PULINET Journal

การสกัดข้อมูลและสร้างระเบียนรายการหลักฐานจากระเบียนสหบรรณานุกรม
Data Extracting and Generating the Authority Records from Union Catalog Bibliographic Records

วิทยา เทวรังษี* ศิวนาถ นันทพิชัย ยุทธนา เจริญรื่น
สำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์
e-mail: twittaya@wu.ac.th*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ (1) เพื่อศึกษา วิเคราะห์รูปแบบความสัมพันธ์ของข้อมูลการลงรายการบรรณานุกรม และการลงรายการหลักฐานตามมาตรฐาน MARC 21 Format (2) เพื่อสร้างระเบียนรายการหลักฐานจากระเบียนบรรณานุกรมโดยอัตโนมัติและ (3) เพื่อศึกษาวิธีการควบคุมความถูกต้องของระเบียนรายการหลักฐานจากระเบียนบรรณานุกรมโดยอัตโนมัติวิธีการศึกษาประกอบด้วย (1) การวิเคราะห์เอกสาร (Document Analysis) (2) การสัมภาษณ์เชิงลึก (In-Depth Interview) จากผู้เชี่ยวชาญของคณะกรรมการเครือข่าย ThaiLIS และ (3) การประชุมกลุ่ม (Focus Group) โดยการจัดประชุมกลุ่มร่วมกับห้องสมุดสมาชิกเครือข่าย ThaiLIS จำนวน 24 แห่ง

ผลการศึกษา พบว่า (1) สามารถกำหนดวิธีการสร้างเป็นข้อมูลรายการหลักฐาน (Authority Data) จากข้อมูลบรรณานุกรม (Bibliographic Data) ของห้องสมุดสมาชิกที่นำเข้า โดยนำข้อมูลมาสกัดข้อมูล (Extract) และจำแนกเป็นรายการหลักฐาน 3 ประเภทคือ ชื่อ (Name) หัวเรื่อง (Subject) และชื่อชุด (Series) (2) สามารถสร้างระเบียนรายการหลักฐาน (Authority record) ตามรูปแบบ MARC 21 Format for Authority Data ประกอบด้วย Name 594,845ระเบียน Subject 413,980 ระเบียน Series 23,269 ระเบียน และ Subdivision 35,507 ระเบียน (3) สามารถสร้างกระบวนการตรวจสอบความซ้ำซ้อน และกฎเกณฑ์เงื่อนไขสำหรับการตรวจสอบความถูกต้องของระเบียนเช่น การตรวจสอบรูปแบบขอบตัวอักษรพิมพ์ การตรวจสอบอักขระพิเศษ การตรวจสอบซ้ำระเบียนจำแนกตามประเภทและข้อกำหนดรายการเขตข้อมูลสำหรับการสร้างระเบียนรายการหลักฐาน ซึ่งผลการศึกษาทั้งหมดนี้จะนำไปใช้เป็นพื้นฐานสำหรับการพัฒนาโปรแกรมและเครื่องมือในการตรวจสอบความถูกต้องของระเบียนรายการหลักฐานต่อไป

คำสำคัญ : ระเบียนรายการหลักฐาน,การสกัดข้อมูล,ระเบียนบรรณานุกรม

Keywords: Authority Record, Data Extracting, Bibliographic Record

บทนำ

การพัฒนาระบบฐานข้อมูลสหบรรณานุกรมสำหรับห้องสมุดสถาบันอุดมศึกษาของไทยภายใต้การดำเนินงานของคณะทำงาน ThaiLISมุ่งพัฒนาองค์ประกอบที่สำคัญ 3 ส่วนคือ (1) ฐานข้อมูลเลขมาตรฐานสากลของหนังสือ/วารสาร เพื่อรองรับการตรวจสอบเลขมาตรฐานสากล และการทำรายการสิ่งพิมพ์สำเร็จรูป (2) ฐานข้อมูลสหบรรณานุกรม เพื่อรองรับการทำรายการตามหลักเกณฑ์การทำรายการ (AACR2, MARC21) และตามมาตรฐานของการสื่อสารและแลกเปลี่ยนข้อมูล (ISO2709, Z39.50) และ (3) โปรแกรมรองรับการตรวจสอบรายการ การทำรายการ และการใช้รายการร่วมกัน ประกอบด้วยระบบย่อย (Module) ที่สำคัญคือ ระบบการจัดการเลขมาตรฐานสากลของหนังสือ/วารสาร (ISN Module) ระบบการทำรายการ (Union Cataloging Module) ระบบการสืบค้นฐานข้อมูลสหบรรณานุกรม (Web OPAC for Union Catalog Module) และระบบการตรวจสอบและควบคุมรายการหลักฐาน (Union Catalog Authority Control Module)

ระบบงานควบคุมรายการหลักฐาน (Authority Control) ซึ่งเป็นระบบงานย่อยหนึ่งที่ทำหน้าที่ควบคุมและจัดการความสัมพันธ์ข้อมูลประเภทต่างๆ ได้แก่ ชื่อ (Name) หัวเรื่อง (Subject) และชื่อชุด (Series) จัดเก็บอยู่ในรูปแบบของระเบียนรายการหลักฐาน (Authority Record) โดยอยู่ในรูปแบบของ MARC21 Format for Authority Data ซึ่งเป็นมาตรฐานในการจัดเก็บข้อมูลระเบียนควบคุมรายการเพื่อให้ข้อมูลอยู่ในรูปแบบเดียวกันตามมาตรฐานที่กำหนด ปัจจุบันระบบสหบรรณานุกรมไม่มีระบบควบคุมรายการหลักฐาน และประกอบกับข้อมูลที่อัพโหลดมาจากห้องสมุดสถาบันต่างๆ ไม่ใช่การลงรายการใหม่และเป็นข้อมูลบรรณานุกรมเดิมของแต่ละสถาบัน ส่งผลให้ข้อมูลควบคุมมีความหลากหลาย และมีความผิดพลาดเกิดขึ้น ซึ่งส่งผลให้การสืบค้นในระบบ OPAC ไม่มีประสิทธิภาพ จากปัญหาดังกล่าวจึงมีแนวคิดที่จะพัฒนาระบบงานควบคุมรายการสำหรับสหบรรณานุกรมเพื่อใช้ในการจัดการข้อมูลควบคุมให้ถูกต้องและอยู่ในรูปแบบเดียวกัน เพื่อทำให้ข้อมูลบรรณานุกรมถูกต้อง ซึ่งส่งผลให้เกิดการสืบค้นที่มีประสิทธิภาพ

วัตถุประสงค์

1. เพื่อศึกษาและวิเคราะห์รูปแบบความสัมพันธ์ของข้อมูลการลงรายการบรรณานุกรม และการลงรายการหลักฐานตามมาตรฐาน MARC 21 Format

2. เพื่อสร้างระเบียนรายการหลักฐานจากระเบียนบรรณานุกรมโดยอัตโนมัติ

3. เพื่อศึกษาวิธีการควบคุมความถูกต้องของระเบียนรายการหลักฐานจากระเบียนบรรณานุกรมโดยอัตโนมัติ

วิธีการศึกษา

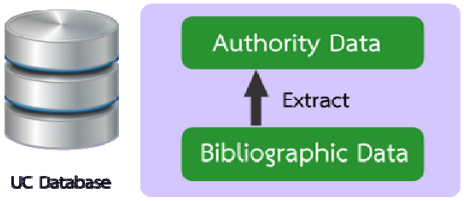
1. วิเคราะห์เอกสาร (Document Analysis) เพื่อทบทวนเอกสารที่เกี่ยวข้องกับรูปแบบ โครงสร้างมาตรฐาน และการดำเนินงานที่เกี่ยวข้องรวมทั้งวิเคราะห์กฎการตรวจสอบข้อมูลที่มีอยู่เดิมจากเอกสารการประชุมของคณะกรรมการThaiLIS

2. สัมภาษณ์เชิงลึก (In-Depth Interview) เพื่อสรุปปัญหาและข้อจำกัดของการดำเนินงาน โดยใช้การสุ่มตัวอย่างแบบเฉพาะเจาะจง (Purposive Sampling) จากผู้เชี่ยวชาญของคณะกรรมการเครือข่าย ThaiLIS จำนวน 1 คน

3. การประชุมกลุ่ม (Focus Group) เพื่อนำเสนอผลการวิเคราะห์และสรุปแนวทางการตรวจสอบความถูกต้องของระเบียนบรรณานุกรมสู่การกำหนดเป็นกฎเกณฑ์สร้างรายการหลักฐานโดยการจัดประชุมกลุ่มร่วมกับห้องสมุดสมาชิกเครือข่าย ThaiLIS จำนวน 24 แห่ง จำนวน 1 ครั้ง

ผลการศึกษา

1. ข้อมูลรายการหลักฐานของระบบฐานข้อมูลสหบรรณานุกรม



ภาพที่ 1 แสดงการสกัดข้อมูลรายการหลักฐาน

ภาพที่ 1 ข้อมูลบรรณานุกรม (Bibliographic Data) ของห้องสมุดสมาชิกที่นำเข้าและจัดเก็บในฐานข้อมูลสหบรรณานุกรมถูกนำมาสกัดข้อมูล (Extract) เพื่อนำมาสร้างเป็นข้อมูลรายการหลักฐาน (Authority Data) สามารถจำแนกรายการหลักฐานออกเป็น 3 ประเภทคือ ชื่อ (Name) หัวเรื่อง (Subject) และชื่อชุด (Series) โดยการสกัดข้อมูลนั้นจะพิจารณาจากเขตข้อมูล (Tag) ในระเบียนบรรณานุกรมเป็นสำคัญ เขตข้อมูลที่นำมาพิจารณาประกอบด้วยเขตข้อมูลกลุ่ม 1XX, 6XX, 7XX และ 8XX ดังตารางที่ 1

ตารางที่ 1 เขตข้อมูลที่นำมาใช้พิจารณาในการจัดทำข้อมูลรายการหลักฐาน

Type	Tag	Fields
Name	100	Main Entry - Personal Name
	110	Main Entry - Corporate Name
	111	Main Entry - Meeting Name
	130	Main Entry - Uniform Title
	700	Added Entry - Personal Name
	710	Added Entry - Corporate Name
	711	Added Entry - Meeting Name
	730	Added Entry - Uniform Title

Subject	600	Subject Added Entry - Personal Name
	610	Subject Added Entry - Corporate Name
	611	Subject Added Entry - Meeting Name
	630	Subject Added Entry - Uniform Title
	648	Subject Added Entry - Chronological Term
	650	Subject Added Entry - Topical Term
	651	Subject Added Entry - Geographic Name
	655	Index Term - Genre/Form
Series	800	Series Added Entry - Personal Name
	810	Series Added Entry - Corporate Name
	811	Series Added Entry - Meeting Name
	830	Series Added Entry - Uniform Title

2. ข้อมูลที่สกัดได้ทั้งหมดถูกสร้างเป็นระเบียนรายการหลักฐาน (Authority Record) โดยสร้างเขตข้อมูลหลัก (1XX) ตามรูปแบบ MARC 21 Format for Authority Data และแยกตามประเภทได้แก่ Subject, Author และ Seriesนอกจากนี้ระบบฯ ยังสร้างหัวเรื่องย่อย (Subdivision) โดยการสกัดข้อมูลจากรายการหลักฐานที่เป็นหัวเรื่อง ในเขตข้อมูลย่อย (Subfield) ได้แก่ v, x, y และ z จากนั้นนำมาสร้างเป็น Authority Record ที่มีเขตข้อมูลหลักเป็น 18X เพื่อใช้ในการควบคุมข้อมูลของ Authority Record อีกระดับหนึ่ง

การนำข้อมูลจากระเบียนบรรณานุกรมไปสร้างเป็นข้อมูลรายการหลักฐาน ข้อมูลในรายการ แต่ละเขตข้อมูลย่อยจะตัดอักขระพิเศษออกทั้งหมดและนำไปตรวจสอบซ้ำกับรายการหลักฐาน ในกรณีที่ไม่มีรายการหลักฐานให้สร้างเป็นระเบียนรายการหลักฐานใหม่ หลังจากนั้นจะนำข้อมูลในเขตข้อมูลย่อย (Subfield) ที่กำหนดไปสร้าง View สำหรับการสืบค้นของ OPAC โดยรายการที่กำหนดได้แก่ a, b, c, d, v, x, y และ z ผลการดำเนินงานโดยระบบฯ ทำให้สามารถแปลงข้อมูลจากระเบียนบรรณานุกรมจำนวน 712,788 ระเบียนไปเป็นระเบียนรายการหลักฐานได้จำนวน 1,032,094 ระเบียน ดังตารางที่ 2

ตารางที่ 2 ผลการดำเนินงานจัดทำระเบียนรายการหลักฐานของระบบฐานข้อมูลสหบรรณานุกรม

Type	**Tag**	**Fields**	**Record**	**Similarity**
Name	100	Main Entry - Personal Name	525,750	30,239
	110	Main Entry - Corporate Name	58,676	245

Type	Tag	Fields	Record	Similarity
(594,845 records)	111	Main Entry - Meeting Name	10,135	50
	130	Main Entry - Uniform Title	284	12
Subject (413,980 records)	100	Subject Added Entry - Personal Name	22,607	208
	110	Subject Added Entry - Corporate Name	25,298	557
	111	Subject Added Entry - Meeting Name	59	0
	130	Subject Added Entry - Uniform Title	820	25
	148	Subject Added Entry - Chronological Term	0	0
	150	Subject Added Entry - Topical Term	337,343	14,872
	151	Subject Added Entry - Geographic Name	27,804	1,569
	155	Index Term - Genre/Form	49	19
Series (23,269 records)	100	Series Added Entry - Personal Name	44	3
	110	Series Added Entry - Corporate Name	504	5
	111	Series Added Entry - Meeting Name	7	0
	130	Series Added Entry - Uniform Title	22,714	120
Subdivision (35,507 records)	180	Heading - General Subdivision	18,767	-
	181	Heading - Geographic Subdivision	6,186	-
	182	Heading - Chronological Subdivision	2,803	-
	185	Heading - Form Subdivision	7,751	-
		รวม	1,067,601	47,924

3. การตรวจสอบความถูกต้องของระเบียนรายการหลักฐาน

นอกจากการสร้างระเบียนรายการหลักฐานแล้ว ระบบฐานข้อมูลสหบรรณานุกรมยังได้สร้างกระบวนการตรวจสอบความซ้ำซ้อนของข้อมูลรายการหลักฐานด้วย และจากการพบข้อมูลที่คล้ายคลึงกันที่เกิดจากการลงรายการแตกต่างกัน แต่มี view สำหรับการสืบค้นเหมือนกัน (Similarity) ดังนั้นเพื่อแก้ไขปัญหาดังกล่าว ผู้วิจัยจึงนำเสนอปัญหานี้กับที่ประชุมห้องสมุดสมาชิกเครือข่าย ThaiLIS เพื่อทำความตกลงและกำหนดขอบเขตการแปลงข้อมูลเพิ่มเติม ที่จะทำให้ลดจำนวนข้อมูลรายการหลักฐาน (Authority Data) และลดความซ้ำซ้อนของระเบียน

รายการหลักฐาน (Authority Record) ซึ่งจากมติที่ประชุมในวันที่ 20 พฤศจิกายน 2556 ได้มีการกำหนดข้อตกลงร่วมกันดังนี้

1. รูปแบบขอบตัวอักษรพิมพ์ใหญ่ พิมพ์เล็กของข้อมูลดูจากรูปแบบการกำหนดของ LC คือตัวอักษรแรกของแต่ละ Subfield เป็นตัวใหญ่และตัวอักษรที่ตามหลังจะเป็นตัวเล็กทั้งหมด

2. ข้อมูลจะตัดอักขระพิเศษที่อยู่ท้ายสุดของแต่ละ subfield ออกทั้งหมด

3. การตรวจสอบซ้ำของ Authority Record จะจำแนกตามประเภทของ Authority และ Authority Tag

4. รายการ Tag และ Subfield ที่นำไปสร้างเป็น Authority record ดังตารางที่ 3

ตารางที่ 3รายละเอียดของข้อกำหนดการตรวจสอบระเบียนรายการหลักฐาน

Type	Bib Tag	Authority Tag	รายละเอียดของข้อกำหนด
Name	100	100	- นำไปตรวจสอบซ้ำและสร้าง Authority record เฉพาะ $a และไม่สนใจ space
	110	110	- นำข้อมูลทุก subfield ที่ลงรายการไปตรวจสอบซ้ำและสร้าง Authority record โดย subfield นั้นจะต้องเป็น subfield ที่กำหนดให้ใช้ใน MARC21 format for Authority data ด้วย
	111	111	
	130	130	
	700	100	- นำไปตรวจสอบซ้ำและสร้าง Authority record เฉพาะ $a และไม่สนใจ space
	710	110	- นำข้อมูลทุก subfield ที่ลงรายการไปตรวจสอบซ้ำและสร้าง Authority recordโดย subfield นั้นจะต้องเป็น subfield ที่กำหนดให้ใช้ใน MARC21 format for Authority data ด้วย
	711	111	
	730	130	
Subject	600	100	- นำไปสร้าง Authority record เฉพาะ $a, $v, $x, $y, $z
	610	110	- หมายเลขไทยในหัวเรื่องให้แปลงเป็นเลขอาราบิค
	611	111	- กรณีที่หัวเรื่องมีความคล้ายคลึงกัน และมีความแตกต่างเฉพาะการลงรายการระหว่าง $x และ $v ถ้าระเบียนที่พบว่าคล้ายกันนั้นมีการสร้างระเบียนโดยมีการลง $v ไว้ก่อนแล้วให้แก้ $x เป็น $v และไม่ต้องสร้าง Authority record ใหม่
	630	130	
	648	148	
	650	150	- กรณีที่หัวเรื่องมีการลงรายการโดยการใส่ -- ลงใน$a ให้นำข้อมูลดังกล่าวไปค้นหาแบบ Similarity ก่อน ถ้าพบระเบียนที่คล้ายคลึงกันให้แก้ไขข้อมูลการลงรายการหัวเรื่องให้เหมือนกับรายการคล้ายคลึงที่พบและไม่ต้องสร้าง
	651	151	

Type	Bib Tag	Authority Tag	รายละเอียดของข้อกำหนด
	655	155	ระเบียนใหม่
Series	800	100	- นำไปสร้าง Authority record เฉพาะ $a
	810	110	
	811	111	
	830	130	

บทสรุป

ผลการศึกษาในภาพรวมที่นอกจากจะได้รูปแบบ โครงสร้าง และเงื่อนไขต่าง ๆ ที่สามารถนำไปใช้ตรวจสอบและการควบคุมระเบียนรายการหลักฐานแล้ว ยังสามารถนำไปพัฒนาโปรแกรมการจัดการระเบียนบรรณานุกรมได้อีกด้วย ซึ่งก่อให้เกิดประโยชน์ต่อการพัฒนาระบบฐานข้อมูลสหบรรณานุกรมมีโปรแกรมที่สามารถแก้ไขเขตข้อมูลควบคุมอัตโนมัติ ที่สำคัญคือเงื่อนไขการตรวจสอบข้อมูลนี้ยังช่วยเพิ่มประสิทธิภาพการพัฒนาระบบฐานข้อมูลสหบรรณานุกรมได้มากยิ่งขึ้น ทำให้ระเบียนบรรณานุกรมมีความถูกต้องสมบูรณ์ในการนำไปประมวลผลและสร้างเป็นรายการหลักฐาน (Authority Control) ที่มีความถูกต้องสมบูรณ์ซึ่งส่งผลต่อระบบสืบค้น (OPAC) มีประสิทธิภาพมากยิ่งขึ้น อย่างไรก็ตามในการดำเนินงานดังกล่าวจำเป็นต้องใช้ระยะเวลา และความร่วมมือจากห้องสมุดสมาชิก ในการตรวจสอบและแก้ไขข้อมูลเพื่อนำเข้าระบบจำเป็นต้องแสวงหาความร่วมมือกับบุคคลคณะทำงาน และหน่วยงานกลางที่ทำหน้าที่บริหารจัดการข้อมูลสหบรรณานุกรม รวมไปถึงดำเนินการฝึกอบรมและเตรียมความพร้อมของห้องสมุดสมาชิกในการพัฒนาฐานข้อมูลสหบรรณานุกรมต่อไป

เอกสารอ้างอิง

คณะทำงานกลุ่มวิเคราะห์ทรัพยากรสารสนเทศ ห้องสมุดสถาบันอุดมศึกษา. (2547).*หัวเรื่องสำหรับหนังสือภาษาไทย* (พิมพ์ครั้งที่ 4).

สุวันนา ทองสีสุขใส. (2543).*MARC 21 สำหรับระเบียนหนังสือ / เอกสาร*. ขอนแก่น: สำนักวิทยบริการ มหาวิทยาลัยขอนแก่น.

อุหมาด หมัดอาด้ำ.(2551). *โครงสร้างข้อมูล*.นครศรีธรรมราช: สำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์.

Chan, Lois Mai. (1995). *Library of Congress Subject Heading*. (3rd ed.). Englewood, Colorado: LIBRARIES UNLIMITED, INC.

Drozdek, Adam. (2008).*Data Structure and Algorithm in Java*. (3rd ed.). Singapore: Cengage Learning Asia.

Ganendran, Jacki. (1998). Learn SUBJECT ACCESS. (2nd ed.). Canberra: DocMatrix.

Julien, Charles-Antoine, Tirilly, Prerre, Leide, John E., &Guastavino, Catherine. (2012) Constructing a True LCSH Tree of a Science and Engineering Collection. *Journal of The American Society for Information Science and Technology* 63(12), 2405-2418.

Maneewongvatana, Suthathip, &Maneewongvatana, Songrit.(2011). A Similarity Model for Bibliographic Records Using Subject Headings. In Eighth International Joint *Conference on Computer Science and Software Engineering (JCSSE), Institute of Electrical and Electronics Engineers (IEEE)*,263-268.

Miller, Joseph. (1997).*Sear List of Subject Heading* (16th ed.). New York: The H. W. Wilson Company.

Mortimer, Mary. (2543).*Introducing MARC 21*. กรุงเทพฯ: บุ๊คโปรโมชั่น แอนด์ เซอร์วิส.

Yi, Kwan, &Chan, Lois Mai. (2009). Linking Folksonomy to Library of Congress subject heading: an exploratory study. *Journal of Documentation* 65(6), 872-900.