 43-48.

- Snae, Chakkrit; Brückner, Michael. and Wongthongtham, Pornpit. 2007. Local Organization and Business Ontology (LOBO). In **Proceedings of IEEE International Conference on Digital Ecosystems and Technologies**. 292-295.
- Snae, Chakkrit and Diaz, Bernard M. 2002. An Interface for Mining Genealogical Nominal Data Using the Concept of Linkage and a Hybrid Name Matching Algorithm. **Journal of 3D-Forum Society**. Vol. 16, No. 1. 142-147.
- Sylejmani, Kadri and Dika, Agni. 2011. Solving Touristic Trip Planning Problem by Using Taboo Search Approach. **IJCSI International Journal of Computer Science Issues**. Vol. 8. No. 3. 139-149.
- Winchester, Ian. 1970. The Linkage of Historical Records by Man and Computer: Techniques and Problems. **Journal of Interdisciplinary History**. Vol. 1. No. 1. 107-124.

เอกสารอ้างอิงจากภาษาไทยเป็นภาษาอังกฤษ (Translated Thai References)

- Panawong, Naruepon and Snae Namahoot, Chakkrit. 2013. Performance Analysis of an Ontology-Based Tourism Information System with ISG Algorithm and Name Variation Matching. **NU Science Journal**. Vol. 9. No. 2. 47-64. (In Thai)
- Panawong, Naruepon and Snae, Chakkrit. 2010. Search System for Attractions in Thailand with Ontology and Name Matching. **Journal of Information Science and Technology**. Vol. 1. No. 2. 60-69. (In Thai)
- Panawong, Naruepon and Snae, Chakkrit. 2011. System Development of Search Tourism Ontology with SPARQL. **Proceedings of the 7th Research Naresuan**. 195-204. (In Thai)
- Chowkoksung, Surangrat and Snae, Chakkrit. 2011. The Development of a Web-based System on GIS for Tourist Information System and Tourism Packaging. **Journal of Humanities and Social Sciences, Ubon Ratchathani University**. Vol. 2. No. 1. 31-45. (In Thai)