ฐานข้อมูลพิเศษ : ฐานข้อมูลศูนย์เอกสารประเทศไทย

สุนทรี ศุภวงศ์ *

ประวัติศูนย์เอกสารประเทศไทย

ศูนย์เอกสารประเทศไทย เป็นหน่วยงานที่จัดขึ้นด้วยความร่วมมือระหว่างศูนย์วิจัยและพัฒนาการทหาร กองบัญชาการทหารสูงสุดของไทย กับองค์การวิจัยโครงการขั้นสูง (ARPA) กระทรวงกลาโหมสหรัฐฯ เมื่อปี พ.ศ. 2511 โดยได้มอบหมายให้ Battelle Memorial Institute, Columbus, Ohio เป็นผู้วางระบบสารสนเทศ (information system) และดำเนินการจัดตั้งให้เป็นแหล่งรวบรวมเอกสารทางสังคมศาสตร์เกี่ยวกับประเทศไทย เพื่อให้บริการแก่นักวิจัยของรัฐบาลไทยและรัฐบาลสหรัฐฯ เกี่ยวกับความมั่นคงและการพัฒนา

ต่อมาเมื่อต้นปี พ.ศ. 2514 องค์การวิจัยโครงการขั้นสูง (ARPA) ได้เตรียมปิดการดำเนินงานวิจัยในประเทศไทย ฉะนั้นศูนย์เอกสารประเทศไทยซึ่งเป็นส่วนหนึ่งของ ARPA จึงจำเป็นต้องยุติงานตามไปด้วย จุฬาลงกรณ์มหาวิทยาลัยได้แจ้งความจำนงขอรับศูนย์เอกสารประเทศไทยมาดำเนินการเมื่อวันที่ 30 ธันวาคม พ.ศ. 2514 ต่อจากนั้น ศูนย์เอกสารประเทศไทยได้รวมเข้าเป็นหน่วยงานหนึ่งในสถาบันวิทยบริการ ตามพระราชกฤษฎีกาซึ่งให้ไว้ ณ วันที่ 17 พฤษภาคม 2521

ศูนย์เอกสารประเทศไทย สถาบันวิทยบริการ เป็นแหล่งบริการสารสนเทศทางด้านสังคมศาสตร์และพฤติกรรมศาสตร์เกี่ยวกับประเทศไทยและประเทศในภูมิภาคเอเชียตะวันออกเฉียงใต้ ให้บริการทางวิชาการแก่ นิสิตบัณฑิตศึกษา คณาจารย์ และบุคลากรของจุฬาลงกรณ์มหาวิทยาลัย ตลอดจนนักวิจัยทั่วไป และนักศึกษาระดับบัณฑิตศึกษาจากสถาบันอื่น ๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ความเป็นมาของฐานข้อมูลศูนย์เอกสารประเทศไทย

ระบบสารสนเทศของศูนย์เอกสารประเทศไทย หรือ Thailand Information Center (TIC) ซึ่ง Battelle Memorial Institute เป็นผู้จัดวางระบบไว้เมื่อปี พ.ศ.2511 นั้น เป็นระบบที่พร้อมที่จะรับสถานการณ์ "กระแสสารสนเทศท่วมท้น" ได้อย่างมีประสิทธิภาพ และความสามารถ ปรับเปลี่ยนเข้าสู่ระบบห้องสมุดอัตโนมัติได้อย่างคล่องตัว มีการวิเคราะห์สารสนเทศและนำเสนอสาร

* นักเอกสารสนเทศ ชำนาญการ 8 ศูนย์เอกสารประเทศไทย จุฬาลงกรณ์มหาวิทยาลัย

สนเทศในรูปแบบใหม่ (repackaging of information) เพื่อให้ผู้ใช้บริการเข้าถึงข้อสนเทศได้อย่างสะดวก รวดเร็ว และตรงกับความต้องการ ระบบการจัดเก็บและสืบค้นข้อสนเทศ (information storage & retrieval system) จะแตกต่างจากห้องสมุดทั่วไป กล่าวคือ เอกสารของ TIC จัดเรียงหมายเลขตามลำดับก่อนหลังที่ได้รวบรวมเข้ามาไว้ใน TIC และจัดเก็บไว้ในชั้นปิด (closed shelves) สำหรับระบบการสืบค้นข้อสนเทศนั้นได้มีการพัฒนาเป็น 2 ระยะ ก่อนที่จะปรับเข้าสู่ระบบเครือข่ายห้องสมุด จุฬาลงกรณ์มหาวิทยาลัย หรือ Chulalinet ดังนี้

1. ยุคของบัตรรายการ (ระยะแรกถึงปี พ.ศ. 2530) ได้จัดทำบัตรรายการขนาด 5 x7 นิ้ว เป็นเครื่องมือช่วยค้น บัตรรายการประกอบด้วย หมายเลขของเอกสาร (Accession number) รายละเอียดทางบรรณานุกรม และที่สำคัญคือ มีสาระสังเขปของเอกสาร ซึ่งจะช่วยให้นักวิจัยสามารถคัดเลือกเอกสารที่ต้องการได้โดยไม่ต้องเสียเวลาอ่านเอกสารทั้งหมด บัตรรายการนี้จะจำแนกตามหัวเรื่องดรรชนีบัญญัติศัพท์ (Key to use category file) ที่กำหนด ขึ้นใช้ใน TIC มีประมาณ 300 หัวเรื่อง นอกจากนี้ยังได้จัดเรียงบัตรรายการไว้ตามชื่อผู้แต่ง ชื่อเรื่อง ประเภทของเอกสาร และแหล่งผลิตเอกสารอีกด้วย เพื่อความสะดวกในการค้นคว้าวิจัย นอกจากนี้ยังได้จัดทำเครื่องมือช่วยค้นอื่น ๆ ที่จะอำนวยความสะดวกแก่ผู้ใช้บริการได้แก่

1.1 บรรณานุกรม ซึ่งรวบรวมและคัดเลือกบัตรรายการในหัวข้อที่มีผู้สนใจค้นคว้ากันมาก มาถ่ายเอกสารและจัดทำเป็นบรรณานุกรมเฉพาะเรื่อง ในระยะแรกนี้มีบรรณานุกรมประมาณ 120 เรื่อง มีการเพิ่มเติมข้อมูลใหม่ที่รวบรวมเข้ามาใน TIC อยู่ตลอดเวลา รวมทั้งพิจารณาจัดทำบรรณานุกรมหัวข้อใหม่ ๆ ที่นักวิจัยสนใจค้นคว้าเพิ่มขึ้นด้วย

1.2 ดรรชนีภูมิภาค (locator) รวบรวมหมายเลขเอกสาร ที่มีข้อมูลเกี่ยวกับทุกอำเภอทุกจังหวัด และทุกภาคของประเทศไทยไว้อย่างครบถ้วน ตัวอย่างเช่น ถ้านักวิจัยต้องการศึกษาสถิติหรือข้อมูลบางอย่างของอำเภอแม่สอด จังหวัดตาก นักเอกสารสนเทศสามารถนำดรรชนีภูมิภาคของอำเภอแม่สอดมาให้ใช้ ซึ่งจะมีหมายเลขเอกสารที่กล่าวถึงอำเภอแม่สอดรวบรวมไว้ทุกรายการ อันเป็นบริการพิเศษของ TIC

1.3 รายละเอียดของชาวเขาและชนกลุ่มน้อยในประเทศไทย ได้รวบรวมหมายเลขเอกสารทุกรายการที่เกี่ยวข้องกับชาวเขาและชนกลุ่มน้อยไว้ต่างหาก โดยจำแนกข้อมูลของชาวเขาเผ่าต่าง ๆ 36 เผ่า และชนกลุ่มน้อยอีก 38 กลุ่ม

เมื่อปริมาณของเอกสารสิ่งพิมพ์มีเพิ่มมากขึ้น การสืบค้นข้อมูลจากบัตรรายการเริ่มใช้เวลาเพิ่มมากขึ้น และมีปัญหาเรื่องจำนวนตู้บัตรรายการและสถานที่ตั้งตู้บัตรตามมา ประกอบกับเริ่มมีการใช้เทคโนโลยีกับงานห้องสมุด ศูนย์เอกสารประเทศไทยจึงเริ่มพัฒนาฐานข้อมูล TIC ขึ้น

2. ยุคของฐานข้อมูล (พ.ศ.2530 ถึง พ.ศ. 2537) เพื่อให้ระบบสารสนเทศมีความสมบูรณ์และทันสมัย TIC จึงเริ่มพัฒนาระบบตั้งแต่เดือนมกราคม พ.ศ.2530 โดยเริ่มจากการศึกษา

ข้อมูลที่สำคัญต่าง ๆ เพื่อกำหนดการลงรายการในแผ่นงาน (worksheet) สำหรับนำข้อมูลบัญทึกลงฐานข้อมูล ITIC และใช้โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS จัดการกับฐานข้อมูล เริ่มโครงการบันทึกข้อมูลเอกสารใหม่เข้าสู่ฐานข้อมูล TIC ในเดือน เมษายน พ.ศ.2530

ฐานข้อมูล TIC สามารถเอื้ออำนวยประโยชน์ให้แก่ผู้ใช้บริการอย่างเต็มที่ ทำให้ผู้ใช้เลือกค้นสารสนเทศตามต้องการได้ อาทิเช่น ชื่อผู้แต่ง ชื่อเรื่อง ชื่อวารสาร/หนังสือพิมพ์ การประชุม/สัมมนา ผู้จัดพิมพ์/หน่วยงานที่รับผิดชอบ ปีที่พิมพ์ นอกจากนี้ ยังสามารถสืบค้นสารสนเทศได้เฉพาะเจาะจงยิ่งขึ้นเพราะ TIC ได้สร้างและควบคุมคำสำคัญ (descriptors) โดยใช้ Macrothesaurus for Information Processing in the Field of Economic and Social Development ของ UN/OECD เป็นคู่มือ และที่สำคัญยิ่งกว่านี้ ยังสามารถสืบค้นจากคำสำคัญที่เป็นชื่อบุคคล ชื่อนิติบุคคล ดรรชนีภูมิภาคของประเทศไทย ชาวเขาและชนกลุ่มน้อยต่าง ๆ ซึ่งสามารถสืบค้นโดยตรงจากชื่ออำเภอ จังหวัด หรือชื่อกลุ่มชนดังกล่าวมาแล้วได้โดยสะดวกและรวดเร็ว

อย่างไรก็ตาม การจัดการฐานข้อมูล TIC ด้วยโปรแกรม Mini - Micro CDS/ISIS มีข้อจำกัดหลายประการ อาทิเช่น การทำงานของโปรแกรมมีความจำเป็นต้องใช้เนื้อที่ว่างใน hard disk เป็นจำนวนมาก ไม่ว่าจะเป็นการจัดการกับแฟ้มข้อมูลหลัก แฟ้มดรรชนี รวมถึงการประมวลผลข้อมูลต่าง ๆ ประกอบกับข้อมูลแต่ละระเบียนของฐานข้อมูล TIC ก็มีปริมาณมากเช่นกัน เพราะนอกจากจะมีรายละเอียดจากบรรณานุกรมแล้ว ยังมีสาระสังเขปและดรรชนีคำสำคัญต่าง ๆ ทำให้ไม่เหมาะกับการใช้งานในไมโครคอมพิวเตอร์

ฐานข้อมูล TIC มีข้อมูลทั้งหมดประมาณ 47,000 ระเบียน สามารถที่จะให้บริการในขณะนั้นเฉพาะข้อมูลของเอกสารใหม่ประมาณ 6,000 ระเบียนเท่านั้น การสืบค้นข้อมูลจึงไม่สะดวกตามที่ต้องการเพราะผู้ใช้บริการจะต้องค้นทั้งจากบัตรรายการและจากฐานข้อมูลจึงจะได้ข้อมูลครบถ้วน

ฐานข้อมูล TIC กับ Chulalinet

ฐานข้อมูล TIC มีลักษณะเป็นฐานข้อมูลทางบรรณานุกรมของเอกสารสิ่งพิมพ์ที่รวบรวมอยู่ใน TIC ซึ่งปัจจุบันมีมากกว่า 50,000 รายการ เอกสารสิ่งพิมพ์ส่วนใหญ่เป็นภาษาอังกฤษ ข้อมูลแต่ละระเบียนในฐานข้อมูลนี้ จัดทำเป็นภาษาอังกฤษทั้งหมดไม่ว่าจะเป็นรายละเอียดทางบรรณานุกรม สาระสังเขปและคำสำคัญ (descriptors) แม้เอกสารภาษาไทยที่มีอยู่ก็จะแปลข้อมูลทางบรรณานุกรมเป็นภาษาอังกฤษด้วยเช่นกัน ซึ่งรายการเหล่านี้เป็นประโยชน์ต่องานวิจัยด้านสังคมศาสตร์เกี่ยวกับประเทศไทย นอกจากนี้ยังเป็นประโยชน์ในด้านการเรียนการสอนของนิสิตบัณฑิตศึกษา และนิสิตชาวต่างประเทศในโครงการแลกเปลี่ยนต่าง ๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ข้อมูลทางบรรณานุกรมของเอกสารเฉพาะส่วนที่เป็นหนังสือ/เอกสาร (mongraph) จากฐานข้อมูล TIC ได้รับการพิจารณาให้รวมอยู่ในฐานข้อมูลหลัก (main database) ของ Chulalinet สำหรับข้อมูลที่สมบูรณ์อันได้แก่รายละเอียดทางบรรณานุกรมพร้อมสาระสังเขปและคำสำคัญของเอกสารทั้งที่เป็นหนังสือ/เอกสาร และเป็นบทความจากวารสารและหนังสือพิมพ์ (serials) จะอยู่ในฐานข้อมูลพิเศษ ฐานข้อมูพิเศษ (Special database) เป็นฐานข้อมูลที่พัฒนาขึ้นโดยมีวัตถุประสงค์ที่จะเพิ่มประสิทธิภาพของงานบริการสารสนเทศ เพื่อรับสถานการณ์ที่เรียกว่า" กระแสสารสนเทศท่วมท้น (information explosion) หรือ การทะลักทะลายของเอกสาร" มีคุณลักษณะที่แตกต่างจากฐานข้อมูลหลัก คือ นอกจากรายละเอียดทางบรรณานุกรมแล้ว ยังมีการวิเคราะห์สารสนเทศและนำเสนอสารสนเทศในรูปแบบใหม่ ได้แก่ การจัดทำสาระสังเขปและดรรชนีเป็นต้น ประเภทของสิ่งพิมพ์ (material types) ที่รวบรวมไว้ในฐานพิเศษนี้มีความหลากหลาย ทั้งนี้ขึ้นอยู่ในวัตถุประสงค์ของหน่วยงานที่รับผิดชอบแต่ละฐานข้อมูล เนื้อหาวิชาที่รวบรวมค่อนข้างเฉพาะเจาะจง และมุ่งเน้นสารสนเทศในระดับลึก เพื่อประโยชน์แก่งานวิจัยในด้านนั้น ๆ อย่างเต็มที่

ด้วยเหตุผลดังกล่าวแล้วข้างต้น TIC จึงต้องเตรียมข้อมูลถึง 2 ชุด เพื่อปรับเข้าระบบใหม่ (INNOPAC) โดยมีขั้นตอนในการดำเนินงานดังนี้

1. ตรวจสอบความถูกต้องในการลงรายการของข้อมูลที่มีอยู่ทั้งหมด ไม่ว่าจะเป็นเรื่องกฎเกณฑ์เกี่ยวกับรูปแบบทางบรรณานุกรมของหนังสือ/เอกสาร (CUMARC) และรูปแบบทางบรรณานุกรมของสิ่งพิมพ์ต่อเนื่อง (CUMARC - Serials) รวมถึงการพิสูจน์อักษร

2. จัดทำ Database profile โดยการศึกษาเปรียบเทียบระหว่างเขตข้อมูลแต่ละเขตที่เราลงรายการไว้กับเขตข้อมูลของระบบใหม่ ที่สำคัญคือศึกษาว่าเขตข้อมูลใดบ้างที่สามารถกำหนดให้มีการสืบค้นได้

3. แก้ไขเพิ่มเติมรายละเอียดบางรายการที่ระบบใหม่ไม่สามารถจัดการได้ อาทิเช่น เพิ่มเครื่องหมายวรรคตอนตามที่กำหนดไว้ใน AACR II เนื่องจากการบันทึกข้อมูลด้วยโปรแกรม Mini - Micro CDS/ISIS ไม่ได้ใส่เครื่องหมายต่าง ๆ ลงไว้ เพราะสามารถเขียนคำสั่งเพิ่มเติมใน display format ได้ แต่ระบบใหม่จะแสดงรายการตามที่บันทึกเท่านั้น หรือกรณีของชื่อเรื่อง ซึ่งรูปแบบทางบรรณานุกรมได้กำหนดไว้ว่าถ้ามีชื่อเรื่อง 2 ภาษา ให้ใส่ชื่อเรื่องภาษาแรกไว้ที่ tag 245 ^a : main entry และชื่อเรื่องอีกภาษาหนึ่งให้ไว้ที่ tag 245 ^I : parallel title แต่เมื่อจะเข้าระบบใหม่ ต้องย้ายข้อมูลจาก tag 245 ^I ไปไว้ที่ tag 740 : variant title จึงจะสืบค้นได้ เพราะเขาจัดการกับข้อมูลตามรายการที่ปรากฏใน USMARC เท่านั้น ดังนี้เป็นต้น

4. รูปแบบการแสดงข้อมูล ซึ่งจะต้องพิจารณาทั้งส่วนที่เป็น Public display หรือการแสดงผล สำหรับผู้ใช้บริการ และ Staff mode หรือการแสดงผลสำหรับเจ้าหน้าที่ TIC การพิจารณา ในเรื่องนี้ต้องทำอย่างรอบคอบ เพื่อให้ผู้ใช้บริการสามารถอ่านรายละเอียดต่าง ๆ และเห็น

ความแตกต่าง ของ ข้อมูลได้อย่างชัดเจน และเพื่อให้เจ้าหน้าที่สามารถปฏิบัติงานได้อย่างคล่องตัว เพราะฐานข้อมูล TIC มีรายการทั้ง หนังสือ/เอกสาร และบทความอยู่รวมกัน

โครงสร้างข้อมูลในฐานข้อมูล TIC

รายละเอียดในแต่ละระเบียนของฐานข้อมูล TIC ลงรายการโดยยึดถือตาม CUMARC เป็นหลักแต่ต้องปรับเปลี่ยนให้เหมาะสมกับลักษณะของข้อมูลซึ่งมีทั้งเอกสารที่เป็นหนังสือ/เอกสาร และบทความจากวารสารและหนังสือพิมพ์ ดังนั้นแผ่นงานของฐานข้อมูล TIC จึงต้องเตรียมพร้อมสำหรับ หนังสือ/เอกสาร และ บทความ โดยมีรายการดังนี้

008 Fixed-length data elements
020 ISBN/ISSN
041 Language code
099 Accession Number
100 Personal name - main entry
110 Corporate name - main entry
111 Meeting name - main entry
244 Article name
245 Title statement / Journal or Newspaper name
260 Imprint
300 Physical description
344 Article sources
500 General note
502 Dissertation note
520 Abstract
600 Personal name descriptors
610 Corporate name descriptors
650 Thesaurus descriptors
651 Geographic name descriptors
690 Local descriptors
691 Local geographic descriptors
693 Subject category field & broader term
700 Personal name - added entries

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

710 Corporate name - added entries

711 Meeting name - added entries

740 Variant title

850 Holding institutions

สรุป

ระบบการจัดเก็บและสืบค้นข้อสนเทศของศูนย์เอกสารประเทศไทยที่ Battelle Memorial Institute เป็นผู้จัดวางระบบไว้นั้น ได้เตรียมพร้อมที่จะปรับเปลี่ยนเข้าสู่ระบบอัตโนมัติตั้งแต่แรก ข้อมูลแต่ละรายการในบัตรรายการของ TIC ตรงกับเขตข้อมูลหลักในระบบอัตโนมัติสามารถถ่ายข้อมูลเข้าระบบอัตโนมัติโดยตรงได้ ถ้าไม่กังวลถึงรูปแบบทางบรรณานุกรม เพื่อให้เป็นมาตรฐานเดียวกับฐานข้อมูลหลัก

ข้อดีของระบบอัตโนมัติที่เห็นได้อย่างชัดเจนก็คือ เมื่อครั้งที่ใช้บัตรรายการ การกำหนดคำสำคัญมีข้อจำกัดที่การเรียงบัตรรายการในตู้บัตร จึงกำหนดให้ใช้คำสำคัญในลักษณะที่เป็น Broader term เช่น politics economics industry แล้วใช้วิธีการจัดด้วยเทคนิคพิเศษเพื่อให้ผู้ใช้บัตรรายการทราบว่าเอกสารนั้น ๆ เป็นเรื่องของ international relations หรือ local government เป็นเรื่องของ economic policy หรือ economic conditions เป็นเรื่องของ energy หรือ textile industry ดังนี้เป็นต้น เมื่อมีการปรับเปลี่ยนเข้าสู่ระบบอัตโนมัติ การกำหนดคำสำคัญสามารถใช้คำที่เป็น Narrower term ที่ตรงกับเนื้อหาของเอกสารนั้น ๆ ได้โดยตรงทำให้การสืบค้นข้อมูลสะดวกรวดเร็วและตรงกับความต้องการของผู้ใช้บริการนับว่าเป็นการสนองนโยบายของจุฬาลงกรณ์มหาวิทยาลัยในด้านการใช้เทคโนโลยีสารสนเทศเพื่อพัฒนาระบบห้องสมุดและศูนย์เอกสารเพื่อให้การบริการวิชาการดำเนินไปอย่างมีประสิทธิภาพ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย