



การสร้างแฟ้มข้อมูลรายการหลักฐานโดยอัตโนมัติจากฐานข้อมูลบรรณานุกรม ด้วยการใช้ Faceted Application of Subject Terminology (FAST)

Automatically creating Authority Files from bibliographic databases using the Faceted Application of Subject Terminology (FAST)

ยุทธนา เจริญรื่น¹Yuttana Jaroenruen¹

(Received: May 16, 2019; Revised: May 27, 2019; Accepted: June 19, 2019)

บทคัดย่อ

งานวิจัยชิ้นนี้มีวัตถุประสงค์ คือ (1) ศึกษาและวิเคราะห์รูปแบบการทำรายการหลักฐาน (Authority files) ตามมาตรฐาน MARC21 format (2) เชื่อมโยงระเบียบบรรณานุกรมในฐานข้อมูล WalaiAutoLib กับรายการหลักฐานในรูปแบบ FAST และ (3) เพื่อสร้าง Lookup file ที่เชื่อมโยงกับรายการหลักฐานในรูปแบบ FAST กับฐานข้อมูล WalaiAutoLib โดยอัตโนมัติ วิธีการศึกษาประกอบด้วย (1) วิเคราะห์เอกสารที่เป็นมาตรฐานโครงสร้าง MARC21 และ FASTlinked data (2) วิเคราะห์โครงสร้างข้อมูลจากฐานข้อมูลระบบห้องสมุดอัตโนมัติ WALAI AutoLib โดยใช้ข้อมูลย้อนหลัง 5 ปี ของศูนย์บรรณสารและสื่อการศึกษา มหาวิทยาลัยวลัยลักษณ์ จำนวน 40,000 ระเบียบเป็นข้อมูลตัวอย่าง และ (3) สร้างกลไกเพื่อเชื่อมต่อกับแหล่งข้อมูลที่เป็น Facet Application 8 แหล่ง เพื่อทดสอบการปรับปรุงรายการหลักฐานในรูปแบบ FAST โดยอัตโนมัติ

ผลการศึกษาพบว่า สามารถกำหนดกลไกในการสร้าง Lookup files ของข้อมูลรายการหลักฐาน 3 กลุ่ม ได้แก่ ชื่อ (Names) หัวเรื่อง (Subjects) และชื่อชุด (Series) (2) สามารถเชื่อมต่อกับแหล่งข้อมูลในรูปแบบ FAST และเปรียบเทียบข้อมูลรายการหลักฐานได้โดยอัตโนมัติ โดยจำแนกออกเป็น Names 72,506 เซตข้อมูล Subjects 127,322 เซตข้อมูล และ Series 7,729 เซตข้อมูล (3) สามารถสร้าง Lookup files ที่เป็น FAST Facet Application และอัปเดตลงในฐานข้อมูลโดยอัตโนมัติ ได้แก่ Name 44,393 รายการ Subject 34,208 รายการ และ Series 37,666 รายการ ซึ่งผลการศึกษาในครั้งนี้จะนำไปใช้เป็นพื้นฐานในการพัฒนาชุดข้อมูล (data set) สำหรับการพัฒนา Linked data สำหรับศัพท์ควบคุมภาษาไทยเป็นลำดับถัดไป

คำสำคัญ: ระเบียบบรรณานุกรม ระเบียบรายการหลักฐาน แฟซิท ลิงค์ เดต้า

ABSTRACT

This research aimed to (1) explore and analyze authority files based on MARC21 format, (2) to connect bibliographic data in the WalaiAutoLib database with schema of Faceted Application of Subject Terminology (FAST), and (3) to create Lookup files that associate records in the WalaiAutoLib database with FAST linked data. The steps of this research consisted of (1) document analysis for MARC21 structure, standard and Facet linked data, (2) analyzing the data structure from WALAI AutoLib database system; using 40,000 examples of bibliographic record from CLM Walailak University, and (3) coding Linked data scripts to connect to 8 Data Sources of the Facet Application to automatically update the Authority files.

¹สำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์

²School of Informatics, Walailak University



The results of the study revealed that (1) there were three types of authority files: Name, Subject, and Series. The study also found that the files can be linked to FAST's Data Sources and were automatically compared, categorized by Name 72,506 fields, Subject 127,322 fields and Series 7,729 fields. Furthermore, this study could automatically create 13 sets of lookup files and automatically update them into the MARC21 record (Name 44,393 fields, Subject 34,208 fields and Series 37,666 fields. Therefore, the results of this study will be used as a basis for developing the data sets for Thai control vocabulary linked data in the future.

Keywords: Bibliographic record, Authority record, Facet, Linked data

บทนำ

การให้หัวเรื่องแบบหอสมุดรัฐสภาอเมริกัน หรือ The Library of Congress Subject Headings (LCSH) เป็นเครื่องมือที่ได้รับการยอมรับที่สุดและใช้กันอย่างแพร่หลายที่สุดในห้องสมุดสถาบันอุดมศึกษา (Hjorland, 2016) จนอาจกล่าวได้ว่า LCSH คือเครื่องมือสากลในการจัดการศัพท์ควบคุม (controlled vocabulary) และเป็นตัวแบบตั้งต้นในการพัฒนาระบบในการให้หัวเรื่อง (Subject Heading) ของห้องสมุดสถาบันอุดมศึกษาในระดับนานาชาติ (Childress & Vizine-Goetz, 2017) อย่างไรก็ตาม ด้วยความซับซ้อนและกฎเกณฑ์ที่เจาะจงของตัวคู่มือ LCSH เอง ทำให้การควบคุมหัวเรื่องเป็นงานที่เหมาะสมกับบรรณารักษ์หรือนักทำรายการ (cataloger) ที่มีทักษะและประสบการณ์สูงเท่านั้น

สาเหตุที่เป็นเช่นนั้นเนื่องมาจาก การควบคุมหัวเรื่องเป็นกระบวนการที่ต้องอาศัยผู้เชี่ยวชาญในการวิเคราะห์ วินิจฉัย และตัดสินใจในการใช้ศัพท์ควบคุม (controlled vocabulary) ให้ถูกต้องเพื่อการนำผลการตัดสินใจนั้นมาใช้ในการควบคุมหัวเรื่อง โดยอาศัยกลไกการควบคุมรายการหลักฐาน (automated authority control) เป็นมาตรฐานการควบคุมอย่างไรก็ตามกลไกหรือฟังก์ชันเชิงระบบ (functional system) ในการควบคุมรายการหลักฐานนั้น ยังถือว่ามีความจำกัดอยู่มากสำหรับระบบห้องสมุดอัตโนมัติที่ใช้กันอยู่ในปัจจุบัน (Library of Congress, 2012) ดังนั้น เพื่อให้กระบวนการจัดการข้อมูลของรายการหลักฐานในระบบมีความถูกต้อง ห้องสมุดสถาบันอุดมศึกษาจึงอาศัยเครื่องมือจากมอดูลควบคุมรายการหลักฐาน (Authority control module) ซึ่งเป็นระบบงานย่อยระบบหนึ่งของระบบห้องสมุดอัตโนมัติมาเป็นเครื่องมือหลัก (วิทยา เทวรังษี, คีวนาถ นันทพิชัย และยุทธนา เจริญรัตน์, 2557) เครื่องมืองดดังกล่าวนี้ประกอบด้วยฟังก์ชันและข้อเสนอแนะ ที่ทำหน้าที่ควบคุมและจัดการความสัมพันธ์ (relation) ของข้อมูลประเภทต่างๆในระเบียบบรรณานุกรม เช่น ชื่อ (Name) หัวเรื่อง (Subject) และชื่อชุด (Series) โดยข้อมูลดังกล่าวจะถูกจัดเก็บเป็นระเบียบรายการหลักฐาน (Authority record) ตามโครงสร้างของ MARC21 Format for Authority Data (Library of Congress, 2019) นอกจากนี้แล้ว การจัดทำระเบียบรายการหลักฐานจำเป็นต้องมีการเชื่อมโยง (linking) การใช้ความสัมพันธ์ประเภท see หรือ see also ตลอดจนการเชื่อมโยงไปยังระบบภายนอก (external system) เพื่อแสดงผลข้อมูลที่เกี่ยวข้องกับคำที่เป็นศัพท์ควบคุมนั้นๆ ซึ่งเป็นงานที่ต้องอาศัยผู้เชี่ยวชาญด้านการควบคุมรายการหลักฐานโดยเฉพาะ และเป็นปัญหาสำหรับห้องสมุดที่ขาดแคลนผู้เชี่ยวชาญในด้านดังกล่าว (Walailak University, 2018)

เมื่อพิจารณาในแง่ของการจัดเก็บและค้นคืนประกอบกันไปด้วยจะพบว่าการสืบค้นระเบียบบรรณานุกรมตลอดจนรายการหลักฐานของทรัพยากรสารสนเทศห้องสมุดสถาบันอุดมศึกษาผ่านเครือข่ายอินเทอร์เน็ต ยังมีประเด็นปัญหาลำคัญสำหรับผู้ใช้งานไปอยู่ เนื่องจากโดยปกติแล้วระเบียบบรรณานุกรมในระบบห้องสมุดจะไม่สามารถถูกนำมาแสดงผลผ่านกลไกการสืบค้น (search engine) บนอินเทอร์เน็ตได้ สาเหตุสำคัญมาจากเมทาดาทาที่ใช้เก็บข้อมูลบรรณานุกรม เช่น MARC21 นั้นจะถูกเก็บอยู่ในฐานข้อมูลของระบบห้องสมุดอัตโนมัติในรูปแบบของระเบียบฐานข้อมูล และไม่สามารถเข้าถึงผ่านกลไกการสืบค้นทั่วไป

ได้ด้วยวิธีปกติ นอกเสียจากผู้ใช้ทำการสืบค้นผ่านหน้า Web OPAC ของระบบห้องสมุดเท่านั้น

ปัญหาหรือข้อจำกัดที่กล่าวมาแล้วในข้างต้นถือเป็นผลเสียต่อการบริการของห้องสมุดในยุคดิจิทัลโดยตรง เนื่องจากการปิดกั้นช่องทางในการเข้าถึงข้อมูล และไม่ตอบสนองกับพฤติกรรมของผู้ใช้ในบริบทของยุคดิจิทัล ดังนั้น องค์กรตลอดจนหน่วยงานระดับสากลที่เกี่ยวข้องกับระบบห้องสมุดจึงมีข้อตกลงร่วมกันในการสร้างมาตรฐาน (standards) โครงสร้างข้อมูล (data structures) กลไก (mechanism) ตลอดจนเกณฑ์วิธี (protocols) เพื่อให้ผู้ใช้สามารถเข้าถึงข้อมูลบรรณานุกรมของทรัพยากรสารสนเทศโดยไม่ต้องผ่านหน้าจอ Web OPAC ซึ่งวิธีที่ได้รับการยอมรับว่ามีประสิทธิภาพที่สุด คือการเข้าถึงชุดข้อมูลผ่าน linked data เนื่องจากข้อมูลในรูปแบบดังกล่าวใช้เพียงคุณสมบัติพื้นฐานของ Web โดยไม่ต้องอิงกับเทคโนโลยีอื่น เช่น ระบบฐานข้อมูลหรือระบบห้องสมุดอัตโนมัติ ซึ่งชุดข้อมูลรายการหลักฐานที่เป็น linked data ที่ใหญ่ที่สุดในโลกตอนนี้คือ FAST ของ OCLC (Gonzales, 2014)

Faceted Application of Subject Terminology หรือ FAST เป็นชุดของคำศัพท์ที่ได้มาจากหัวเรื่องของหอสมุดรัฐสภาอเมริกัน (Library of Congress Subject Headings - LCSH) โดยมีวัตถุประสงค์ในการพัฒนาขึ้นเพื่อเป็นเค้าร่าง (schema) ของชุดคำศัพท์ที่สามารถนำไปใช้งานได้ง่ายสำหรับบรรณารักษ์ ตลอดจนบุคคลทั่วไปที่ไม่ได้เป็นผู้เชี่ยวชาญด้านการควบคุมรายการหลักฐานซึ่งเค้าร่างดังกล่าวนี้ห้องสมุดหรือสถาบันสารสนเทศสามารถที่จะเชื่อมโยงและนำไปใช้งานได้โดยอัตโนมัติ โดยผ่านฟังก์ชันของ linked data ของระบบห้องสมุดอัตโนมัติที่สถาบันใช้งานอยู่

ดังนั้น ผู้วิจัยจึงมีแนวคิดในการศึกษาการประยุกต์ใช้ FAST กับการสร้างและควบคุมรายการหลักฐานของระบบบรรณานุกรม เพื่อให้ผู้ใช้ทั่วไปสามารถนำชุดคำศัพท์ของ FAST ไปใช้ในการสร้างและควบคุม Facet ผ่านกลไกการทำงานของ linked data และทำให้ห้องสมุดที่ขาดแคลนบรรณารักษ์ที่เป็นนักทำรายการหรือนักทำดัชนี (indexer) ยังสามารถที่จะจัดทำ Facet รายการหลักฐานที่เป็น ชื่อ (Name) ชื่อเรื่อง (Title) หัวเรื่อง (Subject) และศัพท์ควบคุม (Control vocabulary) สำคัญอื่นๆ ที่สามารถใช้เป็นช่องทางในการเข้าถึงสารสนเทศ โดยใช้วิธีการเชื่อมโยงจากแหล่งข้อมูลภายนอก ที่ให้บริการเชื่อมโยงรายการหลักฐานระดับนานาชาติ เช่น LCSH (Library of Congress, 2017), OCLC FAST (Online Computer Library Center, 2018) หรือ VIAF (VIAF, 2018) ได้อย่างสะดวกและมีประสิทธิภาพยิ่งขึ้น

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาและวิเคราะห์รูปแบบการทำรายการหลักฐาน (Authority file) ตามรูปแบบ MARC21
2. เพื่อเชื่อมโยงระบบบรรณานุกรมในฐานข้อมูลระบบ WalaiAutoLib กับรายการหลักฐานในรูปแบบ Faceted Application of Subject Terminology (FAST)
3. เพื่อสร้างแฟ้มข้อมูลอ้างอิง (Lookup file) ที่เชื่อมโยงกับรายการหลักฐานในรูปแบบ FAST สำหรับรายการหลักฐานของข้อมูลบรรณานุกรมในฐานข้อมูล WalaiAutoLib โดยอัตโนมัติ

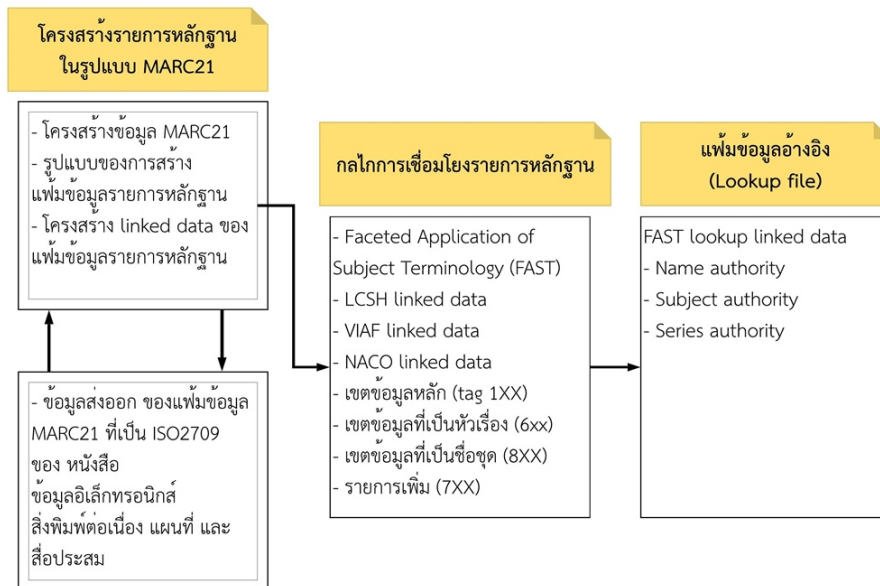
ระเบียบวิธีวิจัย

1. วิเคราะห์เอกสาร (Document Analysis) เพื่อทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับรูปแบบ โครงสร้าง และมาตรฐานของการทำรายการหลักฐานในรูปแบบ MARC21 และการประยุกต์ใช้ FAST ในการทำรายการหลักฐานทั้งที่เป็นภาษาไทยและภาษาต่างประเทศ
2. วิเคราะห์โครงสร้างข้อมูลจากฐานข้อมูลระบบห้องสมุดอัตโนมัติ WALAI AutoLib โดยใช้ข้อมูลของศูนย์บรรณสารและสื่อการศึกษา มหาวิทยาลัยวลัยลักษณ์เป็นข้อมูลตัวอย่าง จำนวน 40,000 ระเบียบ เพื่อสร้างตาราง mapping ระหว่าง



MARC21 Authority tags และ FAST linked data โดยกลไกการเปรียบเทียบระเบียบประกอบด้วย (1) ทำเชื่อมต่อไปยัง Data Sources ที่กำหนดไว้ โดยใช้สคริปต์ linked data tools ที่เขียนโดยภาษา Python และใช้โปรแกรมประยุกต์ MARC Editor เป็น front-end ในการทำงาน (2) เปรียบเทียบข้อมูลในเขตข้อมูลควบคุม (1XX, 7XX, 6XX, 8XX) หากข้อมูลตรงกับ Data Sources แหล่งใดแหล่งหนึ่ง ให้ทำการ mapping ข้อมูล โดยเพิ่ม \$0 - Authority record control number or standard number และ \$1 - Real World Object URI ในทุกเขตข้อมูลที่เป็นเขตข้อมูลควบคุม (3) แสดงผลการเปรียบเทียบระเบียบ จำแนกตามเขตข้อมูลควบคุมที่เป็น Name, Subject และ Series โดยรายละเอียดและผลการเปรียบเทียบข้อมูล

3. ทำการเชื่อมต่อกับแหล่งข้อมูล (Data Sources) ของ Facet Application จำนวน 8 แหล่ง กับโปรแกรมระบบห้องสมุดอัตโนมัติ WalaiAutoLIB เพื่อทดสอบการปรับปรุงรายการหลักฐานที่เป็น FAST ในรูปแบบอัตโนมัติ ซึ่งขั้นตอนในการศึกษาทั้ง 3 ขั้นที่กล่าวมาข้างต้น สามารถแสดงเป็นกรอบแนวคิดในการวิจัยได้ ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดในการวิจัย

ผลการวิจัย

สามารถสรุปผลได้เป็น 3 ประเด็นหลัก ประกอบด้วย 1) รูปแบบ โครงสร้าง และมาตรฐานของการทำรายการหลักฐานในรูปแบบ MARC21 2) กลไกในการเชื่อมโยงระเบียบบรรณานุกรมในฐานะข้อมูลระบบ WalaiAutoLib กับรายการหลักฐานในรูปแบบ FAST และ 3) ได้แฟ้มข้อมูลอ้างอิง (Lookup file) ที่เชื่อมโยงกับรายการหลักฐานในรูปแบบ FAST ที่สามารถใช้รวมกันกับฟังก์ชัน linked data ของระบบห้องสมุดอัตโนมัติโดยมีรายละเอียดของผลการศึกษาดังนี้

1. รูปแบบ โครงสร้าง และมาตรฐานของการทำรายการหลักฐานในรูปแบบ MARC21 มีลักษณะดังนี้

ในด้านรูปแบบของแฟ้มข้อมูล รายการหลักฐานในรูปแบบ MARC21 หรือ MARC21 Authority record เป็นระเบียบรายการที่อ้างอิงตามข้อกำหนดที่เป็นมาตรฐานด้านรายการหลักฐานของหอสมุดรัฐสภาอเมริกัน (Library of Congress, 2015) ถูกสร้างขึ้นให้มีรูปแบบเป็นแฟ้มข้อมูลอักษรที่มีโครงสร้างสำคัญ 3 ส่วน คือ ส่วนนำระเบียบ (Leader) ส่วนนามานุกรมข้อมูล (Directory) และส่วนเขตข้อมูลที่มีความยาวไม่คงที่ (Variable fields) ส่วนรูปแบบที่เป็นเนื้อหาภายใน

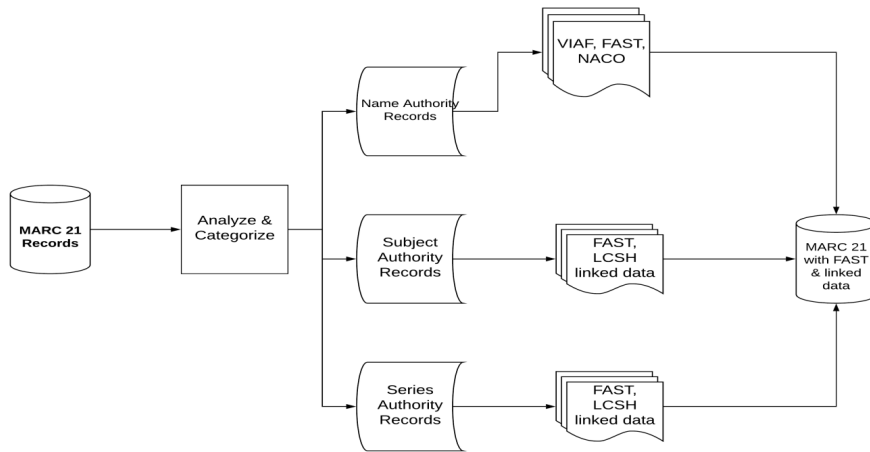
ระเบียบรายการหลักฐานนั้น จะถูกระบุเนื้อหาตามหัวเรื่อง (Heading) ซึ่งจำแนกออกเป็น names, name/title combinations, uniform titles, และ subjects

ในด้านโครงสร้างนั้น รายการหลักฐานในรูปแบบ MARC21 ประกอบด้วยโครงสร้าง 3 ส่วนดังที่กล่าวมาแล้ว และมีรายละเอียดในแต่ละส่วน คือ 1) องค์ประกอบของส่วนป้ายระเบียบจะเป็นข้อมูลที่เป็นตัวเลข หรือคำการเข้ารหัสและมีการระบุตำแหน่งอักขระของข้อมูลในระเบียบ ส่วนป้ายระเบียบนั้นจะมีความยาวคงที่ คือ 24 อักขระ ซึ่งใช้สำหรับการอ่านและประมวลระเบียบด้วยคอมพิวเตอร์ 2) องค์ประกอบของส่วนนามาุกรมข้อมูลมีโครงสร้างเป็นเป็นชุดรายการที่ประกอบด้วยแท็ก (tag) ความยาว (length) และตำแหน่งเริ่มต้น (start portion) ของแต่ละเขตข้อมูลภายในระเบียบ ส่วนนามาุกรมข้อมูลจะมีความยาว 12 อักขระเสมอ และสร้างขึ้นโดยอัตโนมัติจากระบบคอมพิวเตอร์ และ 3) ส่วนเขตข้อมูลที่มีความยาวไม่คงที่ เป็นเขตข้อมูลที่กำกับตัวข้อมูลด้วยเลขประจำเขตข้อมูล 3 อักขระ หรือหมายเลขประจำแท็กที่ถูกเก็บไว้ในส่วนนามาุกรมข้อมูล และแต่ละเขตข้อมูลจะลงท้ายด้วยอักขระตัวยกเลิกเขตข้อมูล (field terminator character) เสมอ ในแต่ละเขตข้อมูลจะประกอบด้วยเนื้อหาข้อมูลที่มีความยาวไม่คงที่ขึ้นอยู่กับเนื้อหาในแต่ละระเบียบ

ด้านมาตรฐานในการทำรายการหลักฐานในรูปแบบ MARC21 ในห้องสมุดสถาบันอุดมศึกษาในประเทศไทยนั้น จะยึดตาม LC MARC Authority Records (Library of Congress, 2015) และ ข้อตกลงในการทำรายการของเครือข่ายสหบรรณานุกรมห้องสมุดสถาบันอุดมศึกษาไทย หรือ UCTAL เป็นเกณฑ์ (Walailak University, 2018)

2. ผลการวิเคราะห์โครงสร้างข้อมูลจากฐานข้อมูลระบบห้องสมุดอัตโนมัติ WALAI AutoLib กับรายการหลักฐานในรูปแบบ FAST เป็นผลการศึกษาที่เกิดจากการนำข้อมูลบรรณานุกรมของศูนย์บรรณสารและสื่อการศึกษา มหาวิทยาลัยวลัยลักษณ์จำนวน 40,000 ระเบียบ ซึ่งถูกส่งออก (export) ในรูปแบบ ISO2709 (Sahu, 2002) ข้อมูลตัวอย่างดังกล่าว เป็นระเบียบที่มีการจัดทำตั้งแต่ปี พ.ศ. 2556 - พ.ศ. 2561 หรือ 5 ปีย้อนหลังนับจากวันที่ส่งออกข้อมูล โดยสามารถจำแนกประเภท (Type of record) ตามที่ปรากฏใน MARC Leader ออกเป็น (1) หนังสือ 31,114 ระเบียบ (2) แฟ้มข้อมูลคอมพิวเตอร์ 1300 ระเบียบ (3) สิ่งพิมพ์ต่อเนื่อง 2,185 ระเบียบ (4) แผนที่ 1,008 ระเบียบ (5) สื่ออิเล็กทรอนิกส์ 396 ระเบียบ และ (6) สื่อประสม 3,997 ระเบียบ

การจัดกลุ่มเขตข้อมูล (tag) ที่ใช้ในการทำรายการหลักฐานเป็นกระบวนการที่ทำหลังจากตรวจสอบความถูกต้องของโครงสร้าง ISO2709 ของระเบียบ MARC21 ที่ถูกส่งออกจากระบบ โดยใช้โปรแกรมประยุกต์เพื่อควบคุมรายการหลักฐานหัวเรื่องสำหรับฐานข้อมูลสหบรรณานุกรมสถาบันอุดมศึกษาไทย (ศิวานถ หนักพิชัย, พรพิมล วัชรกุล และณรงค์ ทองรักจันทร์, 2557) ผลลัพธ์ที่ได้จะเป็นชุดข้อมูลที่เกิดจากการจัดกลุ่ม tag เพื่อใช้ในการตรวจสอบรายการหลักฐานนั้น และนำมาประมวลผลด้วยภาษาสคริปต์ที่พัฒนาขึ้น ให้เป็นไปตามการจัดกลุ่ม Facet ของ FAST ที่จำแนกออกเป็นชุดข้อมูลของ PersonalNames, Corporate Names, Event, Uniform Titles, Chronological, Topical, Geographic และ Form/Genre รวมทั้งสิ้น 8 กลุ่ม (Bennett, O Neill, & Kammerer, 2016) จากนั้นจึงทำการ mapping เขตข้อมูลในระเบียบบรรณานุกรมแต่ละรายการของศูนย์บรรณสารและสื่อการศึกษา มหาวิทยาลัยวลัยลักษณ์ เข้าไปในกลุ่มที่เป็นไปตามข้อกำหนดของ FAST กลไกในการศึกษาปรากฏตามรายละเอียดในภาพที่ 2



ภาพที่ 2 ขั้นตอนการส่งออก วิเคราะห์ และเปรียบเทียบระเบียบบรรณานุกรม

หลังจากวิเคราะห์เปรียบเทียบแล้ว ขั้นตอนต่อไปคือการนำผลการวิเคราะห์ไปจัดกลุ่มตามกลุ่มข้อมูลของระเบียบรายการหลักฐาน (Authority record) ซึ่งมีทั้งสิ้น 4 กลุ่ม ได้แก่ เขตข้อมูลหลัก (tag 1XX) เขตข้อมูลที่เป็นหัวเรื่อง (6xx) เขตข้อมูลที่เป็นชื่อชุด (8XX) และรายการเพิ่ม (7XX) จากนั้นจึงใช้ภาษาสคริปต์ทำการเชื่อมต่อกับชุดข้อมูล (Data Sources) ที่ให้บริการรายการหลักฐานในรูปแบบ linked data โดยเลือกแหล่งเชื่อมต่อไปยังผู้ให้บริการซึ่งเป็นองค์กรหรือหน่วยงานระดับสากลและเป็นเจ้าของหรือผู้กำหนดเกณฑ์หรือโครงสร้างของข้อมูลที่ปรากฏอยู่ในรายการหลักฐานทั้ง 4 กลุ่ม เช่น เขตข้อมูลที่เป็นหัวเรื่องจะเชื่อมต่อกับ Web service ของหอสมุดรัฐสภาอเมริกันและเครือข่าย OCLC ซึ่งเป็นเครือข่ายของผู้ให้บริการด้านระเบียบบรรณานุกรมและระบบห้องสมุดที่ใหญ่ที่สุดในโลก (Childress & Vizine-Goetz, 2017) โดยรายละเอียดของ Data Sources หรือแหล่งข้อมูลที่ใช้ในการเชื่อมต่อทั้งหมด สามารถแสดงรายละเอียดได้ ตามตารางที่ 1

ตารางที่ 1 Data Sources ที่ใช้ในการเชื่อมต่อเพื่อตรวจสอบรายการหลักฐาน

Name	Data sets	URL
LCSH	ชุดข้อมูลที่เป็น Open data และ Linked data ของ Library of Congress Subject Headings	http://id.loc.gov/authorities/subjects.html
LCTGM	ชุดข้อมูลที่เป็น Open data และ Linked data ของ Thesaurus of Geographic Materials	http://id.loc.gov/vocabulary/graphicMaterials.html
LCAC	ชุดข้อมูลที่เป็น Open data และ Linked data ของ Authorities & Controlled Vocabularies	http://id.loc.gov/authorities/names.html
VIAF	ชุดข้อมูลที่เป็นนามบุคคลของ The Virtual International Authority File	https://viaf.org/
ULAN	ชุดข้อมูลที่เป็นนามบุคคลและชื่อศิลปินผู้สร้างสรรค์ผลงานที่เป็น Open data ของ Union List of Artist Names	http://www.getty.edu/research/tools/vocabularies/ulan/
NACO	ชุดข้อมูลที่เป็น Open data และ Linked data ของ LC - Name Authority Cooperative Program	https://www.loc.gov/aba/pcc/naco/

ตารางที่ 1 Data Sources ที่ใช้ในการเชื่อมต่อเพื่อตรวจสอบรายการหลักฐาน (ต่อ)

Name	Data sets	URL
MESH	ชุดข้อมูลหัวเรื่องควบคุมทางการแพทย์ (Medical Subject Headings)	https://meshb.nlm.nih.gov/search
FOAF	ชุดข้อมูลที่เป็น Linked data ของ Friend of a Friend vocabulary	https://lov.linkeddata.es/dataset/lov/vocabs/foaf

โดยผลการเชื่อมต่อ Data Sources เพื่อนำรายการเขตข้อมูลที่ตรงกัน มาทำการอัปเดตเพิ่มข้อมูลหลักฐาน (Authority files) ได้โดยอัตโนมัติปรากฏตามตารางที่ 2

ตารางที่ 2 ผลการเปรียบเทียบและรายการเขตข้อมูลที่มีการอัปเดต Authority files โดยอัตโนมัติ

Types	Tags ID	Fields	Original	Lookup
Name (72,147)	100	Main Entry - Personal Name	32,839	19,366
	110	Main Entry - Corporate Name	3,056	1,181
	111	Main Entry - Meeting Name	985	368
	700	Added Entry - Personal Name	28,712	20,060
	710	Added Entry - Corporate Name	6,591	3,319
	711	Added Entry - Meeting Name	323	99
Subject (103,890)	600	Subject Added Entry - Personal Name	5,286	734
	610	Subject Added Entry - Corporate Name	2,942	459
	611	Subject Added Entry - Meeting Name	49	14
	630	Subject Added Entry - Uniform Title	574	380
	650	Subject Added Entry - Topical Term	107,736	30,534
	651	Subject Added Entry - Geographic Name	10,244	2,079
	655	Index Term - Genre/Form	491	8
Series (6,207)	800	Series Added Entry - Personal Name	58	0
	810	Series Added Entry - Corporate Name	5	0
	811	Series Added Entry - Meeting Name	0	0
	830	Series Added Entry - Uniform Title	7,666	3,458
Total of Tags			207,557	82,059

3. การสร้างแฟ้มข้อมูลอ้างอิง (Lookup file) ที่เชื่อมโยงกับ FAST สำหรับรายการหลักฐานของข้อมูลบรรณานุกรมในฐานข้อมูล WalaiAutoLib เป็นกระบวนการที่นำเทคโนโลยีด้านการสืบค้น 2 อย่างมาประยุกต์ใช้ และทำการอัปเดตรายการหลักฐานในระบบโดยอัตโนมัติได้แก่ (1) Autosuggest technology ที่ใช้ในการสืบค้นข้อมูลผ่านเว็บตลอดจนการป้อนค่าข้อมูลผ่านโปรแกรมประยุกต์ (text entry application) และ (2) Linked data display ที่ใช้เพื่อแสดงผลข้อมูลที่มีความ



คล้ายคลึง (similarity) หรือมีความสัมพันธ์กัน (relation) มาแสดงผลเพื่อให้สารสนเทศที่มาจากระบบภายนอกเพิ่มเติมแก่ผู้ใช้ นอกเหนือจากข้อมูลที่สืบค้นได้จาก Web OPAC ซึ่งเทคโนโลยีทั้งสองอย่างนี้มีส่วนช่วยในการลดข้อผิดพลาดที่เกิดจากการทำรายการ เช่น การสะกดคำ การเว้นวรรค หรือใช้อักษรพิเศษ ทำให้ข้อมูลบรรณานุกรมมีความถูกต้องเที่ยงตรง รวมถึงแสดงผลของสารสนเทศที่มีความสัมพันธ์กับเขตข้อมูลควบคุมในลักษณะของ Discovery service for library ได้ รายการ Lookup files ที่ถูกเพิ่มโดยอัตโนมัติเข้าไปในระบบนี้ ปรากฏตามตารางที่ 3 และตารางที่ 4

ตารางที่ 3 ประเภทของ Lookup files ที่ถูกเพิ่มโดยอัตโนมัติเข้าไปในระบบ

Types	Tags	Lookup files
Name	100, 110, 111, 700, 710, 711	1. Preferred name สำหรับรายการหลัก (Main entry) 2. รูปแบบการลงรายการ Name authority ในภาษาอื่น 3. Alternate name forms สำหรับ tag 4XX 4. งานสร้างสรรค์หรือหนังสืออื่นที่แต่งโดยผู้แต่งคนคนนี้ 5. ผู้แต่งรวมที่เคยมีผลงานร่วมกับผู้แต่งคนนี้ 6. ประเทศหรือภูมิภาคที่เคยตีพิมพ์ผลงานของผู้แต่งคนนี้ 7. สถิติการตีพิมพ์และเผยแพร่ 8. สำนักพิมพ์ที่เคยจัดพิมพ์ผลงานของผู้แต่งคนนี้
Subject	600, 610, 611, 630, 651, 655	9. รูปแบบการลงรายการตาม LCSH และ FAST 10. Suggestion terms ของหัวเรื่อง ประกอบด้วย Variants, Broader terms, Narrower terms, Related terms และ Closely Matching Concepts from Other Schemes
Series	800, 810, 830	11. รูปแบบการลงรายการตาม LCSH Series (Publications) และ LCSH Monographic series 12. คอลเล็กชันหรือชุดที่ทรัพยากรสารสนเทศรายการนี้ปรากฏอยู่ (Collection Memberships)

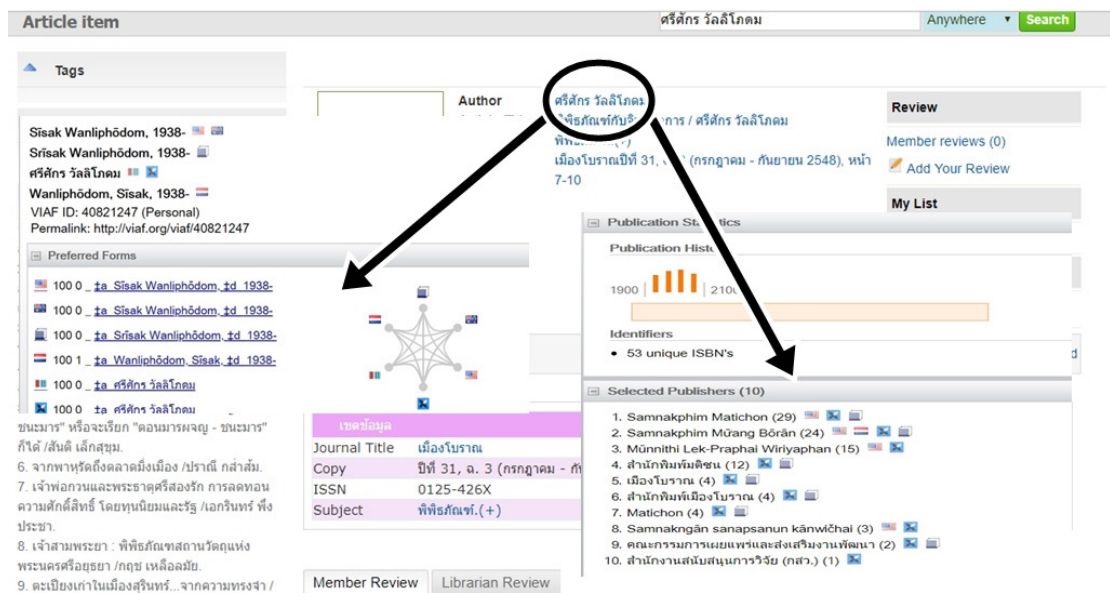
ตารางที่ 4 ตัวอย่างข้อมูล linked ของ Lookup files ที่ปรากฏใน MARC record

Types - Tags	Data categories	Data values
Name	Original	1\$aGamon, David.
100	FAST lookup	1\$aGamon, David. \$0http://id.loc.gov/authorities/names/n85118544 \$1http://viaf.org/viaf/24889383/
Name	Original	1\$aBragdon, Allen D.
700	FAST lookup	1\$aBragdon, Allen D. \$0http://id.loc.gov/authorities/names/n79108360 \$1http://viaf.org/viaf/10135827/

ตารางที่ 4 ตัวอย่างข้อมูล linked ของ Lookup files ที่ปรากฏใน MARC record (ต่อ)

Types - Tags	Data categories	Data values
Subject	Original	\0\$aBrain \$vPopular works.
650	FAST lookup	\0\$aBrain \$vPopular works. \$0http://id.loc.gov/authorities/subjects/sh2007102129
Series	Original	\0\$aBooks in soils, plants, and the environment
830	FAST lookup	\0\$aBooks in soils, plants, and the environment \$0http://id.loc.gov/authorities/names/n90667428

แฟ้มข้อมูลอ้างอิงที่ถูกสร้างขึ้นนั้น สามารถนำไปแสดงผลบนหน้าจอ Web OPAC ของระบบห้องสมุดอัตโนมัติได้หลายรูปแบบ ขึ้นอยู่กับความต้องการในการนำเสนอข้อมูลของห้องสมุดนั้น เป็นต้นว่า แสดงผลในรูปของรายการเชื่อมโยง (link list) ลิงค์ เดตา กราฟ (linked data graph) หรือบูรณาการเข้ากับหน้าจอแสดงผลของ Web OPAC โดยตรง เพื่อให้ผู้ใช้สามารถเชื่อมต่อไปยังหน้าเว็บที่ใช้ในการให้คำอธิบายรายการหลักฐาน ใช้ในการให้ข้อมูลเพิ่มเติมที่โดยปกติไม่ปรากฏในระบบห้องสมุดอัตโนมัติของสถาบัน (เช่น ผลงานอื่นของผู้แต่งที่ปรากฏอยู่ในแหล่งข้อมูลภายนอก หรือ ชื่อเรื่อง ชื่อบุคคล ชื่อสถานที่ ที่ถูกบันทึกไว้ในภาษาอื่น ๆ ที่อยู่แหล่งข้อมูลภายนอก เป็นต้น) ดังตัวอย่างในภาพที่ 3



ภาพที่ 3 ตัวอย่างการเชื่อมโยงไปยัง Lookup files เมื่อแสดงผลบนหน้า Web OPAC



อภิปรายผล

การศึกษาในครั้งนี้เป็นข้อมูลพื้นฐานที่สำคัญต่อการพัฒนาโปรแกรมประยุกต์ด้านการควบคุมรายการหลักฐาน (Authority control) โดยอัตโนมัติซึ่งสามารถสรุปและอภิปรายผลการวิจัยได้ 3 ด้านคือ

1) ด้านรูปแบบ โครงสร้าง และมาตรฐานของการทำรายการหลักฐานในรูปแบบ MARC21 สอดคล้องกับเกณฑ์ที่เป็นข้อกำหนดของมาตรฐานด้านรายการหลักฐานของหอสมุดรัฐสภาอเมริกัน โดยรายการหลักฐานในระบบ ประกอบด้วยส่วน ป้ายระเบียบ (Leader) ส่วนนามาุกรมข้อมูล (Directory) และส่วนเขตข้อมูลที่มีความยาวไม่คงที่ (Variable fields) มีรูปแบบเนื้อหาภายในระเบียบตามหัวเรื่อง (Heading) ซึ่งจำแนกออกเป็น names, name/title combinations, uniform titles, และ subjects

2) ด้านการวิเคราะห์โครงสร้างข้อมูลจากฐานข้อมูลระบบห้องสมุดอัตโนมัติ WALAI AutoLib กับรายการหลักฐานในรูปแบบ FAST มีการจัดกลุ่มเขตข้อมูล (tag) ที่ใช้ในการทำรายการหลักฐาน โดยใช้โครงสร้างระเบียบนำเข้าที่สอดคล้องกับมาตรฐานสากล ISO2709 นั้น สามารถจัดกลุ่ม Facet ตามกลุ่มเขตข้อมูลเป็น 4 กลุ่ม คือ เขตข้อมูลหลัก (tag 1XX) เขตข้อมูลที่เป็นหัวเรื่อง (6xx) เขตข้อมูลที่เป็นชื่อชุด (8XX) และรายการเพิ่ม (7XX) ซึ่งใช้วิธีแตกต่างจากผลการวิจัยของ Bennet (Bennett, O'Neill, & Kammerer, 2016) ที่แยกประเภทตาม Facet 8 กลุ่ม คือ Personal Names, Corporate Names, Event, Uniform Titles, Chronological, Topical, Geographic และ Form/Genre

3) ด้านการสร้างเพิ่มข้อมูลอ้างอิง (Lookup file) ที่เชื่อมโยงกับ FAST สำหรับรายการหลักฐานของข้อมูลบรรณานุกรมในฐานข้อมูล WalaiAutoLib สามารถสร้าง Lookup files ที่เป็น FAST Facet Application และอัปเดตลงในฐานข้อมูลโดยอัตโนมัติ ได้แก่ Name 44,393 รายการ Subject 34,208 รายการ และ Series 37,666 รายการ

ซึ่งวิธีการและเครื่องมือในการศึกษาทั้งหมดที่กล่าวนี้เหมาะสำหรับห้องสมุดและสถาบันสารสนเทศที่ขาดผู้เชี่ยวชาญที่เป็นบรรณารักษ์นักทำรายการ (cataloger) และนักทำดัชนี (Indexer) เนื่องจากระบบใช้กลไกของ linked data ในการเชื่อมต่อไปยังแหล่งข้อมูลของหน่วยงานที่เป็นเจ้าของมาตรฐานด้านการทำรายการระดับสากลและใช้เทคโนโลยี Auto suggestion และ Linked data display เพื่อให้ระบบตรวจสอบและปรับปรุงข้อมูลในเขตข้อมูลควบคุมโดยอัตโนมัติทำให้ห้องสมุดสามารถลดขั้นตอนลดกำลังคนเวลาและทรัพยากรที่ต้องใช้ในการบวนการทำรายการทรัพยากรสารสนเทศได้อย่างมีนัยสำคัญ

ข้อเสนอแนะ

ผู้วิจัยพบว่าแหล่งข้อมูลที่เป็น Faceted Application of Subject Terminology (FAST) ที่เป็นภาษาไทยนั้นปรากฏเฉพาะใน data source ของ OCLC และ VIAF เท่านั้นยังไม่มีการนำหัวเรื่องหรือนามบุคคลที่เป็นศัพท์ควบคุมของทรัพยากรสารสนเทศภาษาไทยมาทำเป็นชุดข้อมูล (data set) และให้บริการผ่านอินเทอร์เน็ตในลักษณะของ Linked data ผู้วิจัยจึงจะนำผลการวิจัยในครั้งนี้มาปรับปรุงและพัฒนาเทคโนโลยีหรือโปรแกรมประยุกต์ที่ใช้ในการสร้างชุดข้อมูลของ Faceted Application of Subject Terminology (FAST) ของศัพท์ควบคุมภาษาไทยเพื่อประโยชน์ในการรองรับการทำงานร่วมกันของห้องสมุดและสถาบันสารสนเทศในประเทศไทยเป็นลำดับถัดไป

**รายการอ้างอิง**

- วิทยา เทวรังษี, คิวนาถ นันทพิชัย และยุทธนา เจริญรัตน์. (2557) . การสกัดข้อมูลและสร้างระเบียบรายการหลักฐานจากระเบียนสหบรรณานุกรม. *PULINET Journal*, 1(1), 115-122.
- คิวนาถ นันทพิชัย, พรพิมล วัชรกุล และณรงค์ ทองรักจันทร์. (2557) . การพัฒนาโปรแกรมประยุกต์ เพื่อการควบคุมระเบียบบรรณานุกรม. *วารสารสารสนเทศศาสตร์*, 32(1), 1-33.
- Bennett, R., O'Neill, E. T., & Kammerer, K. (2016) . assignFAST: An Autosuggest based tool for FAST Subject Assignment. *Information Technology and Libraries*, 33(1), 34-43.
- Childress, E. R., & Vizine-Goetz, D. (2017) . Faceted Application of Subject Terminology (FAST) . *Encyclopedia of Library and Information*, 539-548.
- Gonzales, B. M. (2014) . Linking libraries to the web: linked data and the future of the bibliographic record. *Information Technology and Libraries*, 33(1), 10-23. Retrieved June 12, 2019, from <https://ejournals.bc.edu/index.php/ital/article/view/5631>
- Hjorland, B. (2016) . Knowledge organization (KO) . *KO KNOWLEDGE ORGANIZATION*, 43(6), 475--484.
- Library of Congress. (2012) . *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*. Washington, DC: Library of Congress. Retrieved October 25, 2018, from <http://www.loc.gov/bibframe/pdf/marcl-d-report-11-21-2012.pdf>
- Library of Congress. (2015, October 28) . *What is a MARC Authority Record*. Retrieved January 17, 2019, from WHAT IS A MARC RECORD, AND WHY IS IT IMPORTANT?: <https://www.loc.gov/marc/uma/index.html>
- Library of Congress. (2017, March 29) . *ID.LOC.GOV - Linked Data Service*. Retrieved January 14, 2019, from Linked Data Service: <https://id.loc.gov/>
- Library of Congress. (2019, June 5) . *Resource Description and Access (RDA)* . Retrieved June 14, 2019, from Information and Resources in Preparation for RDA: <https://www.loc.gov/aba/rda/>
- Online Computer Library Center. (2018, May 10) . *FAST (Faceted Application of Subject Terminology)* . Retrieved April 24, 2019, from OCLC FAST: <https://fast.oclc.org/>
- Sahu, S. R. (2002) . Converting bibliographic records into ISO 2709 format. In S. R. Sahu, *Training course on Information Technology Application to library and Information Services* (pp. 2-36) . Bangalore: Indian Institute of Science.
- VIAF. (2018, June 17) . *The Virtual International Authority File*. Retrieved January 20, 2019, from VIAF: The Virtual International Authority File: <http://viaf.org/viaf/>
- Walailak University. (2018) . *Implementation of new validation rules in UC-TAL (UC Connexion Client Ver. 3.1.2) Rev. May 2018*. Nakhon Si Thammarat: Walailak University, Innovation Information Research Unit.