NU Science Journal 2013; 9(2) : 47 - 64

การวิเคราะห์ประสิทธิภาพของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยวด้วยอัลกอริทึม ISG และ Name Variation Matching

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์ นมะหุต*

Performance Analysis of an Ontology-Based Tourism Information System with ISG Algorithm and Name Variation Matching

Naruepon Panawong and Chakkrit Snae Namahoot*

ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก 65000

*Corresponding author. Email: chakkrits@nu.ac.th

บทคัดย่อ

เมื่อนักท่องเที่ยวต้องการสืบค้นข้อมูลท่องเที่ยวส่วนมากจะทำการสืบค้นผ่านเสิร์ชเอนจินที่ได้รับความนิยมอย่างกูเกิ้ล แต่ผลลัพธ์ที่ได้นั้นมีจำนวนมากอาจต้องเสียเวลาในการเข้าถึงข้อมูลจากหลายเว็บไซต์และผลลัพธ์ที่ได้อาจไม่ตรงกับความต้องการ อีกทั้งในบางเว็บไซต์นำเสนอข้อมูลบนเว็บไซต์ที่อาจมีการเขียนไม่ถูกต้องตามหลักภาษาไทยหรือภาษาอังกฤษก็จะทำให้ได้ผลของการสืบค้นไม่พบเว็บไซต์ที่เขียนไม่ถูกต้อง ดังนั้นผู้วิจัยจึงได้นำออนโทโลจีมาจัดเก็บข้อมูลท่องเที่ยว ซึ่งทำให้ข้อมูลออนไลน์ที่เนื้อหาเดียวกันแต่มีรูปแบบการนำเสนอที่แตกต่างกันเพื่อสามารถเชื่อมโยงข้อมูลการท่องเที่ยวต่าง ๆ อย่างมีความสัมพันธ์ต่อกัน อีกทั้งอัลกอริทึม ISG (Index of Similarity Group) จะถูกนำมาใช้ในการกรองข้อมูลชื่อที่สามารถเขียนได้หลายรูปแบบหรือมีความคล้ายคลึงกัน ทั้งคำพ้องรูป พ้องเสียงหรือสะกดผิดในภาษาไทยหรือภาษาอังกฤษ เพื่อให้ได้เว็บไซต์ที่เขียนผิดถูกค้นพบ การวัดประสิทธิภาพของอัลกอริทึม ISG และ Name Variation Matching ในงานวิจัยนี้จะใช้ F-Measure ซึ่งจากผลการวิจัยพบว่า อัลกอริทึม ISG และ Name Variation ที่ใช้ในการกรองข้อมูลชื่อแหล่งท่องเที่ยวของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยวทั้งภาษาไทยและภาษาอังกฤษนั้นมีประสิทธิภาพไม่แตกต่างกัน ซึ่งอยู่ในเกณฑ์ดี

คำสำคัญ: การสืบค้นข้อมูล ออนโทโลจี ข้อมูลออนโทโลจีท่องเที่ยว อัลกอริทึมเนมแมทชิ่ง การวัดประสิทธิภาพด้วย F-Measure

Abstract

To extract tourism information, most travelers rely on web search engines such as, Google, so they will face with a ton of web pages from results. Consequently, they have time consuming to access multiple websites and might not get what they need. Moreover, if keywords or Web site contents are misspelled in Thai or in English, the search results will consist of incomplete or incorrect information. To deal with these problems, we adopt an ontology to connect similar information for different platforms. The ISG algorithm (Index of Similarity Group) is used to make connections several forms in both Thai and English words, such as homograph, homophone and misspelling. In order to validate this algorithm and Name Variation Matching, F-Measure is used. The results show that ISG and Name Variation can classify attraction places information with names in Thai language as well as in English language effectively.

Keywords: Information Retrieval, Ontology, Tourism Ontology, Name Matching Algorithm, F-Measure

บทนำ

การท่องเที่ยวถือเป็นอุตสาหกรรมบริการที่ทำให้ประเทศมีรายได้เพิ่มขึ้นและยังมีบทบาทต่อการพัฒนาเศรษฐกิจของชาติอีกด้วย ทั้งนี้ในปัจจุบันพบว่ามีข้อมูลท่องเที่ยวเป็นจำนวนมากที่ถูกนำเสนอบนอินเทอร์เน็ตและมีรูปแบบการนำเสนอที่แตกต่างกันออกไป เช่น รูปแบบข้อความ รูปแบบมัลติมีเดีย เป็นต้น ซึ่งในการค้นหาข้อมูลท่องเที่ยวเพื่อให้ได้ตรงกับความต้องการของนักท่องเที่ยวนั้นจึงมีความสำคัญต่อนักท่องเที่ยวเป็นอย่างมาก โดยเฉพาะการป้อนคำค้นหาที่บางครั้งมีการเขียนผิดโดยไม่ตั้งใจหรือการเขียนหรือสะกดคำไม่ถูกต้องตามหลักภาษาไทยหรือภาษาอังกฤษ หรือมีการใช้คำค้นหาที่เป็นชื่อเรียกอื่นๆ ของสถานที่นั้นๆ เช่น วัดใหญ่ หมายถึง วัดพระศรีรัตนมหาธาตุวรมหาวิหาร สองแคว หมายถึง จังหวัดพิษณุโลก เป็นต้น ซึ่งทำให้ได้ผลลัพธ์ตรงและไม่ต้องกับความต้องการปะปนกันอยู่

ในงานวิจัยนี้ได้วิเคราะห์ประสิทธิภาพของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยวที่นำอัลกอริทึม Name Variation Matching มาสร้างคำที่คล้ายคลึงกันตามตัวสะกดที่สามารถเขียนแทนกันได้ เช่น แม่กก แม่กน แม่กบ แม่กด และคำพ้องรูป พ้องเสียง เช่น พ คล้าย ภ หรือ ร คล้าย ล เพื่อแก้ปัญหาคำที่มีการเขียนผิดหรือไม่ถูกต้องตามหลักภาษาไทยและภาษาอังกฤษ ทำให้เว็บไซต์ที่เขียน

ผิดสามารถถูกค้นพบ แต่เนื่องจาก Name Variation ที่ถูกสร้างขึ้นมีจำนวนมาก ผู้วิจัยจึงใช้อัลกอริทึม ISG เพื่อกรองข้อมูลก่อนนำไปสืบค้นในเว็บไซต์กูเกิ้ล และวัดประสิทธิภาพของอัลกอริทึม ISG และ Name Variation ด้วย F-Measure

วิจารณ์วรรณกรรม

ออนโทโลจี

ออนโทโลจีได้กลายมาเป็นที่รู้จักและใช้กันอย่างแพร่หลายในปัจจุบัน เช่น สาขาปัญญาประดิษฐ์ สาขาวิชาคอมพิวเตอร์และสาขาวิทยาศาสตร์สารสนเทศ โดยเฉพาะอย่างยิ่งในขอบเขตสาขาทางด้านระบบสารสนเทศ การรวบรวมสารสนเทศทางปัญญา การสืบค้นข้อมูล การแทนความรู้ การออกแบบฐานข้อมูล การจัดการพื้นฐานองค์ความรู้ และระบบการจัดการฐานข้อมูล (Guarino, 1998; Andrade and Saltz, 1999, 2000) โดย Gruber (1993) ได้นิยามความหมายของออนโทโลจีซึ่งเป็นที่รู้จักกันอย่างกว้างขวางมากที่สุดว่าออนโทโลจีเป็นรายละเอียดที่ชัดเจนแน่นอนของแนวความคิด ดังนั้นเมื่อต้องการแสดงหรือระบุแนวความคิดของวัตถุหรือสิ่งของนั้น ๆ ควรจะแสดงรายละเอียดที่ชัดเจนและแน่นอน Chandrasekaran และคณะ (1999) อีกทั้ง Zhang และคณะ (2010) อธิบายว่าออนโทโลจีเป็นการรวมศาสตร์ทางด้านปัญญาประดิษฐ์และภาษาเครื่องเข้าไว้ด้วยกัน เพื่อที่จะแบ่งปันความรู้และนำความรู้ต่าง ๆ กลับมาใช้ได้อีก นอกจากนี้ยังมีการประมวลผลภาษาธรรมชาติ การแทนความรู้ (Smith and Welty, 2001) ที่สามารถนำมาใช้ระหว่างพฤติกรรมของมนุษย์และระบบคอมพิวเตอร์ได้ อีกทั้งยังเป็นประโยชน์ในการสืบค้นข้อมูลและการจัดการความรู้โดยรายละเอียดแสดงด้วยคอนเซ็ปต์ต่างๆ ในโดเมนที่สามารถกำหนดการอธิบายได้ด้วยคลาส สล็อต และเงื่อนไขของสล็อตที่เรียกว่าฟาเซตส์ (Noy and McGuinness, 2001)

OWL (McGuinness and Harmelen, 2004) ถูกสร้างขึ้นโดย W3C Web Ontology Working Group (WebOnt) เพื่อเป็นส่วนขยายต่อจากภาษา RDF (Klyne and Carroll, 2004) ใช้ในการแก้ปัญหาข้อจำกัดในการระบุเงื่อนไขให้กับความสัมพันธ์ระหว่างคลาสในภาษา RDF ซึ่ง OWL นั้นจัดเป็นองค์ประกอบหนึ่งของเว็บเชิงความหมายที่ใช้ในการบรรยายข้อมูลเชิงความหมาย สามารถกำหนดโครงสร้างข้อมูลในลักษณะลำดับชั้นและอธิบายข้อมูล (Metadata) ที่มีความสัมพันธ์ในระบบฐานข้อมูลได้ อีกทั้งยังสามารถใช้ผสมผสานกันระหว่าง RDF และ XML (Extensible Markup Language) ขึ้นอยู่กับลักษณะของการใช้งานและรายละเอียดของข้อมูล ซึ่งถูกจัดเก็บเป็นไฟล์นามสกุล .owl, .rdf และ .xml ตามลำดับ

ปัจจุบันมีนักวิจัยที่ได้ออกแบบและพัฒนาออนโทโลจีดังตัวอย่างต่อไปนี้ Snae และคณะ (2007) ได้นำเสนอระบบ LOBO (Local Organization Business Ontology) ที่ใช้หลักการออนโทโลจีมาใช้กับองค์กรธุรกิจต่างๆ เพื่อช่วยในการสืบค้นสารสนเทศแบบพลวัตและเป็นการผสมผสานขององค์ประกอบที่แตกต่างกันในด้านความหมาย เช่น ชื่อองค์กร สถานที่ ชั่วโมงธุรกิจ จำนวนพนักงาน ซึ่งสามารถคัดลอกข้อมูลโดยตรงจากแหล่งข้อมูลที่เป็นแบบพลวัตได้ อีกทั้งระบบนี้ได้ใช้วิธีเสนอแบบคำถามคำตอบในจังหวัดพิษณุโลก เช่น ขอบเขตของธุรกิจคืออะไร จำนวนของบริษัทที่ให้บริการอินเทอร์เน็ตเท่าไหร่ โรงแรมที่มีมาตรฐานระดับสูงสุดได้แก่ มีห้างสรรพสินค้าที่อยู่ใกล้กับสถานีรถไฟหรือไม่ (รัศมี 1 กิโลเมตร) จำนวนบริษัทที่เลิกกิจกรรมในปี 2005 และจำนวนบริษัทที่มีพนักงานมากกว่า 1000 คน ซึ่งระบบสามารถรวบรวมและสุ่มคำที่น่าจะอธิบายถึงสิ่งที่อยู่ภายใต้ขอบเขตขององค์กรและธุรกิจในพิษณุโลก ทำให้ได้ข้อมูลบริเวณขององค์กรธุรกิจ สถานที่ คำศัพท์และชื่อของธุรกิจ อีกทั้งยังได้รวบรวมคำนิยามหรือคำอธิบายสำหรับคำหรือชื่อเหล่านั้นได้ เพื่อให้ระบบสามารถจับคู่และแสดงความสัมพันธ์ของคำที่เกี่ยวกับองค์กรธุรกิจได้

Snae และ Brueckner (2007) ได้พัฒนาระบบ LOWCOST (Local Organization Search With Consolidated Ontologies for Name, Space and Time) ที่ใช้สำหรับการค้นหาองค์กรระดับท้องถิ่นด้วยชื่อด้วยหลักการออนโทโลจีในการออกแบบเพื่อแก้ปัญหารูปแบบเวลาที่เขียนในรูปแบบต่างกันให้อยู่ในรูปแบบมาตรฐานเดียวกันและนำทฤษฎีเนมแมทชิ่งมาแก้ปัญหาชื่อ ชื่อสถานที่ ที่มีการเขียนแตกต่างกันหรือคล้ายคลึงกัน ซึ่งใช้ทั้งด้านการออกเสียง (Phonetic Variation) และความแตกต่างด้านโครงสร้างหรือรูปแบบ (Form Variation) หรือข้อผิดพลาดในการสะกดชื่อที่แตกต่างกัน (Spelling Variation) มาช่วยจากคำที่สามารถเขียนได้หลากหลาย โดยใช้เครื่องมือที่เรียกว่า IT-TELL (Snae *et al*., 2006) รวมถึงการเขียนชื่อในรูปแบบชื่อเต็มและชื่อย่อ โดยระบบสามารถช่วยในการค้นหาประวัติขององค์กร ชื่อขององค์กรที่มีการเขียนแตกต่างกันหรือคล้ายคลึงกันและการค้นหาแบบ Temporal และ Spatial ได้ เพื่อให้ได้คำตอบที่ถูกต้อง

Brueckner และคณะ (2008) ได้นำเสนอการออกแบบระบบการค้นหาสถานที่ต่างๆ ในแผนที่กูเกิ้ล โดยใช้หลักการออนโทโลจีเพื่อสนับสนุนการค้นหาจากคำค้นหาที่ใช้ภาษาไทย ภาษาอังกฤษและภาษาญี่ปุ่นและมีใช้หลักการเนมแมทชิ่งด้วยอัลกอริทึม Canonical Name Matching มาช่วยในการแก้ปัญหาการพิมพ์ชื่อที่สามารถเขียนได้หลากหลายหรือคล้ายคลึงกันเพื่อให้ได้คำตอบที่ใกล้เคียงมากที่สุดและแสดงผลลัพธ์ได้อย่างรวดเร็ว อีกทั้งยังจะได้แสดงผลการค้นหาด้วยแผนที่กูเกิ้ลโดยใช้ Geographic Markup Language (GML) and the Keyhole Markup Language (KML)

Dema (2008) ได้นำหลักการออนโทโลจีมาใช้ในการสร้างองค์ความรู้สำหรับการเดินทางและวางแผนท่องเที่ยวให้กับนักท่องเที่ยวในประเทศภูฐาน (eTourPlan) ที่จัดเก็บอยู่ในรูปแบบไฟล์

RDF และใช้ Partonomy Rules ในการเข้ารหัสพื้นที่ทางภูมิศาสตร์ พร้อมทั้งจัดการพื้นฐานของการค้นหากิจกรรมท่องเที่ยว เทศกาล สถานที่ท่องเที่ยว ที่พักและแนะนำเส้นทางในการเดินทางไปยังสถานที่ๆ ต้องการ อีกทั้งยังได้นำ OO jDREW (Object Oriented java Deductive Reasoning Engine for the Web) ซึ่งเป็นซอฟต์แวร์รหัสเปิด (Open Source) มาใช้สำหรับจัดการกฎต่างๆ ที่มีเพื่อแสดงผลการค้นหาข้อมูลการท่องเที่ยวและวางแผนการท่องเที่ยว ซึ่งองค์ความรู้ที่ใช้ในงานวิจัยนี้ประกอบไปด้วย 115 คลาส 73 แฟกซ์ และ 37 กฎ ที่เก็บข้อมูลการท่องเที่ยวในประเทศภูฐาน รวมไปถึงการคำนวณเส้นทางระหว่างจังหวัด โดยมีองค์ความรู้ 3 ส่วน คือ การค้นหาสถานที่ท่องเที่ยว (Search Engine) ระบบแนะนำการท่องเที่ยว (Recommender System) และวางแผนการเดินทาง (Travel Planner) ซึ่งทำการค้นหาผ่านทาง OO jDREW เท่านั้น

จุฑามาศ ศิริรัชนีกร และ ฤทธิกร เล้าอรุณ (2549) ยังได้นำเสนอระบบการค้นหาข้อมูลแบบซับซ้อนสำหรับข้อมูลการท่องเที่ยวในประเทศไทย โดยใช้นำเสนอในรูปแบบของเว็บเชิงความหมาย และมีการออกแบบข้อมูลให้มีความเชื่อมโยงความสัมพันธ์กัน เพื่อให้ผู้ใช้ได้ผลลัพธ์ของค้นหาข้อมูลในสถานที่ใกล้ ๆ กันแต่การนำเสนอเป็นเพียงข้อมูลที่ใช้ภาษาอังกฤษเท่านั้นยังไม่รองรับการใช้ภาษาไทย อีกทั้งยังไม่รองรับการใช้คำค้นหาด้วยชื่อเรียกอื่นๆ ได้

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์ (2553) ได้ออกแบบออนโทโลจีท่องเที่ยว ประกอบด้วย 10 คลาสคือคลาสจังหวัด คลาสอำเภอ คลาสตำบล คลาสเทศกาล คลาสสถานที่ท่องเที่ยว คลาสที่พัก คลาสการเดินทาง คลาสร้านอาหาร คลาสร้านขายของฝาก และคลาสร้านขายสินค้าหนึ่งผลิตภัณฑ์หนึ่งตำบล อีกทั้งยังได้พัฒนาระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยว โดยหากผู้ใช้ต้องการรู้ข้อมูลแหล่งท่องเที่ยวในประเทศไทย ผู้ใช้จะป้อนคำค้นหาจากนั้นระบบจะดึงความรู้และข้อมูลต่างๆ ที่ถูกเก็บไว้ในออนโทโลจีท่องเที่ยวโดยผลลัพธ์ที่ได้นั้นจะถูกนำไปค้นหาผลลัพธ์ที่มีการเขียนหรือสะกดคล้ายคลึงกันในเว็บไซต์ต่างๆ โดยใช้ Name Variation Matching Algorithm ประสิทธิภาพของงานนี้ประเมินได้จากการพยายามทำให้ข้อมูลต่างๆ ที่ได้จากการค้นหาที่มีการเขียนหรือสะกดคล้ายคลึงกันหรือที่เกี่ยวข้องสามารถค้นหาเจอได้ครบถ้วนมากที่สุด และข้อมูลพื้นฐานของจังหวัดนั้นๆ จะถูกแสดงอย่างมีความสัมพันธ์กันอย่างถูกต้องครบถ้วน เช่น คำขวัญประจำจังหวัด สัญลักษณ์ประจำจังหวัด ต้นไม้ประจำจังหวัด เป็นต้น

ISG (Index of Similarity Group)

ISG เป็นอัลกอริทึมของแนมแมทชิ่ง (Name Matching Algorithm) ซึ่งจะใช้ในการวัดหรือเปรียบค่าความคล้ายคลึงกันระหว่างชื่อโดยมีสเกลจาก 0-1 (Snae, 2007) โดยหากค่าความคล้ายคลึงระหว่างชื่อมีค่าเป็น 1 หมายถึงชื่อนั้นเหมือนกัน ค่าความคล้ายคลึงระหว่างชื่อเข้าใกล้ 1 มากเท่าใด

หมายถึงชื่อทั้งสองมีความคล้ายคลึงกันมากเท่านั้น ค่าความคล้ายคลึงระหว่างชื่อเข้าใกล้ 0 หมายถึงชื่อทั้งสองมีความคล้ายคลึงกันน้อยเท่านั้น และค่าความคล้ายคลึงระหว่างชื่อมีค่าเป็น 0 แสดงว่าชื่อทั้งสองไม่มีความคล้ายคลึงกันเลย อัลกอริทึมนี้ยังถูกนำไปใช้ในระบบตรวจสอบชื่อดีโดยใช้หลักการตั้งชื่อแบบทักษาปกรณ์และการจัดกลุ่ม (จักรกฤษณ์ เสน่ห์ และ กนกกาญจน์ นมะหุต, 2550) ซึ่งเป็นวิธีการเทียบตัวอักษรอย่างง่าย ไม่ยุ่งยากและไม่ซับซ้อน ทำให้การประมวลผลของการเปรียบเทียบชื่อและรายชื่อข้อมูลเป็นไปอย่างรวดเร็ว โดยมีสมการในการคำนวณดังต่อไปนี้

$$ISG = \frac{I}{I + D} \quad \text{.....(1)}$$

เมื่อ I คือ จำนวนตัวอักษรที่เหมือนกันระหว่างชื่อ

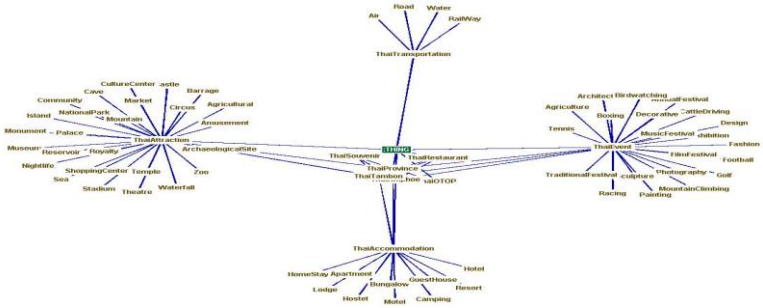
D คือ ตัวอักษรที่แตกต่างกันระหว่างชื่อ

วิธีการและการออกแบบระบบ

การออกแบบระบบสืบค้นออนโทโลจีท่องเที่ยวนั้นประกอบไปด้วย 3 ส่วนคือ ออนโทโลจีท่องเที่ยว อัลกอริทึม ISG และสถาปัตยกรรมของระบบ

การออกแบบออนโทโลจีท่องเที่ยว

ในงานวิจัยนี้ผู้วิจัยได้ทำการสร้างออนโทโลจีท่องเที่ยวและจัดเก็บข้อมูลท่องเที่ยวทั้งหมดโดยใช้โปรแกรม Protégé และบันทึกไว้ในไฟล์ OWL ดังรูป 1 ซึ่งมีลักษณะของการจัดเก็บแบบ XML ที่สามารถกำหนดคุณสมบัติของคลาสและมีการสืบทอดเป็นคลาสย่อยทำให้ OWL มีประโยชน์ต่อรูปแบบของ XML Schema ด้วย เพราะแสดงออกถึงความหมายและความจริงได้มากกว่า RDF, RDFS และ XML ด้วยเหตุนี้จึงจำเป็นต้องมีการจัดกลุ่มข้อมูลตามความสัมพันธ์ไว้ด้วยกันด้วยหลักการออนโทโลจี โดยแสดงตัวอย่างการเก็บข้อมูลของจังหวัดพิษณุโลกในไฟล์ OWL ดังแสดงในรูป 2



รูป 1 ออนโทโลจีท่องเที่ยวโดยใช้โปรแกรม Protégé

รูป 1 แสดงออนโทโลจีของข้อมูลท่องเที่ยวที่ทำการแบ่งเป็นคลาสต่างๆ เช่น คลาสจังหวัด (ThaiProvince) คลาสเทศกาล (ThaiEvent) คลาสสถานที่ท่องเที่ยว (ThaiAttraction) คลาสที่พัก (ThaiAccommodation) คลาสการเดินทาง (ThaiTransportation) เป็นต้น

```
<myont:ThaiProvince rdf:about="http://www.owl-ontologies.com/tourism.owl#Phitsanulok">
<myont:hasNameOfProvince xml:lang="th">พิษณุโลก</myont:hasNameOfProvince>
<myont:hasNameOfProvince xml:lang="en">Phitsanulok</myont:hasNameOfProvince>
<myont:hasTraditionalNameOfProvince xml:lang="th">เมืองสองแคว พิคโลก</myont:hasTraditionalNameOfProvince>
<myont:hasTraditionalNameOfProvince xml:lang="en">SongKhwae PhitLok</myont:hasTraditionalNameOfProvince>
<myont:hasMotto xml:lang="th">พระพุทธชินราชงามเลิศ ถิ่นกำเนิดพระนเรศวร สองฝั่งน่านล้วนเรือนแพ หวานฉ่ำแท้กล้วยตาก ถ้ำและน้ำ
ตกหลากตระการตา</myont:hasMotto>
<myont:hasMotto xml:lang="en">Buddha Chinnarat beauty excellence. King Naresuan origin. Nan Fang are two houseboat real sweet
dried banana and various spectacular water fall.</myont:hasMotto>
<myont:hasFlowers xml:lang="th">ดอกนนทรี</myont:hasFlowers>
<myont:hasFlowers xml:lang="en">Nonsi flowers</myont:hasFlowers>
<myont:hasSeals xml:lang="th">พระพุทธชินราช</myont:hasSeals>
<myont:hasSeals xml:lang="en">Shin Buddhist Kingdom</myont:hasSeals>
<myont:hasTree xml:lang="th">ปีบ</myont:hasTree>
<myont:hasTree xml:lang="en">Millingtonia hortensis</myont:hasTree>
<myont:hasURLOfProvince
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">www.phitsanulok.go.th</myont:hasURLOfProvince>
<myont:hasImageOfProvince rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Phitsanulok</myont:hasImageOfProvince>
<myont:hasLatitudeOfProvince rdf:datatype="http://www.w3.org/2001/XMLSchema#string">17.015936</myont:hasLatitudeOfProvince>
<myont:hasLongtitudeOfProvince
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">100.66983</myont:hasLongtitudeOfProvince>
</myont:ThaiProvince>
```

รูป 2 ตัวอย่างการเก็บข้อมูลในไฟล์ OWL (tourism.owl)

รูป 2 แสดงตัวอย่างการเก็บข้อมูลจังหวัดพิษณุโลก โดยในแต่ละ Tag นั้นจะมีค่า Property ที่แสดงถึงการเก็บข้อมูลภาษาไทย (xml:lang="th") และภาษาอังกฤษ (xml:lang="en") หรือหากต้องการเก็บภาษาอื่นๆ ก็ให้กำหนดในค่า Property นี้ได้ ซึ่งสามารถอธิบายในแต่ละ Tag ได้ดังนี้

- myont:hasNameOfProvince เก็บชื่อจังหวัด
- myont:hasTraditionalNameOfProvince เก็บชื่อเดิมหรือชื่อเรียกอื่นๆ ของจังหวัด

- myont:hasMotto เก็บคำขวัญประจำจังหวัด
- myont:hasFlowers เก็บชื่อดอกไม้ประจำจังหวัด
- myont:hasSeals เก็บสัญลักษณ์ประจำจังหวัด
- myont:hasTree เก็บชื่อต้นไม้ประจำจังหวัด
- myont:hasURLOfProvince เก็บที่อยู่เว็บไซต์ประจำจังหวัด
- myont:hasImageOfProvince เก็บชื่อรูปภาพ
- myont:hasLongtitudeOfProvince เก็บค่าลองติจูด
- myont:hasLatitudeOfProvince เก็บค่าละติจูด

ในงานวิจัยนี้ได้ใช้อัลกอริทึม ISG (รูปที่ 3) เพื่อหาค่าความคล้ายคลึงของชื่อจังหวัดที่มีการสะกดผิดหรือเขียนผิดซึ่งปรากฏอยู่ในเว็บไซต์ท่องเที่ยวต่างๆ บนอินเทอร์เน็ต ซึ่งจะช่วยให้ชื่อที่มีการเขียนผิดหรือสะกดผิดและปรากฏอยู่ในข้อมูลท่องเที่ยวสามารถถูกสืบค้นพบในอินเทอร์เน็ตได้ เช่น ถ้าเราจะค้นหาคำว่า พิษณุโลก (Phitsanulok) อัลกอริทึมจะนำคำต่างๆ ที่ได้จากอัลกอริทึม Name Variation Matching มาทำการคัดกรองโดยคำนวณความคล้ายคลึง ดังตัวอย่างที่แสดงในตาราง 1 และตาราง 2 โดยอัลกอรทึม ISG มีหลักการทำงานดังต่อไปนี้

1. รับชื่อ 2 ชื่อและคำนวณหาความยาวของชื่อทั้งสอง โดยยึดชื่อที่มีความยาวสูงสุดเป็นหลัก
2. กำหนดค่าเริ่มต้นของตัวชี้ตำแหน่งอักษร ค่าความเหมือนและค่าความแตกต่าง ให้มีค่าเป็นศูนย์
3. ตรวจสอบตัวชี้ตำแหน่งอักษรกับความยาวสูงสุดของชื่อ หากพบว่าตัวชี้ตำแหน่งมากกว่าความยาวสูงสุดของชื่อแล้วให้ไปทำงานข้อที่ 8
4. ถ้าตัวอักษรของชื่อทั้งสองเหมือนกันให้เพิ่มค่าความเหมือนในตำแหน่งใดๆ
5. ถ้าตัวอักษรของชื่อทั้งสองไม่เหมือนกันให้เพิ่มค่าความแตกต่าง
6. เพิ่มค่าตัวชี้ตำแหน่งอักษร
7. กลับไปยังข้อที่ 3
8. คำนวณหาค่า ISG จากสูตร 1

```
$len1 = strlen($var1); $len2 = strlen($var2);
$pointer = 0; $same = 0; $diff = 0;
Loop:
        if ($pointer > $maxvarlen) return
($same/($same+$diff));
        if ( comparechar($var1,$var2) > 0) $same++;
        else $diff++;
```

รูป 3 อัลกอริทึม ISG

ตาราง 1 ตัวอย่างการคำนวณชื่อจังหวัดพิษณุโลก (ภาษาไทย) ด้วยอัลกอริทึม ISG

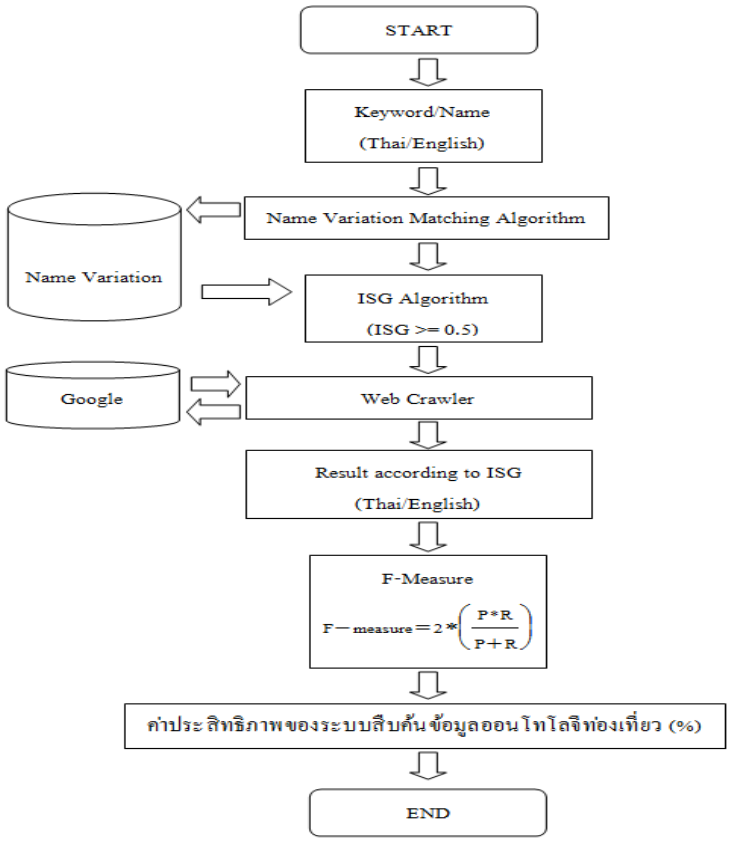
ลำดับ	ชื่อที่คล้ายคลึงกับ จังหวัดพิษณุโลก	จำนวนตัวอักษรที่เหมือนกัน (I)	จำนวนตัวอักษรที่แตกต่างกัน (D)	ค่า ISG I/(I+D)	จำนวนรายการที่พบ (กูเกิ้ล, กันยายน 2554)
1.	พิษณุโลก	8	0	1.0000	5,470,000
2.	พิษะณุโลก	8	1	0.8889	110,000
3.	พิดษะณุโลก	8	2	0.8000	76,200
4.	พิดสะณุโลก	7	3	0.7000	2,290,000
5.	พิจสะณุโลก	7	3	0.7000	2,200,000
6.	พิษนุโลก	7	1	0.8750	931,000
7.	พิตสะณุโลก	7	3	0.7000	106,000
8.	พิจสะนุโลก	6	4	0.6000	10,100,000
9.	พิสสะนุโลก	6	4	0.6000	1,230,000
10.	พิดโลก	5	3	0.6250	5,770,000
11.	ภีฎสะนุโลก	4	6	0.4000	3,000,000
12.	ภีดศะนุโลก	4	6	0.4000	197,000
13.	พีดสะนุโรก	4	6	0.4000	15,300,000
14.	พีดศะนุโรก	4	6	0.4000	1,200,000
15.	พีดโรก	3	5	0.3750	66,700,000

ตาราง 2 ตัวอย่างการคำนวณชื่อจังหวัด Phitsanulok (ภาษาอังกฤษ) ด้วยอัลกอริทึม ISG

ลำดับ	ชื่อที่คล้ายคลึงกับจังหวัด Phitsanulok	จำนวนตัวอักษรที่เหมือนกัน (I)	จำนวนตัวอักษรที่แตกต่างกัน (D)	ค่า ISG I/(I+D)	จำนวนรายการที่พบ (กูเกิ้ล, กันยายน 2554)
1.	phitsanulok	11	0	1.0000	1,240,000
2.	phisanulok	10	1	0.9091	603,000
3.	pitsanulok	10	1	0.9091	117,000
4.	pisanulok	9	2	0.8182	17,200
5.	phitsanurok	10	1	0.9091	1,910
6.	phisanuklok	10	1	0.9091	105
7.	phitzanulok	10	1	0.9091	53
8.	phitsanoolok	10	2	0.8333	6
9.	phitsanoklok	10	2	0.8333	3
10.	phisanooklok	9	3	0.7500	1
11.	pheedzanoorok	6	7	0.4615	0
12.	pedzanoorok	5	6	0.4545	0
13.	pedzanuroc	5	7	0.4167	0
14.	peetzanooroc	5	7	0.4167	28
15.	peedzanooroc	4	8	0.3333	0

ตาราง 1 และตาราง 2 แสดงตัวอย่างการการคำนวณชื่อจังหวัดพิษณุโลกที่มีการเขียนหลากหลายแบบตามหลักการออกเสียงในภาษาไทยและภาษาอังกฤษ พร้อมแสดงตัวอย่างวิธีการคำนวณค่า ISG ระหว่างชื่อ 2 ชื่อในภาษาไทยและชื่อภาษาอังกฤษ โดยการเปรียบเทียบจะใช้วิธีการเปรียบเทียบจากชื่อที่มีจำนวนตัวอักษรมากที่สุด ตัวอย่างเช่น พิษณุโลก กับ พืดสะนุโรก มีตัวอักษรที่เหมือนกัน 4 ตัว คือ พ สระอุ โ ก และมีตัวอักษรต่างกัน 6 ตัว คือ สระอี ด ส สระอะ น ร และ phitsanulok กับ phisanuklok มีตัวอักษรเหมือนกัน 10 ตัว คือ p h i s a n u l o k และมีตัวอักษรต่างกัน 1 ตัวคือ t

สถาปัตยกรรมของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยว



รูป 4 สถาปัตยกรรมของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยว

รูป 4 แสดงสถาปัตยกรรมของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยว โดยมีหลักการทำงานดังต่อไปนี้

1. ป้อนชื่อจังหวัดหรือชื่อแหล่งท่องเที่ยวที่ต้องการค้นหา

2. Name Variation Matching Algorithm จะสร้างชื่อที่สามารถเขียนได้หลากหลายหรือคล้ายคลึงกันชื่อในข้อที่ 1 และจัดเก็บลงฐานข้อมูล การทำงานของ Name Variation Matching มีดังต่อไปนี้

2.1 ตรวจสอบว่าเป็นภาษาไทยหรือภาษาอังกฤษ

2.2 กรณีเป็นภาษาไทยสร้างชื่อจากความคล้ายคลึงของพยัญชนะและตัวสะกด แต่หากเป็นภาษาอังกฤษสร้างชื่อจากความคล้ายคลึงของตัวอักษร (Snae *et al*., 2007; Snae and Brueckner, 2009) ดังแสดงตัวอย่างตามตาราง 3 และตัวอย่างโปรแกรมในรูปที่ 5

ตาราง 3 ตัวอย่างการทำงานของอัลกอริทึม Name Variation Matching

ลำดับ	คำค้นหา	ตำแหน่งตัวอักษร	ตัวอักษรที่เปลี่ยน	เปลี่ยนเป็น	คำที่ได้
1.	พิษณุโลก	1	พ	ภ	ภิษณุโลก
2.	พิษณุโลก	2	ษ	ศ	พิศณุโลก
3.	ภิษณุโลก	2	ษ	ศ	ภิศณุโลก
4.	พิษณุโลก	3	ณ	น	พิษนุโลก
5.	ภิษณุโลก	3	ณ	น	ภิษนุโลก
6.	พิศณุโลก	3	ณ	น	พิศนุโลก
7.	ภิศณุโลก	3	ณ	น	ภิศนุโลก

```
$varname = strtolower($varname);
$vlen = strlen($varname);
for ($i = 0; $i < $vlen; $i++) {
      if ($varname[$i] == "พ")
        $namevar = substr($namevar,0,strpos($namevar,$varname[$i]))."ภ".
substr($varname,$i+1);
      else if ($varname[$i] == "ษ") {
        $tmp = substr($namevar,0,strpos($namevar,$varname[$i]));
        $namevar = $tmp."ศ".substr($varname,$i+1);
        $namevar = $tmp."ส".substr($varname,$i+1);
      }
}
```

รูป 5 ตัวอย่างอัลกอริทึม Name Variation Matching

3. นำชื่อที่ได้จากข้อที่ 2 ไปตรวจสอบชื่อที่มีความคล้ายคลึงกันมากที่สุด และทำการคัดกรองด้วยอัลกอริทึม ISG ซึ่งจะใช้ค่า ISG ที่มีค่ามากกว่าหรือเท่ากับ 0.5 ในการคัดกรอง

4. นำชื่อต่างๆ ที่ได้จากข้อ 3 ไปทำการสืบค้นหาข้อมูลในกูเกิ้ล โดยใช้เทคนิค web crawler ในการดึงข้อมูลจากผลลัพธ์ของการค้นหาจากกูเกิ้ล

5. แสดงผลลัพธ์ของการสืบค้นจากข้อ 4

6. นำผลลัพธ์จากการแสดงในข้อ 5 มาคำนวณหาประสิทธิภาพของระบบด้วย F-Measure

7. แสดงค่าของ F-Measure

การทดสอบและผลลัพธ์

ในการวัดประสิทธิภาพการสืบค้นผู้วิจัยได้ใช้ F-Measure (Christen and Goiser, 2007) สำหรับคำค้นหาในภาษาไทยและภาษาอังกฤษ ซึ่งทำการวิเคราะห์ข้อมูลท่องเที่ยวที่นิยมใน 20 จังหวัด (ตาราง 4 และตาราง 5) แล้วนำชื่อดังตัวอย่างในตาราง 4 และตาราง 5 ไปสืบค้นหาข้อมูลในกูเกิ้ลซึ่งได้ผลลัพธ์ของรายการเว็บไซต์รวมกันทั้งหมดเป็นจำนวน 1,148,348 รายการ แบ่งเป็นเว็บไซต์จากคำค้นหาภาษาอังกฤษจำนวน 387,767 และเว็บไซต์จากคำค้นหาภาษาไทยจำนวน 760,581 จากนั้นทำการวิเคราะห์ประสิทธิภาพการสืบค้นตามสมการ (2) ดังต่อไปนี้

$$F-measure = 2 * \left(\frac{P * R}{P + R} \right) \quad \text{.....(2)}$$

เมื่อ P คือ True Positive/(True Positive + False Positive)

R คือ True Positive/(True Positive + False Negative)

True Positive คือ คำค้นหาที่มีค่า ISG มากกว่าหรือเท่ากับ 0.5

False Positive คือ คำค้นหาที่มีค่า ISG มากกว่าหรือเท่ากับ 0.5 และไม่พบในกูเกิ้ล

True Negative คือ คำค้นหาที่ไม่พบในกูเกิ้ล

False Negative คือ คำค้นหาที่มีค่า ISG น้อยกว่า 0.5 และไม่พบในกูเกิ้ล

ตาราง 4 ตัวอย่างผลจากการกรองข้อมูลด้วยอัลกอริทึม ISG (ภาษาไทย)

ลำดับ	คำค้นหา	Variation	ISG >= 0.5	พบในกูเกิ้ล	ไม่พบในกูเกิ้ล
1.	พิษณุโลก	208	169	208	0
2.	กาญจนบุรี	96	78	96	0
3.	ขอนแก่น	6	2	6	0
4.	ชัยภูมิ	64	38	64	0
5.	เชียงใหม่	16	16	16	0
6.	เพชรบุรี	224	124	220	4
7.	ภูเก็ต	17	4	16	1
8.	สุโขทัย	73	33	72	1
9.	สุรินทร์	588	250	588	0
10.	หนองคาย	13	7	12	1

ตาราง 5 ตัวอย่างผลจากการกรองข้อมูลด้วยอัลกอริทึม ISG (ภาษาอังกฤษ)

ลำดับ	คำค้นหา	Variation	ISG >= 0.5	พบในกูเกิ้ล	ไม่พบในกูเกิ้ล
1.	phitsanulok	341	312	83	258
2.	kanchanaburi	91	86	61	30
3.	khonkaen	58	39	46	12
4.	chaiyaphum	37	36	29	8
5.	chiangmai	43	22	28	15
6.	phetchaburi	341	223	85	256
7.	phuket	72	32	70	2
8.	sukhothai	265	180	174	91
9.	surin	37	19	36	1
10.	nongkhai	109	80	86	23

จากตาราง 4 และตาราง 5 แสดงผลจากการกรองข้อมูลด้วยอัลกอริทึม ISG ทั้งในภาษาไทยและภาษาอังกฤษ โดยในช่อง Variation แสดงจำนวนคำที่สามารถเขียนได้หลากหลายรูปแบบ โดยคำเหล่านี้ได้มาจากการทำงานของ Name Variation Matching Algorithm ซึ่งในแต่ละคำค้นหานั้นจะได้จำนวนของ variation ไม่เท่านั้น ทั้งนี้เนื่องจากความยาวของคำค้นหาที่ไม่เท่ากันและจากการสะกดคำของภาษาไทย ทำให้บางคำที่ได้จากการทำงานของอัลกอริทึม variation นั้นไม่พบในกูเกิ้ลเลย จากนั้นแสดงจำนวนของคำค้นหาที่มีค่า ISG >= 0.5 จำนวนคำที่เขียนได้หลากหลายรูปแบบที่ถูกพบจากผลลัพธ์ของการสืบค้นจากเว็บไซต์กูเกิ้ลและจำนวนคำที่เขียนได้หลากหลายรูปแบบที่ไม่ถูกพบจากผลลัพธ์ของการสืบค้นจากเว็บไซต์กูเกิ้ลตามลำดับ แต่เนื่องจาก Name Variation ที่ถูกสร้างขึ้นมีจำนวนมาก ผู้วิจัยจึงใช้อัลกอริทึม ISG เพื่อกรองข้อมูลก่อนนำไปสืบค้นในเว็บไซต์กูเกิ้ล

ตาราง 6 การวัดประสิทธิภาพของอัลกอริทึมด้วย F-Measure (ภาษาไทย)

ลำดับ	คำค้นหา	T+	F+	T-	F-	P	R	F-Measure
1.	พิษณุโลก	169	0	0	39	1.0000	0.8125	0.8966
2.	กาญจนบุรี	78	0	0	18	1.0000	0.8125	0.8966
3.	ขอนแก่น	2	0	0	4	1.0000	0.3333	0.5000
4.	ชัยภูมิ	38	0	0	26	1.0000	0.5938	0.7451
5.	เชียงใหม่	16	0	0	0	1.0000	1.0000	1.0000

ตาราง 6 (ต่อ)

ลำดับ	คำค้นหา	T+	F+	T-	F-	P	R	F-Measure
6.	เพชรบุรี	124	0	4	96	1.0000	0.5636	0.7209
7.	ภูเก็ต	4	0	1	12	1.0000	0.2500	0.4000
8.	สุโขทัย	33	0	1	39	1.0000	0.4583	0.6286
9.	สุรินทร์	250	0	0	338	1.0000	0.4252	0.5967
10.	หนองคาย	7	0	1	5	1.0000	0.5833	0.7368
	ค่าเฉลี่ย					**1.0000**	**0.5833**	**0.7121**

ตาราง 7 การวัดประสิทธิภาพของอัลกอริทึมด้วย F-Measure (ภาษาอังกฤษ)

ลำดับ	คำค้นหา	T+	F+	T-	F-	P	R	F-Measure
1.	phitsanulok	80	258	0	3	0.2367	0.9639	0.3800
2.	kanchanaburi	60	26	4	1	0.6977	0.9836	0.8163
3.	khonkaen	39	3	4	12	0.9286	0.7647	0.8387
4.	chaiyaphum	29	7	1	0	0.8056	1.0000	0.8923
5.	chiangmai	20	8	7	8	0.7143	0.7143	0.7143
6.	phetchaburi	80	165	86	10	0.3265	0.8889	0.4776
7.	phuket	32	0	2	38	1.0000	0.4571	0.6275
8.	sukhothai	170	27	24	44	0.8629	0.7944	0.8273
9.	surin	19	0	1	17	1.0000	0.5278	0.6909
10.	nongkhai	60	7	16	26	0.8955	0.6977	0.7843
	ค่าเฉลี่ย					**0.7468**	**0.7792**	**0.7049**

จากตาราง 6 และตาราง 7 แสดงการวัดประสิทธิภาพของอัลกอริทึมด้วย F-Measure ที่คำนวณได้จากสมการ (2) ซึ่งแสดงตัวอย่างการคำนวณจังหวัด “พิษณุโลก” ได้ดังนี้

F-Measure = 2*{{[169/(169+0)]*[169/(169+39)]}/ {[169/(169+0)]+[169/(169+39)]}}

= 2*{{1*0.8125}/{1+0.8125}}

= 2*{0.8125/1.8125}

= 2* 0.4482

= 0.8964

สำหรับการคำนวณชื่อภาษาไทยชื่ออื่นและชื่อภาษาอังกฤษนั้นจะมีการคำนวณเช่นเดียวกันกับตัวอย่างชื่อจังหวัด “พิษณุโลก” ในงานวิจัยนี้พบว่าคำค้นหาที่มาจากภาษาไทยมีค่า F-Measure เท่ากับ 0.7121 หรือคิดเป็น 71.21% และคำค้นหาที่มาจากภาษาอังกฤษมีค่า F-Measure เท่ากับ 0.7049 หรือคิดเป็น 70.49% ซึ่งมีค่าใกล้เคียงกันและถือว่ามีประสิทธิภาพของระบบอยู่ในเกณฑ์ดี

บทสรุปและข้อเสนอแนะ

ในงานวิจัยนี้ผู้วิจัยได้นำเสนอผลของการวิเคราะห์ประสิทธิภาพของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยวโดยใช้อัลกอริทึม Name Variation Matching และ ISG มาช่วยในการกรองข้อมูลเพื่อให้ได้ข้อมูลที่มีความคล้ายคลึงหรือใกล้เคียงกับความต้องการของผู้ใช้มากที่สุด โดยจะช่วยให้ข้อมูลท่องเที่ยวที่มีการเขียนผิดรูปผิดเสียง สามารถค้นหาเจอได้ในอินเทอร์เน็ต อีกทั้งผู้วิจัยยังได้ทำการวิเคราะห์ประสิทธิภาพของระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยวด้วย F-Measure ซึ่งพบว่าระบบค้นหาชื่อแหล่งท่องเที่ยวทั้งภาษาไทยและภาษาอังกฤษมีประสิทธิภาพใกล้เคียงกันซึ่งอยู่ในเกณฑ์ดี

ในอนาคตผู้วิจัยจะทำการวิเคราะห์ผลลัพธ์ที่ได้จากการสืบค้นในเว็บไซต์กูเกิ้ลเพื่อเฉพาะเว็บไซต์เกี่ยวกับท่องเที่ยวเท่านั้น ซึ่งจะวิเคราะห์ในแต่ละเว็บด้วยหลักการคำนวณน้ำหนักคำในโครงสร้างของเว็บไซต์ (เกรียงกมล คำมา และ จักรกฤษณ์ เสน่ห์ นมะหุต, 2555) โดยใช้คำค้นหาจากออนโทโลจีท่องเที่ยวที่ออกแบบไว้เพื่อให้ได้ผลที่มีความน่าเชื่อถือมากยิ่งขึ้นและจะใช้อัลกอริทึมนา-อีฟเบย์ (Patil and Pawar, 2012) สำหรับจัดหมวดหมู่ท่องเที่ยวเพื่อให้สะดวกในการเข้าถึงเว็บไซต์ที่เกี่ยวกับท่องเที่ยวอย่างมีประสิทธิภาพ

เอกสารอ้างอิง

จักรกฤษณ์ เสน่ห์ และ กนกกาญจน์ นมะหุต. (2550). ระบบตรวจสอบชื่อดีโดยใช้หลักการตั้งชื่อแบบทักษาปกรณ์และการจัดกลุ่ม. *วารสารวิทยาศาสตร์ มหาวิทยาลัยนเรศวร, 4*(1), 89-106

เกรียงกมล คำมา และ จักรกฤษณ์ เสน่ห์ นมะหุต. (2555). ระบบวิเคราะห์เว็บไซต์อนาจารด้วยกลุ่มคำเชิงความหมายคำอนาจารใน HTML Tags. *การประชุมวิชาการระดับชาติวิทยาศาสตร์วิจัย ครั้งที่ 4*, 17-21

จุฑามาศ ศิริรัชนีกร และ ฤทธิกร เล้าอรุณ. (2549). ระบบการค้นหาข้อมูลแบบซับซ้อนสำหรับข้อมูลการท่องเที่ยวในประเทศไทย. สืบค้นเมื่อ 20 ธันวาคม 2552, จาก http://cpe.kmutt.ac.th/previousproject/2006/26/Index.htm

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์. (2553). ระบบค้นหาสถานที่ท่องเที่ยวในประเทศไทยด้วยหลักการออนโทโลจีและเนมแมทชิ่ง. *Journal of Information Science and Technology, 1*(2), 60-69

Andrade, H. and Saltz, J. (1999). Towards a Knowledge Base Management System KBMS: An Ontology-Aware Database Management System DBMS. *Proceedings of the 14th Brazilian Symposium on Database*, Florianopolis, Brazil, 27-39

Andrade, H. and Saltz, J. (2000). Query Optimization in Kess An Ontology-Based KBMS. *Proceedings of the 15th Brazilian Symposium on Databases (SBBD'2000),* Joao Pessoa, Brazil, 35-48

Brueckner, M., Snae, C. and Payakpate, J. (2008). Ontology-based name matching of toponyms for geographical information systems (ONTO-GIS). *LADIS International Conference ICT*, Society and Human Beings, Amsterdam, The Netherlands, Jul., 108-114

Chandrasekaran, B., Josephson, J.R. and Benjamins, V.R. (1999). What are ontologies And Why Do We Need Them. *IEEE Intelligent System, 14*(1), 20-26

Christen, P. and Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. *Quality Measures in Data Mining*, Studies in Computational Intelligence, *43*, 127-151

Dema, T. (2008). eTourPlan: A knowledge-based tourist route and activity planner. Master thesis, M.Sc., University of New Brunswick, Canada.

Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition, 5*(2), 199-220

Guarino, N. (1998). Formal Ontology and Information Systems. *In Proceedings of the Third Conference FOIS'98*, ISO Press, Trento, Italy, 3-15

Klyne, G. and Carroll, J.J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation, W3C, February. Retrieved November 25 2010, from : http://www.w3.org/TR/2004/REC-rdf-concepts-20040210

McGuinness, D.L. and Harmelen, F.V. (2004). Owl Web Ontology Language Overview, World Wide Web Consortium (W3C) Recommendation. Retrieved November 25 2010, from : http://www.w3.org/TR/owl-features

Noy, N.F. and McGuinness, D.L. (2001). Ontology Development: A Guide to Creating Your First

Ontology. *Technical Report KSL-01-05*, Stanford Knowledge Systems Laboratory

Patil, A. S. and Pawar, B. V. (2012). Automated Classification of Web Sites using Naïve Bayesian Algorithm. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2012*, *1*, 519-523

Smith, B. and Welty, C. (2001). Formal Ontology in Information Systems. *Proceedings of the International conference on formal Ontology in Information Systems,* Ogunquit, Maine, New York: ACM Press

Snae C. (2007). A Comparison and Analysis of Name Matching Algorithms. *International Journal of Applied Science. Engineering and Technology*, *4*(1), 252-257

Snae, C. and Brueckner, M. (2007). LOWCOST: Local organization search with consolidated ontologies for name, space and time. *in International Conference on Software Engineering*, Innsbruck, Austria, Feb., 13-15

Snae, C. and Brueckner, M. (2009). Novel Phonetic Name Matching Algorithm with a Statistical Ontology for Analysing Names Given in Accordance with Thai Astrology. *Informing Science and Information Technology*, *6*, 497-515

Snae, C., Brueckner, M. and Wongthongtham, P. (2007). Local organization and business ontology (LOBO). *in IEEE International Conference on Digital Ecosystems and Technologies,* Rydges Tradewinds Cairns, Australia, Feb., 292-295

Snae, C., Namahoot, K. and Brueckner, M. (2007). MetaSound: A New Phonetic Based Name Matching Algorithm for Thai Naming System. *International Conference on Engineering, Applied Science and Technology (ICEAST 2007)*, Bangkok, Thailand

Snae, C., Singhadech, N., Emapana B. and Brueckner, M. (2006). Interactive transliteration tools for explanation level language system (IT-TELLS). *In International Technical Conference on Circuits/Systems Computers and Communication,* Chiang Mai, Thailand, Jul., 245-248

Zhang, G., Jia, S., and Wang, Q. (2010). Construct ontology-based enterprise information metadata framework. *Journal of Software*, *5*(3), 312-319