

การพัฒนาเครื่องมือลิงก์ดาต้าวิซวลไลเซชันสำหรับอุตสาหกรรมการท่องเที่ยวไทย

Development of A Linked Data Visualization Tool for Thailand Tourism

นภาพร ศรีนนท์ (Napaporn Srinon)* และ กิตติมา เมฆาบุญชากิจ (Kittima Mekhabunchakij)*

บทคัดย่อ

อุตสาหกรรมการท่องเที่ยวเป็นอุตสาหกรรมหลักที่สำคัญของประเทศไทย โดเมนการท่องเที่ยวมีความซับซ้อนสูงและมีการเปลี่ยนแปลงตลอดเวลา ผู้มีอำนาจตัดสินใจมักพึ่งพารูปแบบการพยากรณ์ที่จะคาดการณ์ความต้องการในอนาคตหรือระบบสนับสนุนการตัดสินใจเพื่อวิเคราะห์และเปรียบเทียบผู้มีส่วนได้เสียที่เกี่ยวข้อง การใช้ประโยชน์สาธารณะจากข้อมูลในปัจจุบันเพื่อสนับสนุนการตัดสินใจยังคงมีข้อจำกัด งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อนำเสนอแนวทางการสร้างโมเดลลิงก์โอเพ่นดาต้า (Linked Open Data, LOD) โดยใช้หลักการสนับสนุนการตัดสินใจ พัฒนาต้นแบบเว็บแหล่งข้อมูลตามมาตรฐาน LOD และวิธีการเพื่อจัดการความซับซ้อนของโดเมนการท่องเที่ยว โดยมุ่งเน้นที่วิธีการและเทคนิคการออกแบบ กระบวนการ Mapping ข้อมูลเชิงสัมพันธ์เป็นลิงก์ดาต้า (RDF/OWL) การพัฒนาออนโทโลยีการท่องเที่ยวด้วยโปรแกรม Protégé เครื่องมือวิซวลไลเซชันข้อมูลตัวชี้วัดการท่องเที่ยวเพื่อแสดงให้เห็นถึงความสำคัญของการสร้างโมเดล LOD ในการสนับสนุนการตัดสินใจในภาคอุตสาหกรรมการท่องเที่ยว โดยงานวิจัยนี้พัฒนาแอปพลิเคชัน Proof of Concept (POC) วิซวลไลเซชันชุดข้อมูลการท่องเที่ยวในประเทศไทยระยะเวลา 18 ปีด้วยข้อมูลทางสถิติและตัวชี้วัดของอุตสาหกรรม

คำสำคัญ: ลิงก์ดาต้า โอเพ่นดาต้า วิซวลไลเซชัน ออนโทโลยีการท่องเที่ยว

Abstract

Tourism is an important industry in Thailand. The tourism domain is a highly complex and dynamic domain where decision-makers often rely on forecasting models to predict

* คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีไทย-ญี่ปุ่น

* Faculty of Information Technology, Thai-Nichi Institute of Technology.

future demand or on decision support systems to analyze and compare the relevant stakeholders. The current issues are data integration from heterogeneous sources. This paper presents an approach to Linked Open Data (LOD) modeling in order to develop Web of Data based on LOD standards methodologies and to manage the complexity of the domain of Tourism. Focusing on the design methods, we propose an approach that uses the process of mapping relational data to Linked Data and key processes of LOD modeling and prototyping of Web of Data. The approach includes designing relational data model from various standards and stakeholders in tourism industry; transformations between relational and RDF/OWL data models; and tourism ontology development mainly using Protégé. Using the approach to modeling LOD for industry decision support. Visualization of key tourism indicators are implemented to illustrate the key processes of LOD modeling. This aims at building of a Proof of Concept (POC) application of visualization of Thailand tourism datasets of a 18-year period with complete measurements and indicators of the industry.

Keywords: Linked Data, Open Data, Data Visualization, Tourism Ontology.

1. บทนำ

ข้อมูลแบบเปิด (Open Data) เป็นแนวคิดที่ภาครัฐบาลและภาคธุรกิจสามารถใช้ข้อมูลที่มีจำนวนมากในการจัดการองค์กร ข้อมูลแบบเปิดของรัฐบาล (Open Government Data) มีข้อดีหลายประการ เช่น ความโปร่งใส และการมีส่วนร่วมของภาครัฐ การเพิ่มขีดความสามารถ ประสิทธิภาพของ

การให้บริการภาครัฐ การวัดผลกระทบของนโยบาย และ ความรู้ใหม่ของคุณลักษณะขนาดใหญ่จากหลายแหล่งที่มาและ รูปแบบที่หลากหลาย [1] บนพื้นฐานของมาตรฐานเว็บแบบ เปิด ลิงก์โอเพ่นดาต้า (Linked Open Data: LOD) ช่วยให้ ผู้บริโภคข้อมูลสามารถใช้เครื่องมือทั่วไปในการเข้าถึง ผสมผสาน และแสดงข้อมูล และมีศักยภาพในการเข้าถึง ข้อมูลภาครัฐได้อย่างกว้างขวาง [2], [3] แนวคิดของลิงก์ ดาตาสันับสนุน และมีวัตถุประสงค์เพื่อทำให้ข้อมูลจากหลาย แหล่งที่มาสามารถนำมารวมกันได้อย่างง่ายดายโดยใช้ เครื่องมือมาตรฐาน [4] ประเทศไทยมีการเผยแพร่เว็บไซต์ ข้อมูลแบบเปิดของรัฐบาลในปี 2014 ขณะนี้อยู่ในขั้นแรกของ นโยบายเศรษฐกิจดิจิทัล [5] อย่างไรก็ตามประเทศไทยยัง ไม่มีการทดลองที่แสดงให้เห็นถึงผลของการใช้ลิงก์ดาต้าที่มี จำนวนชุดข้อมูลในปริมาณมาก

ความสำคัญของตัวชี้วัดการท่องเที่ยว: ตัวชี้วัดการท่องเที่ยว มีความจำเป็นสำหรับการสนับสนุนกระบวนการตัดสินใจใน ภาคอุตสาหกรรม โดเมนการท่องเที่ยวมีความซับซ้อนสูง และมีความเปลี่ยนแปลงอยู่ตลอดเวลา ผู้ตัดสินใจต้องพึ่งพา รูปแบบการพยากรณ์ในการบริหารจัดการ ซึ่งต้องการ ความต้องการในอนาคต หรือใช้ระบบสนับสนุนการตัดสินใจ เพื่อทำการวิเคราะห์ผู้มีส่วนได้เสียที่เกี่ยวข้อง ข้อมูลทางสถิติ เช่น จำนวนนักท่องเที่ยวที่จุดหมายปลายทาง และระยะเวลา ของการเข้าพักที่จุดหมายปลายทางมีความสำคัญสำหรับ อุตสาหกรรมการท่องเที่ยวสำหรับการตัดสินใจต่างๆ [6] เช่น การทำความเข้าใจความมีส่วนร่วมของการท่องเที่ยว ที่มีต่อเศรษฐกิจ ณ จุดหมายปลายทาง ฯลฯ [7] การตลาดที่ จุดหมายปลายทางโดยการพยากรณ์ความต้องการด้าน การท่องเที่ยว และสำรวจแหล่งตลาดที่มีศักยภาพ [8] การวิเคราะห์ความยืดหยุ่นของความต้องการด้าน การท่องเที่ยว และการวิเคราะห์ฤดูกาล [9] นอกจากนี้ ผู้กำหนดนโยบายสาธารณะ และการวางแผนการท่องเที่ยว ยังสามารถใช้สถิติทางด้านอุตสาหกรรมในการตัดสินใจ เกี่ยวกับโครงสร้างพื้นฐานที่เกี่ยวข้องกับการวางแผน การท่องเที่ยว เช่น สถานที่พัก บริการอาหาร สนามบิน การบำบัดน้ำ ฯลฯ [6], [7]

งานวิจัยนี้จึงมุ่งเน้นการนำเสนอแนวทางการสร้างโมเดล LOD ข้อมูลตัวชี้วัดการท่องเที่ยว และพัฒนาเครื่องมือ วิซวลไลเซชันข้อมูลตัวชี้วัดการท่องเที่ยวที่สำคัญของ

ประเทศไทย (Thailand Tourism Linked Data Visualization: TTLDV) รวบรวมข้อมูลจากหลายแหล่งข้อมูลไว้ในที่ แห่งเดียวเพื่อแก้ปัญหาความซ้ำซ้อนของข้อมูล อำนวย ความสะดวกในการเชื่อมต่อ และการใช้ซ้ำชุดข้อมูล การท่องเที่ยวซึ่งมีปริมาณมาก

ในหัวข้อที่สองจะนำเสนอภาพรวมของทฤษฎีและงาน วิจัยที่เกี่ยวข้อง รวมทั้งพื้นฐานของลิงก์ดาต้า และทำงาน เกี่ยวกับลิงก์ดาต้าเพื่อการท่องเที่ยว

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ลิงก์ดาต้า

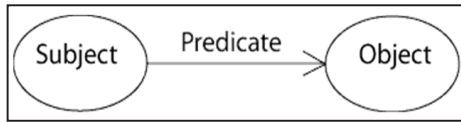
ลิงก์ดาต้า (Linked Data) อธิบายถึงชุดของวิธีการใน การเผยแพร่ข้อมูลที่มีโครงสร้างโดยใช้เทคโนโลยีเว็บ ความหมาย เช่น อาร์ดีเอฟ (Resource Description Framework: RDF) [10] ลิงก์ดาต้าเป็นแนวคิดของ Tim Berners-Lee โดย กำหนดไว้ด้วยสี่หลักการของเว็บเพื่อการเข้าถึงข้อมูล ได้แก่ 1) ใช้ URI ในการระบุตัวตนของทรัพยากร (Things) หรือ แนวคิด (Concepts) 2) ใช้ HTTP URI ที่เป็นที่อยู่ของสิ่งเหล่านั้น 3) การให้ข้อมูลที่เป็ประโยชน์โดยใช้มาตรฐานแบบเปิด (เช่น RDF) และ 4) การรวมของการเชื่อมโยงไปยัง URI อื่นๆ (เพื่อการค้นพบสิ่งของหรือแนวคิดที่มากขึ้น) [11], [12] นั่น ก็คือการเผยแพร่ข้อมูลที่เชื่อมโยงกันบนเว็บที่จะช่วยให้ คอมพิวเตอร์สามารถค้นพบข้อมูลเพิ่มเติมจากข้อมูลที่มีอยู่ แล้วโดยอัตโนมัติ

ในทางเทคนิค ลิงก์ดาต้าเป็นวิธีการแสดงข้อมูลที่มี การเชื่อมโยงภายในจากแหล่งข้อมูลที่แตกต่างกัน จนกลายเป็นมาตรฐานสำหรับที่ทำให้เกิดเว็บเชิงความหมาย เป็นวิธี ที่มีระบบของการค้นพบความรู้โดยแสดงข้อมูลโดเมนสำคัญ ด้วยความสัมพันธ์ ลิงก์ดาต้าใช้ RDF เป็นโมเดลข้อมูล และ ใช้ HTTP เป็นโปรโตคอล คล้ายกับเว็บที่ถูกสร้างขึ้นโดยใช้ HTML, HTTP, และ URL [6]

2.2 Resource Description Framework (RDF)

อาร์ดีเอฟ (RDF) เป็นมาตรฐานโครงสร้างที่ใช้ใน การอธิบายลักษณะของข้อมูลบนเว็บ เช่น ชื่อหัวเรื่อง ผู้เขียน วันเวลา เป็นต้น พัฒนาขึ้นโดยองค์กร W3C (World Wide Web Consortium) ได้รับการออกแบบขึ้นมาเพื่อให้ คอมพิวเตอร์ หรือ แอปพลิเคชันของคอมพิวเตอร์ อ่านและ เข้าใจได้ อาร์ดีเอฟมีโครงสร้างข้อมูลที่เรียกว่า ทริปเปิล

(Triple) ประกอบด้วย 3 ส่วน ได้แก่ ทรัพยากร (Subject) คุณสมบัติของทรัพยากร (Predicate) และค่าของคุณสมบัติ (Object) โดยเซตของทริปเปิล เรียกว่า อาร์ดีเอฟกราฟ (RDF Graph)



ภาพที่ 1 อาร์ดีเอฟกราฟ

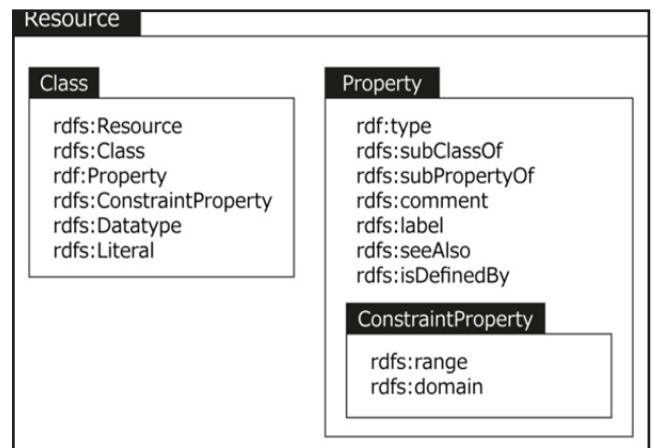
อาร์ดีเอฟเพิ่มคุณค่าให้กับข้อมูลที่มีอยู่ด้วยความสามารถในการเชื่อมโยงข้อมูลข้ามชุดข้อมูล (Linked Data) โดยขยายโครงสร้างการเชื่อมต่อของเว็บเพื่อใช้ URI ในการตั้งชื่อความสัมพันธ์ระหว่างสองวัตถุ โดยการใช้องค์สร้างที่เรียบง่ายนี้จะช่วยให้ข้อมูลที่มีโครงสร้าง และก็มีโครงสร้างสามารถผสมผสาน เปิดเผย และใช้งานร่วมกันในต่างแอปพลิเคชันได้ [13] อาร์ดีเอฟมีคุณสมบัติที่อำนวยความสะดวกในการรวมข้อมูล แม้ว่ารูปแบบของข้อมูลจะต่างกันก็ตาม ทำให้การรวมข้อมูลทำได้โดยง่าย และยังสนับสนุนการวิวัฒนาการของรูปแบบของข้อมูลโดยไม่จำเป็นต้องเปลี่ยนรูปแบบของข้อมูลทั้งหมดให้เหมือนกัน [14] ข้อมูลประเภทอาร์ดีเอฟเป็นข้อมูลที่เหมาะสมกับการใช้ในการพัฒนาเว็บแอปพลิเคชัน ผู้พัฒนาสามารถนำข้อมูลหลายๆชุดข้อมูลมาสร้างเป็นแอปพลิเคชันใหม่ได้ง่ายยิ่งขึ้น

2.2.1 RDF Schema

อาร์ดีเอฟเอส (RDFS) เป็นส่วนขยายเชิงความหมายของอาร์ดีเอฟ มีกลไกสำหรับแบ่งชนิดและอธิบายเกี่ยวกับกลุ่มทรัพยากรที่เกี่ยวข้องกัน และความสัมพันธ์ระหว่างทรัพยากรเหล่านั้น อาร์ดีเอฟเอสเขียนอยู่ในรูปแบบของอาร์ดีเอฟซึ่งประกอบด้วย คลาส (Class) และคุณสมบัติ (Property) ซึ่งคล้ายกับแนวคิดของการเขียนโปรแกรมเชิงวัตถุ (เช่น ภาษาโปรแกรม Java) ทำให้สามารถกำหนดทรัพยากรต่างๆ ให้เป็น Instance ของคลาส และซับคลาสได้

2.3 เว็บเชิงความหมาย

เว็บเชิงความหมาย เป็นส่วนขยายของเว็บในปัจจุบันนี้ทำให้ข้อมูลมีความหมายชัดเจนมากขึ้น (Well-defined meaning) และช่วยให้คอมพิวเตอร์กับคนทำงานร่วมกันอย่างเข้าใจมากขึ้น เว็บเชิงความหมายจะกำหนดตำแหน่งของข้อมูลด้วย “ความหมาย” ของข้อมูล แทนที่จะทำเครื่องหมายกำกับ



ภาพที่ 2 ส่วนประกอบของอาร์ดีเอฟเอส

เอกสารไว้ที่ “ตัวหนังสือ (Text)” ความแตกต่างด้านการค้นหาด้วยเว็ลด์ไวด์เว็บ (WWW) กับเว็บเชิงความหมายคือ ผลการค้นหาของเว็บเชิงความหมายจะได้ผลลัพธ์เป็นชุดข้อมูลที่มีความหมายเฉพาะตรงกับที่ต้องการเท่านั้น ซึ่งเป็นการย่นเวลาในการค้นหาอย่างมากเมื่อเปรียบเทียบกับผลลัพธ์ที่ได้จากค้นหาด้วย WWW ซึ่งได้ผลลัพธ์เป็นลิสต์ของเว็บไซต์จำนวนมาก นอกจากนี้เว็บเชิงความหมายยังหมายถึงส่วนขยายของเว็บปัจจุบันเพื่อทำให้การใช้ข้อมูลบนเว็บสามารถนำมาใช้ซ้ำ และเอื้อต่อการค้นหาข้อมูลโดยอัตโนมัติ จัดเป็นเทคโนโลยีหนึ่งที่จะช่วยในการจัดเก็บ และนำเสนอเนื้อหาแบบมีโครงสร้างที่ช่วยในการวิเคราะห์ จำแนกหรือจัดแบ่งข้อมูลที่มีความสัมพันธ์กับข้อมูลอื่นในแต่ละระดับ โดยมีเป้าหมาย เพื่อเตรียมการให้คอมพิวเตอร์สามารถที่จะอ่าน และทำความเข้าใจความหมายของคำและความคิดรวบยอดที่ผู้พัฒนากำหนดไว้ โดยยินยอมให้ตัวแทน (Software Agents) สามารถเข้าถึงข้อมูล วิเคราะห์ และประมวลผลข้อมูลได้ ซึ่งเว็บเชิงความหมายจะเป็นวิธีการที่ใช้ในการกำหนดโครงสร้างและเนื้อหาของเว็บ กำหนดสภาพแวดล้อมที่ทำให้ตัวแทนสามารถที่จะทำงานแทน Users ได้ ทำให้ คอมพิวเตอร์จะสามารถเข้าใจและประมวลผลข้อมูลระหว่างกันได้โดยอัตโนมัติ [14]

2.4 ลิงก์คำสำหรับการท่องเที่ยวในประเทศไทย

โดเมนการท่องเที่ยวเป็นโดเมนที่มีความซับซ้อนมากและเปลี่ยนแปลงอยู่ตลอดเวลา ซึ่งผู้ทำการตัดสินใจมักจะพึ่งพาการพยากรณ์ที่จะคาดการณ์ความต้องการในอนาคต หรือใช้ระบบสนับสนุนการตัดสินใจเพื่อวิเคราะห์และเปรียบเทียบ

ผู้มีส่วนได้เสียที่เกี่ยวข้อง สถิติการท่องเที่ยวเช่น จำนวนนักท่องเที่ยวที่เดินทางมาถึง และการเข้าพักที่ปลายทางมีความสำคัญสำหรับอุตสาหกรรมสำหรับการตัดสินใจต่างๆ เพื่อทำงานในส่วนที่เกี่ยวข้องเช่น การทำความเข้าใจผลจากการท่องเที่ยวที่มีต่อเศรษฐกิจที่จุดหมายปลายทาง หรือการส่งเสริมและการตลาดที่จุดหมายปลายทาง โดยการพยากรณ์ความต้องการการท่องเที่ยว การตั้งค่าเป้าหมายทางการตลาด และการสำรวจศักยภาพของตลาด

ประโยชน์ของเทคโนโลยีลิงก์ดาต้าสำหรับการอุตสาหกรรมท่องเที่ยว คือ เป็นเทคโนโลยีที่สามารถเปิดใช้งานโครงสร้างพื้นฐานสำหรับการแบ่งปันข้อมูลการท่องเที่ยวที่จะทำให้การเผยแพร่ข้อมูลทำได้ดีกว่าการในปัจจุบันซึ่งมีการละทิ้งข้อมูลและ API ทำให้สามารถอำนวยความสะดวกต่อการใช้ชุดข้อมูลแบบอัตโนมัติ และแบบไดนามิกในเครื่องมือสนับสนุนการตัดสินใจ นอกจากนี้ยังมีประโยชน์ในด้านการเชื่อมโยงที่จะจัดตั้งขึ้นระหว่างชุดข้อมูล โดยการเชื่อมโยงเหล่านี้อำนวยความสะดวกในการรวมข้อมูลไว้ในจุดเดียว [15]

2.5 งานวิจัยที่เกี่ยวข้อง

อิสรา และคณะ [16] ได้ทำการพัฒนาออนโทโลยีการค้นคืนสารสนเทศการท่องเที่ยวในภาคตะวันออกเฉียงเหนือของประเทศไทยแบบบนลงล่าง ล่างขึ้นบน และแบบผสม (Combination) เพื่อทดสอบประสิทธิภาพว่าออนโทโลยีแบบใดจะมีประสิทธิภาพในการค้นคืนข้อมูลมากที่สุด ออกแบบและพัฒนาออนโทโลยีการท่องเที่ยวด้วยโปรแกรม Protégé รุ่น 4.1 มีออนโทโลยีที่พัฒนาอยู่ในรูปแบบ OWL โดยจำนวนคลาสหลักจะแตกต่างกันออกไปตามรูปแบบออนโทโลยีที่ใช้ในการออกแบบ ผลการวิจัยพบว่า ออนโทโลยีแบบผสมให้ค่าประสิทธิภาพในการค้นคืนมากที่สุด แต่ใช้เวลานานที่สุด และแบบล่างขึ้นบน ให้ค่าประสิทธิภาพการค้นคืนต่ำที่สุด แต่ใช้เวลาน้อยที่สุด

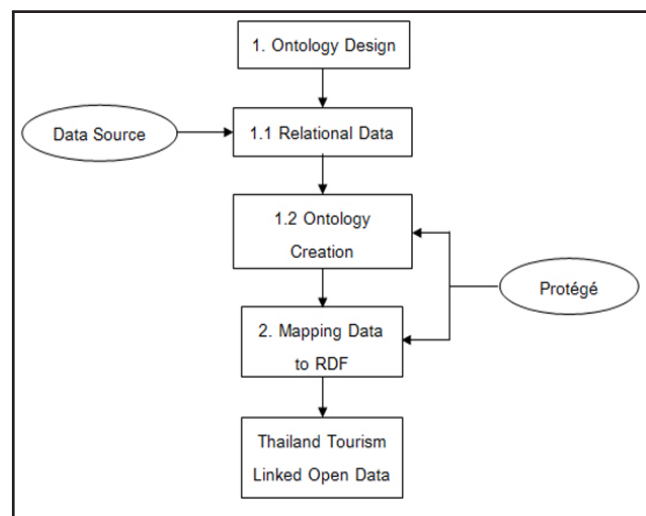
Vohnout และคณะ [17] ได้นำเสนอแนวคิดเกี่ยวกับ SmartTouristData ระบบแบ่งปันข้อมูลการท่องเที่ยวร่วมกันสำหรับสนับสนุนธุรกิจการท่องเที่ยว ข้อมูลที่สามารถนำกลับมาใช้ซ้ำและการแบ่งปันร่วมกัน เป็นวิธีการและเครื่องมือที่จะทำให้จะประหยัดทั้งเวลาและค่าใช้จ่าย เนื่องจากไม่จำเป็นต้องเก็บรวบรวมข้อมูลด้วยตัวเองหรือเสียค่าใช้จ่ายกับข้อมูลเหล่านั้น นอกจากนี้ยังไม่จำเป็นต้องทำการปรับเปลี่ยนหรือ

เปลี่ยนแปลงรูปแบบของข้อมูล เพราะสามารถประมวลผลข้อมูลในรูปแบบใหม่ เช่นการทำเหมืองข้อมูล เป็นระบบที่จะเชื่อมต่อระหว่างกลุ่มหลักของอุตสาหกรรมการท่องเที่ยวด้วยการให้บริการข้อมูลที่มีคุณภาพสูง

Sabou และคณะ [7] เห็นว่าข้อมูลเกี่ยวกับตัวชี้วัดการท่องเที่ยว เช่น ข้อมูลที่ถูกจัดเก็บไว้ในระบบ MIS มีความสำคัญสำหรับผู้มีหน้าที่ตัดสินใจเกี่ยวกับการท่องเที่ยว การท่องเที่ยวเป็นโดเมนที่มีความซับซ้อนมาก และเปลี่ยนแปลงได้ตลอดเวลา นักวิจัยจึงนำเสนอ TourMISLOD ของชุดข้อมูลเชิงสถิติจากระบบ MIS ซึ่งมีแหล่งที่มาของข้อมูลการท่องเที่ยวในยุโรป TourMISLOD มีข้อมูลเกี่ยวกับการมาของนักท่องเที่ยว (Arrivals) จำนวนคืนที่พัก (Bednights) และตัวชี้วัดความสามารถในการรองรับนักท่องเที่ยวที่จุดหมายปลายทาง (Capacity tourism indicators) ซึ่งจัดเก็บข้อมูลตั้งแต่ปี ค.ศ. 1985 จากกว่า 150 เมืองทั่วยุโรปและเชื่อมต่อกับ 19 ตลาดหลัก แต่เป็นเพราะปัญหาลิขสิทธิ์ข้อมูลโดยสมาคม TourMIS การพัฒนาชุดข้อมูลจึงมีข้อมูลที่จำกัด อย่างไรก็ตามตัวต้นแบบนี้ได้้นำออกมาเพื่อการทดสอบ และความพยายามนี้ประสบความสำเร็จเป็นอย่างมาก เป็นประโยชน์สำหรับการสนับสนุนการตัดสินใจ

3. วิธีดำเนินงานวิจัย

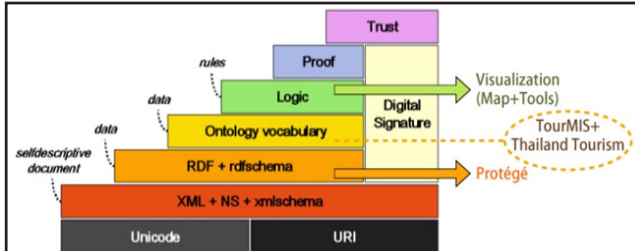
การออกแบบและสร้างโมเดล LOD จากข้อมูลเชิงสถิติด้านการท่องเที่ยว มีแนวคิดของขั้นตอนการออกแบบซึ่งอ้างอิงมาจาก [7] โดยมีขั้นตอนดังภาพที่ 3



ภาพที่ 3 ขั้นตอนการออกแบบและสร้างโมเดล LOD

3.1 สถาปัตยกรรมระบบ

งานวิจัยนี้ออกแบบสถาปัตยกรรมของระบบ Thailand Tourism Linked Data Visualization (TTLDV) โดยใช้หลักการของลิงก์ดาต้า ซึ่งประกอบด้วยชั้นของสถาปัตยกรรมเว็บเชิงความหมาย (Semantic Layers) [18] ภาพที่ 3 แสดงขั้นตอนในแต่ละชั้นของสถาปัตยกรรมเว็บเชิงความหมาย

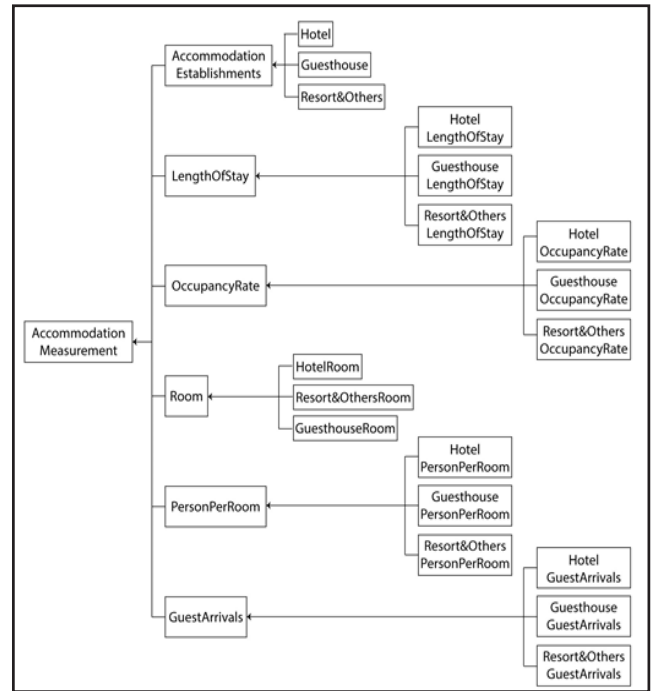


ภาพที่ 4 สถาปัตยกรรมระบบ TTLDV อ้างอิงจากชั้นของสถาปัตยกรรมเว็บเชิงความหมาย(ปรับปรุงจากงานวิจัยของ Koivunen และ Miller [19])

RDF Layer: ทำการ Mapping ข้อมูลเข้ากับลิงก์ดาต้าการท่องเที่ยวแห่งประเทศไทยที่ได้จาก Ontology Layer ด้วยโปรแกรม Protégé

Ontology Layer: ในการออกแบบโครงสร้างข้อมูลของออนโทโลยีที่อธิบายถึงตัวชี้วัดการท่องเที่ยวแห่งประเทศไทย ผู้วิจัยได้ใช้แนวคิดจากออนโทโลยีของ TourMISLOD [7] เป็นรูปแบบเบื้องต้น และเพิ่มจำนวนของตัวชี้วัดที่จำเป็นต่ออุตสาหกรรมการท่องเที่ยวของประเทศไทย เช่น รายได้ และความสามารถในการรองรับของที่พัก ภาพที่ 5 แสดงถึงออนโทโลยีซึ่งแสดงรูปแบบและลักษณะของการวัด (ตัวชี้วัดการท่องเที่ยว)

งานวิจัยนี้พัฒนาออนโทโลยีโดยใช้โปรแกรม Protégé [19] ซึ่งเป็นซอฟต์แวร์ซึ่งเป็นแพลตฟอร์มโอเพนซอร์สสำหรับการจัดการกับออนโทโลยีให้บริการ Graphic User Interface (GUI) ที่ใช้ในการนิยามออนโทโลยี สร้างโมเดลและการประยุกต์ใช้โดเมนความรู้ร่วมกับออนโทโลยีเพื่อกำหนดรูปแบบโดเมนของสถิติการท่องเที่ยวและข้อมูลการสำรวจ รวมทั้งหมด 34 ตัวชี้วัดสำคัญที่ใช้ในปัจจุบันจากการท่องเที่ยวแห่งประเทศไทย (ททท.) และ 17 ตัวชี้วัดเศรษฐกิจและประชากรที่เกี่ยวข้องกับการท่องเที่ยวโดยสำนักงานสถิติแห่งชาติของประเทศไทย ภาพที่ 4 แสดงลำดับชั้นของการพัฒนาแนวคิดออนโทโลยีสำหรับการท่องเที่ยวของไทย การพัฒนานี้ใช้แนวคิดจาก TourMISLOD [7] EUROSTAT [20] และ



ภาพที่ 5 แนวคิดโครงสร้างข้อมูลของออนโทโลยีการท่องเที่ยวแห่งประเทศไทย

ชุดข้อมูลการท่องเที่ยวที่มีอยู่และคำจำกัดความโดยการท่องเที่ยวแห่งประเทศไทย เพื่อรองรับแนวความคิดที่หลากหลาย (Entity, Property, และ Relationship) ผู้วิจัยจึงนำเสนอ 7 งานบทบาทหลัก (ดูแหล่งข้อมูล A) ของการใช้ LOD ในแพลตฟอร์มการทดลองของงานวิจัยนี้ (แอปพลิเคชัน TTLDV) ซึ่งจะช่วยให้ผู้ใช้สามารถทำการเปรียบเทียบข้อมูลหลายชุด การบูรณาการข้อมูล และการวิซวลไลเซชันข้อมูล Logic Layer: ชั้นนี้ของสถาปัตยกรรมทำงานเกี่ยวข้องกับวิธีการวิซวลไลเซชันของข้อมูลการท่องเที่ยวที่อยู่ในรูปแบบ RDF ที่กำหนดไว้บน RDF Layer สำหรับชั้นนี้การพัฒนาของงานวิจัยใช้เซิร์ฟเวอร์ Apache HTTP ในการโฮสต์โปรแกรม PHP และจัดการพื้นที่จัดเก็บข้อมูล RDF / JSON / JSON-LD และพัฒนา Python library สำหรับการประมวลผลข้อมูลสถิติที่ซับซ้อน

3.2 การสร้างข้อมูลแบบเปิด

งานวิจัยนี้เป็นครั้งแรกในประเทศไทยที่มีการนำข้อมูลทางสถิติทั้งหมดจากแหล่งข้อมูลสาธารณะเกี่ยวกับการท่องเที่ยวมาแปลงเป็นให้อยู่ในรูปแบบลิงก์ดาต้าของประเทศ เพื่อวัตถุประสงค์ของการตัดสินใจในอุตสาหกรรมการท่องเที่ยว

แหล่งข้อมูลที่น่ามาใช้ประกอบด้วย 4 เว็บไซต์ ได้แก่



- ข้อมูลการท่องเที่ยว, สำนักงานสถิติแห่งชาติ (สสช) [21]
- ข้อมูลภาครัฐแบบเปิดของประเทศไทย [22]
- การท่องเที่ยวแห่งประเทศไทย (ททท) [23]
- กรมการท่องเที่ยว [24]

ตารางที่ 1 แสดงชุดข้อมูลที่รวบรวมจากเว็บไซต์ของรัฐบาลสำหรับใช้ในแอปพลิเคชัน ในรูปแบบที่ให้บริการของแต่ละเว็บไซต์ ชุดข้อมูลส่วนใหญ่เป็นข้อมูลแบบเปิด แต่อยู่ในรูปแบบ Excel การออกแบบลิงก์ดาต้านี้เพื่อการเข้ารหัสสถิติการท่องเที่ยวไทยที่ได้จากแหล่งข้อมูลที่ระบุไว้ในตารางที่ 1 ในระหว่างการพัฒนาออนไลน์ [9]

ตารางที่ 1 แหล่งที่มาของข้อมูล

Data ^{1,2}	Provenance	Format	Time period
Population	NSO	CSV	Yearly
Province map data	OGD	CSV	Regular updates
Destinations and Divisions ³	TAT	Excel	Regular updates
Tourist Profiles ⁴	TAT	Excel	1998 - present
International Visitor Arrivals ⁵	TAT	Excel	1998 - present
Domestic Visitors	TAT	Excel	1998 - present
Length of Stay	TAT	Excel	2002 - present
Accommodation Establishments	TAT	Excel	2002 - present
Nights Spent	TAT	Excel	2002 - present
Room Rates	TAT	Excel	2002 - present
Various tourism data standards	DOT	PDF	Regular updates

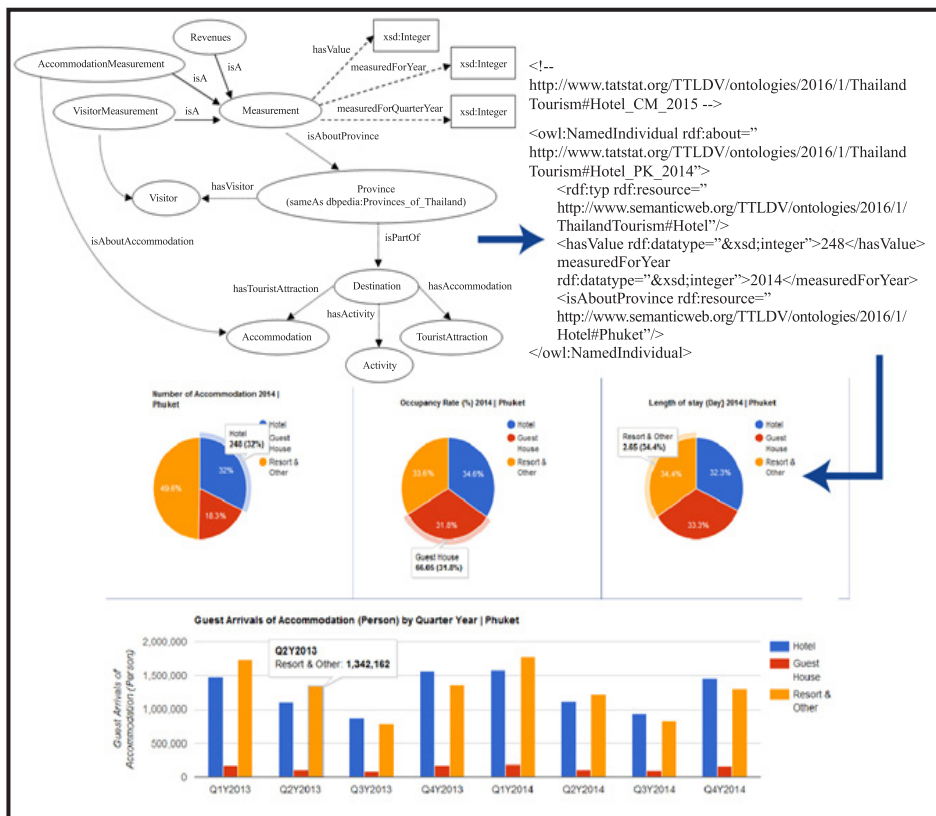
หมายเหตุ:

- 1) ข้อมูลสถิติคัดเลือกโดยข้อกำหนดการใช้งานตามที่กำหนดไว้ในคำศัพท์การท่องเที่ยวของ EUROSTAT [20]
- 2) ข้อมูลส่วนใหญ่อยู่ในช่วงเวลารายเดือน ชุดข้อมูลบางชุดมีการปรับปรุงรายไตรมาส
- 3) การท่องเที่ยวแห่งประเทศไทยกำหนดจุดหมายปลายทางเป็นจังหวัด ยกเว้นพื้นที่ทางภูมิศาสตร์ หรือเมืองที่มีความหนาแน่นสูงของนักท่องเที่ยว เช่น เกาะสมุย กระบี่ และอำเภอหัวหิน ฯลฯ
- 4) ข้อมูลนักท่องเที่ยว เช่น เพศ อายุ วัตถุประสงค์ของการมาเยือน การจัดการการท่องเที่ยว ประเทศที่อยู่อาศัย อาชีพ
- 5) การมาถึงของนักท่องเที่ยวต่างชาติ รวมไปถึงรูปแบบการเดินทาง

3.3 การแปลงข้อมูลเชิงสัมพันธ์เป็นลิงก์ดาต้า

กระบวนการแปลงข้อมูลเชิงสัมพันธ์เป็นลิงก์ดาต้ามีดังนี้: [24], [28]

1. แปลงข้อมูลเชิงสัมพันธ์เป็นออนไลน์: (1) ระบุ “things” (หรือ ประเภทของ Object): ยกตัวอย่างเช่น แต่ละตารางสัมพันธ์จะถูกแปลงให้เป็นประเภทของ Object เช่น “Visitor”,



ภาพที่ 6 ตัวอย่างการจำลองเว็บแอปพลิเคชัน TTLDV

“Destination”, “Measurement” (2) กำหนด Predicates เช่น “v:VisitorArrivals”, “v:Country”, “v:ModeOfTransport” เป็นต้น

2. การสร้าง RDF โดยการจับคู่: การสร้าง Instance กำหนด URI ที่ไม่ซ้ำกัน เช่น แต่ละแถวในตาราง “VisitorArrivals” ถูกจับคู่กับ Instances “V:VisitorArrivals” ตั้งค่าพร้อมกับกำหนด URI ลงท้ายด้วยรหัสของจุดหมายปลายทาง และระยะเวลาของ VisitorArrivals นี้ (คีย์หลักของตาราง)

3. การจัดการข้อมูล RDF ที่ได้: กระบวนการแปลที่มีประสิทธิภาพและความสามารถในการกระบวนการเรียกค้น JSON-LD / SPARQL [25], [26]

ภาพที่ 6 แสดงให้เห็นถึงตัวอย่างกระบวนการวิซวลไลเซชันลิงก์ดาต้าจากข้อมูล RDF การใช้ข้อมูล และทรัพยากรตัวอย่างของจำนวนที่พักรักษาแบ่งแยกตามประเภท จำนวนคืนที่เข้าพักโดยเฉลี่ย จำนวนนักท่องเที่ยวที่มาอยู่ที่พักรักษาแต่ละประเภทของจังหวัดภูเก็ตในแต่ละไตรมาสปี 2013-2014

Generic Task: ผู้วิจัยได้พัฒนาแอปพลิเคชันและทดสอบการเข้าถึงชุดข้อมูลของการท่องเที่ยวไทยในรูปแบบวิซวลไลเซชัน ดังนี้

1. จำนวนนักท่องเที่ยว (Visitor arrivals) จำแนกตามสถานที่ท่องเที่ยว (จังหวัด) และช่วงเวลา (เดือน ไตรมาส หรือปี)

2. จำนวนนักท่องเที่ยวต่างชาติ (International visitor arrivals) จำแนกตามประเทศต้นทาง สถานที่ท่องเที่ยว และช่วงเวลา

3. จำนวนนักท่องเที่ยว (Visitor arrivals) จำแนกตามสถานที่ท่องเที่ยว (จังหวัด) และช่วงเวลา

4. จำนวนนักท่องเที่ยวต่างชาติ (International visitor arrivals) จำแนกตามประเทศต้นทาง สถานที่ท่องเที่ยว และช่วงเวลา

5. ความจุของที่พัก (Capacity of accommodation) จำแนกตามสถานที่และช่วงเวลา

6. รายรับที่ได้จากการท่องเที่ยว และค่าใช้จ่ายของนักท่องเที่ยว จำแนกตามสถานที่และช่วงเวลา

7. รายรับโดยเฉลี่ยจากนักท่องเที่ยวต่อคน (The average revenues per visitor) จำแนกตามสถานที่และช่วงเวลา

8. ทั้งหมดข้างต้น มีตัวเลือกสำหรับแสดงผล 10 และ 20 อันดับสูงสุด

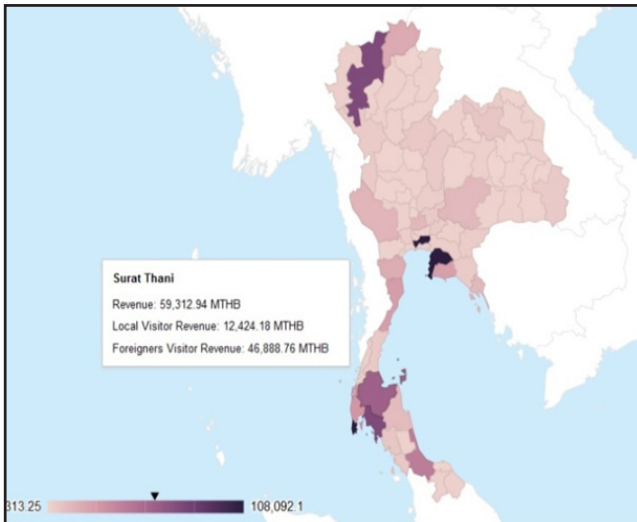
9. ทั้งหมดข้างต้น มีการเปรียบเทียบปีต่อปี

ตัวอย่างเช่น จาก Generic Task ข้อ 4 ภาพที่ 7 แสดงตัวอย่างสามตัวอย่างของ TTLDV ซึ่งแสดงผลการบูรณาการลิงก์ดาต้า เข้ากับ JSON-LD-based เพื่อเรียงลำดับข้อมูล RDF ภาพที่ 7 (ก) เป็นตัวอย่างที่แสดงให้เห็นรายได้ตามสถานที่ (จังหวัดสุราษฎร์ธานี) และช่วงเวลาที่เลือก (ปี 2014) ในการแสดงวิซวลไลเซชันทางภูมิศาสตร์จะใช้ค่าความอ่อนแก่ของสีแทนความมากมายน้อยของรายได้จากการท่องเที่ยวของแต่ละจังหวัด เป็นค่าที่ใช้กันทั่วไปในการวิเคราะห์รายได้จากการท่องเที่ยว ภาพที่ 7 (ข) แสดงสถิตินักท่องเที่ยวในช่วงไฮซีซั่น (ธันวาคมและเมษายนในพื้นที่ที่เลือก เช่น ภาคเหนือ และปีที่น่าสนใจ 2014) และภาพที่ 7 (ค) สถิติของค่าใช้จ่ายในแต่ละวันโดยเฉลี่ยของนักท่องเที่ยว (พื้นที่ที่เลือกเป็น จ.ขอนแก่น ในปี 2014) สองตัวอย่างนี้ใช้การบูรณาการ LOD กับแสดงผลด้วยไลบรารีของภาษา Python

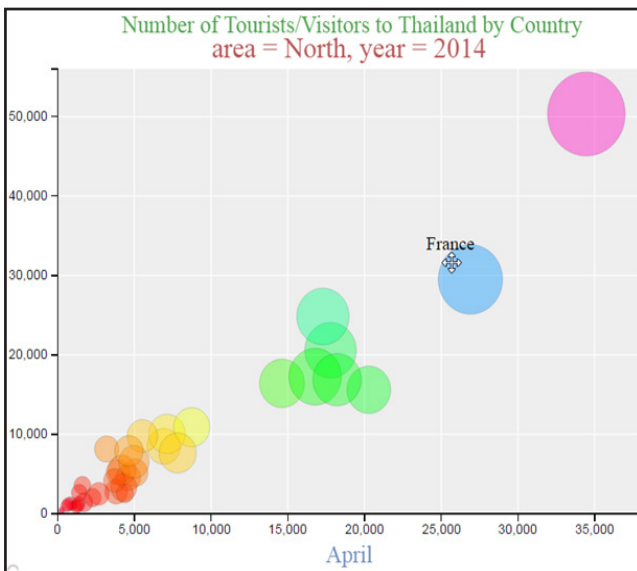
การพัฒนาแอปพลิเคชัน: ผู้วิจัยพัฒนาโดยใช้ Google Charts API [27] และไลบรารีของภาษา Python ในการแสดงแผนภูมิต่างๆ สำหรับข้อมูลสถิติพื้นฐาน (ดังเช่นในภาพที่ 7 (ก)) สร้างขึ้นด้วยภาษา PHP ซึ่งฝัง JavaScript โดยใช้ Google Chart สำหรับข้อมูลสถิติหลายตัวแปรที่ต้องมีตัวเลือกในการกรอง และการขุดเจาะข้อมูลในการวิซวลไลเซชัน (ดังตัวอย่างในภาพที่ 7 (ข) และ 7 (ค) พัฒนาด้วยภาษา Python และไลบรารี PyLD [28] และ Mpd3 [29]

4. บทสรุปและแนวทางในอนาคต

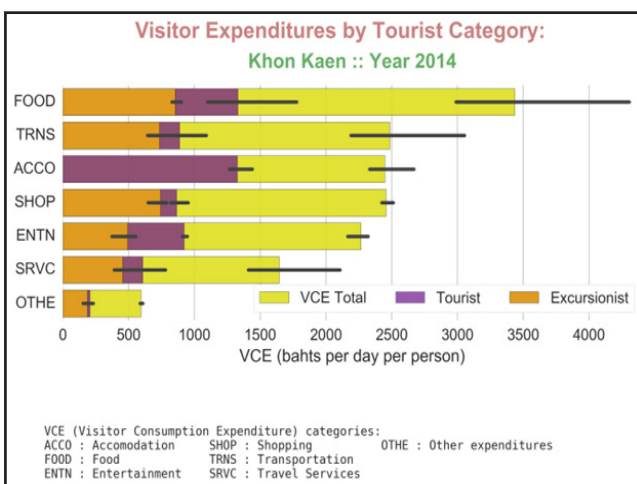
จากแนวทางการสร้างโมเดล LOD ข้อมูลตัวชี้วัดการท่องเที่ยว และพัฒนาเครื่องมือวิซวลไลเซชัน TTLDV โดยมุ่งเน้นการสร้างโมเดล LOD เพื่อสนับสนุนการตัดสินใจของอุตสาหกรรมการท่องเที่ยว ทำการวัดคุณภาพเครื่องมือด้วยการกำหนด General Task ที่เหมาะสมกับการวิซวลไลเซชันข้อมูลการท่องเที่ยว ซึ่งผลที่ได้ก็ออกมาถูกต้องตาม General Task ที่กำหนดไว้ แต่เนื่องจากตัวแบบข้อมูลเชิงสัมพันธ์และการทำ Serialization ของ RDF ที่ใช้ JSON-LD ในการเชื่อมโยงข้อมูลใน TTLDV นั้นอยู่ระหว่างการพัฒนา จึงต้องมีการทำการออกแบบใหม่สำหรับ JSON-LD และ SPARQL นอกจากนี้การสร้างโมเดล LOD ยังต้องอาศัยการพัฒนาออนโทโลยีที่เหมาะสมสำหรับอุตสาหกรรมการท่องเที่ยว ซึ่งมีความสำคัญในการเป็นภาษากลาง และ



(ก) แผนที่แบบมีปฏิสัมพันธ์กับผู้ใช้ (Interactive map) รายได้จากการท่องเที่ยวในแต่ละจังหวัดในปี 2014



(ข) สถิตินักท่องเที่ยวในช่วงไอซีซี (ธันวาคม - เมษายน 2014)



(ค) ค่าใช้จ่ายในแต่ละวันโดยเฉลี่ยของนักท่องเที่ยว (ปี 2014)
ภาพที่ 5 Sample uses of the resulting LOD for tourism data analysis.

เป็นกลไกในการประสานการแลกเปลี่ยนข้อมูลในทุกภาคส่วนของอุตสาหกรรมการท่องเที่ยว เครื่องมือและโมเดล LOD ที่พัฒนาขึ้นในงานวิจัยนี้อาจเป็นเพียงขั้นตอนเริ่มต้น แต่จะเป็นจุดเริ่มต้นของการพัฒนาวิธีการต่างๆ ของการสร้าง LOD ได้ต่อไป ทั้งยังจะช่วยส่งเสริมให้เกิดการพัฒนาลิงก์ดาต้าแบบเปิดเพิ่มมากขึ้นในอนาคต และความร่วมมือในการสนับสนุนนโยบายเศรษฐกิจดิจิทัล

ในอนาคตงานวิจัยจะมี 2 หัวข้อหลัก ได้แก่ (1) การใช้วิธีการต่างๆ ของการสร้างโมเดลที่มีจำนวนชุดข้อมูลเพิ่มมากขึ้น ซึ่งเป็นปัญหาของการเชื่อมโยงข้อมูลที่ซับซ้อนมากขึ้น และขยายการพัฒนาออนไลน์เพื่อให้อุปกรณ์แนวคิด (คลาส) สำหรับชุดข้อมูลใหม่ๆ ของการท่องเที่ยว (2) สำหรับวิชวลไลเซชันกับข้อมูล LOD ผู้วิจัยจะวางแผนการทดลองต่อโดยพัฒนาเครื่องมือให้รองรับตัวชี้วัดที่มากขึ้น และการแสดงผลภาพ (เช่น แผนที่ที่มีโต้ตอบที่หลากหลาย) และพัฒนาพื้นที่จัดเก็บเว็บเชิงความหมาย และบริการซอฟต์แวร์ที่จัดการกับชุดข้อมูลทางสถิติได้ดีมากขึ้น ด้วยมาตรฐาน SDMX (Statistical Data and Metadata eXchange)

5. เอกสารอ้างอิง

- [1] Open Knowledge Foundation. *Open Data Handbook 1.0 (Thai Edition)*. Available Online at <http://www.ega.or.th/th/content/890/877/>.
- [2] C. Bizer. "The Emerging Web of Linked Data." *IEEE intelligent systems*, Vol. 24, Issue 5, pp. 87-92, 2009.
- [3] T. Heath and C. Bizer. "Linked Data: Evolving the Web into a Global Data Space." *Synthesis lectures on the semantic web: Theory and Technology*, Vol. 1, pp. 1-136, 2011.
- [4] F. Bauer and M. Kaltenböck. *Linked open data: The essentials*, Edition mono/monochrom, Vienna, 2011.
- [5] Ministry of Information and Communication Technology. *Thailand Digital Economy*. Available Online at <http://www.mict.go.th/view/1/Digital%20Economy>.
- [6] L. Dwyer and C. Kim. "Destination competitiveness: determinants and indicators." *Current issues in tourism*, Vol. 6, pp. 369-414, 2003.
- [7] M. Sabou and et al. "TourMISLOD: A tourism linked data set." *Semantic Web*, Vol. 4, No. 3, pp. 271-276, 2013.



- [8] H. Song and G. Li. "Tourism demand modelling and forecasting-A review of recent research." *Tourism Management*, Vol. 29, Issue 2, pp. 203-220, 2008.
- [9] D. Frechtling. *Forecasting tourism demand: methods and strategies*. Butterworth Heinemann, Oxford, 2001.
- [10] D. Brickley, R. V. Guha, and B. McBride. *RDF vocabulary description language 1.0: RDF Schema. W3C recommendation (2004)* Available Online at <http://www.w3.org/tr/2004/rec-rdf-schema-20040210>.
- [11] C. Bizer, T. Heath, and T. Berners-Lee. "Linked data-the story so far." *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pp. 205-227, 2009.
- [12] T. B. Lee. *Linked data-design issues*. Available Online at <http://www.w3.org/DesignIssues/LinkedData.html>.
- [13] Semanticweb.org.edu. *Semantic Web*. Available Online at http://semanticweb.org/wiki/Semantic_Web.html.
- [14] M. Sabou and et al. "Supporting tourism decision making with linked data." *Proceedings of the 8th International Conference on Semantic Systems, I-SEMANTICS 2012, USA*, pp. 201-204, 2012.
- [15] I. Chuenta, J. Thongkam, and J. Phuboon-ob. "Information Retrieval in Thai Northeast Travel Utilizing Ontology." *Information Technology Journal*, Vol. 10, Issue 2, pp. 15-25, 2014.
- [16] P. Vohnout and et al. "SmartTouristData approach for connecting local and global tourist information systems." *Proceedings of the IST-Africa 2014 Conference and Exhibition (IST-Africa)*, Le Meridien Ile Maurice, Mauritius, pp. 1-6, 2014.
- [17] M. Koivunen and E. M. Miller. "W3C Semantic Web activity." *Proceedings of the Semantic Web Kick-off Seminar in Finland*, Finland, pp. 27-44, 2001.
- [18] Stanford University. *Protégé*. Available Online at <http://protege.stanford.edu/>.
- [19] EUROSTAT. *Tourism Glossary*. Available Online at http://ec.europa.eu/eurostat/statistics-explained/index.php/Category:Tourism_glossary.
- [20] National Statistics Office of Thailand. *Dataset: tourism*. Available Online at <http://service.nso.go.th/nso/web/statseries/statseries23.html>.
- [21] Open Government Data of Thailand. *Dataset: tourism*. Available Online at <https://data.go.th/Datasets.aspx?kw=tourism>.
- [22] Tourism Authority of Thailand. *Tourism Statistics*. Available Online at http://www2.tat.or.th/stat/web/static_index.php.
- [23] Department of Tourism, *Statistics*. Available Online at <http://www.tourism.go.th/home/listcategory/11/217>.
- [24] JSON-LD. *JSON-LD - JSON for Linking Data*. Available Online at <http://json-ld.org/>.
- [25] W3C. *SPARQL Query Language for RDF*. Available Online at <https://www.w3.org/TR/rdf-sparql-query/>.
- [26] Google Developers. *Google Visualization API Reference*. Available Online at <https://developers.google.com/chart/interactive/docs/reference>.
- [27] Digital Bazaar, Inc. *PyLD: JSON-LD processor written in Python*. Available Online at <https://github.com/digitalbazaar/pyld>.
- [28] mpld3 developers. *MPLD3 -Bringing Matplotlib to the Browser*. Available Online at <http://mpld3.github.io/>.