การพัฒนาคลังเอกสารดิจิทัลระบบเปิดด้วยโอเพนซอร์สซอฟต์แวร์ Drupal

บุญเลิศ อรุณพิบูลย์
หัวหน้างาน งานพัฒนาและบริการสื่อสาระดิจิทัล
ฝ่ายบริการความรู้ทางวิทยาศาสตร์และเทคโนโลยี
สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.)

แนวคิดการพัฒนาคลังเอกสารดิจิทัลระบบเปิด หรือ Open Digital Repositories มีมาระยะหนึ่งแล้วในประเทศไทย โดยเฉพาะเมื่อจุฬาลงกรณ์มหาวิทยาลัยนำซอฟต์แวร์บริหารจัดการเอกสารดิจิทัลชื่อ DSpace มาพัฒนาคลังปัญญาสถาบันของตนเอง ส่งผลให้สถาบันการศึกษาต่างๆ สนใจกับแนวคิดนี้มากขึ้น

คลังเอกสารดิจิทัลระบบเปิด คือ ฐานข้อมูลที่ถูกพัฒนามาเพื่อจัดเก็บเอกสารในรูปแบบดิจิทัล และมีการเผยแพร่ด้วยเทคโนโลยีอินเทอร์เน็ต ทั้งนี้มีฟังก์ชันของฐานข้อมูลและระบบที่อนุญาต Robot[1] ของโปรแกรมค้นหาและฐานข้อมูลอื่นๆ สามารถเข้าถึงฐานข้อมูลนี้เพื่อจัดเก็บเมทาดาทาของเอกสารที่จัดเก็บ รวมทั้งมีการกำหนดสิทธิการเข้าถึง การใช้งานที่ชัดเจน เอื้อต่อการใช้งานของผู้ใช้ หรือที่เรียกว่าการเข้าถึงแบบเปิด (Open Acess)

การเข้าถึงแบบเปิด หมายถึง การเผยแพร่สารสนเทศให้สาธารณะสามารถเข้าถึงได้ผ่านทางเครือข่ายอินเทอร์เน็ต โดยอนุญาตให้ผู้ใช้สามารถกระทำการต่างๆ ได้ ได้แก่ การอ่าน ดาวน์โหลด คัดลอก แจกจ่าย พิมพ์ออกมา ค้นหา หรือเชื่อมโยงไปยังบทความฉบับเต็ม จัดทำดรรชนี ส่งผ่านในลักษณะเป็นข้อมูลไปยังโปรแกรมต่างๆ หรือใช้เพื่อวัตถุประสงค์ทางกฎหมาย โดยปราศจากค่าใช้จ่ายหรืออุปสรรคทางด้านเทคนิคต่างๆ อันจะช่วยกำจัดอุปสรรคทางด้านค่าใช้จ่ายและลิขสิทธิ์ ทั้งนี้สารสนเทศที่เผยแพร่มักจะเป็นสารสนเทศทางด้านวิชาการ โดยเฉพาะบทความวิจัยในรูปแบบดิจิทัล

การจัดทำคลังเอกสารดิจิทัลแบบเปิด จำเป็นต้องใช้โปรแกรมสำเร็จรูปที่ถูกพัฒนาบนฐานของเทคโนโลยีที่สนับสนุนการเข้าถึงแบบเปิด โดยโปรแกรมสำเร็จรูปที่นิยมใช้กันได้แก่ DSpace, EPrints , Fedora, Open Repository, IR+, CONTENTdm และ DigiTool โดยในประเทศไทยพบว่ามีการเลือกใช้

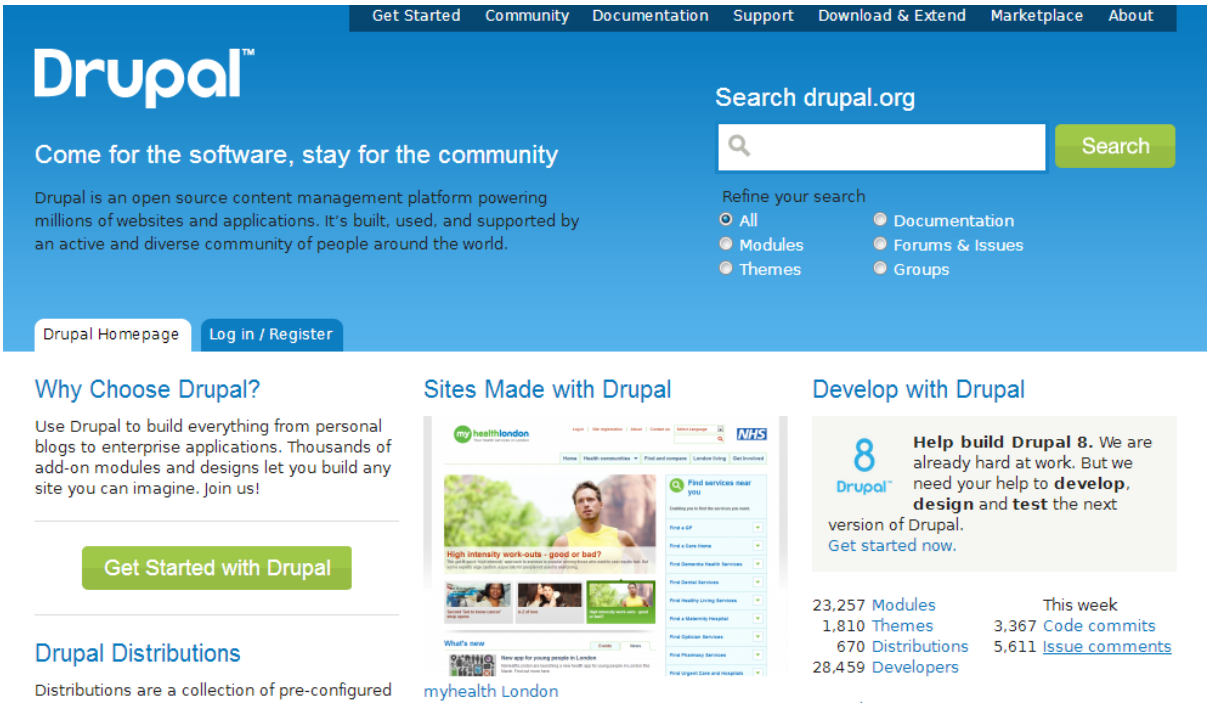
---

[1] Robot เป็นคำเรียกเครื่องมือของโปรแกรมค้นหา (Search Engine) ที่ถูกพัฒนาขึ้นมาเพื่อวัตถุประสงค์ใดๆ เช่น เพื่อจัดเก็บเมทาดาทาของสารสนเทศในระบบฐานข้อมูล

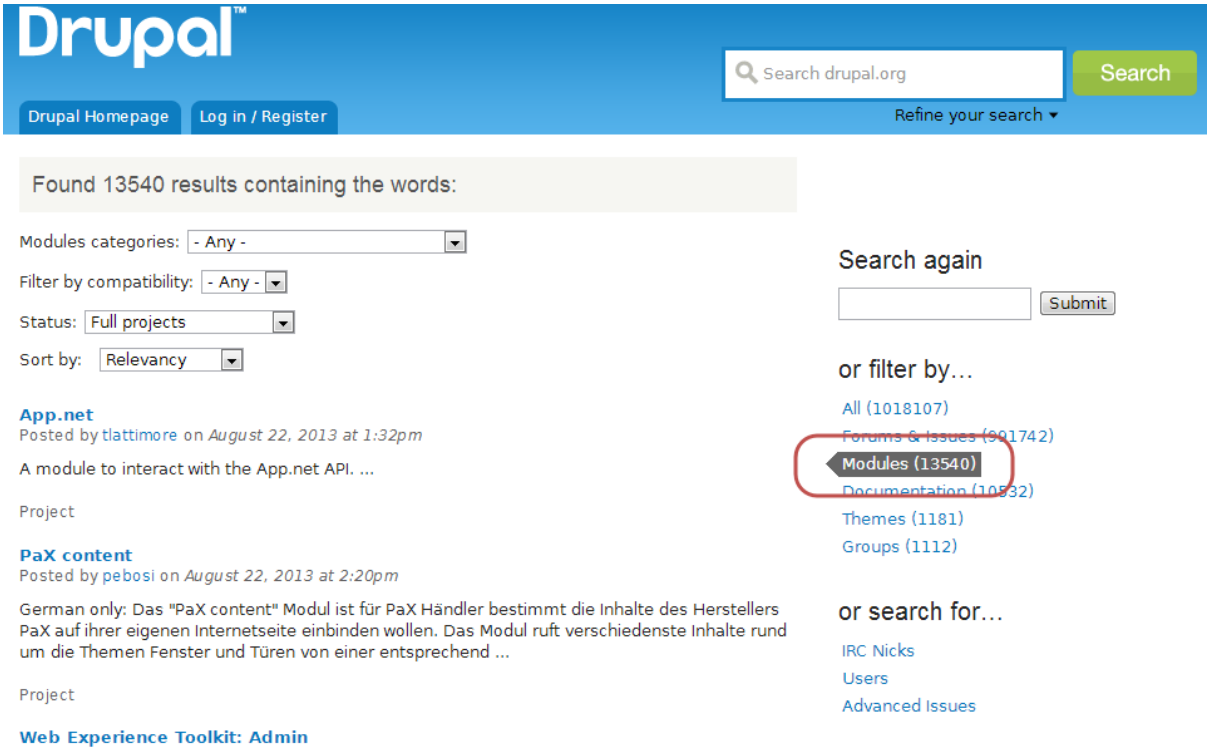
DSpace มากที่สุด อย่างไรก็ตามก็พบว่ามีหลายสถาบันในประเทศไทยที่สนใจพัฒนาคลังเอกสารดิจิทัลแบบเปิดแต่ยังไม่ประสบความสำเร็จ อันเนื่องจากสาเหตุด้านเทคนิคของโปรแกรม DSpace เช่น ความยากในการติดตั้ง ความรู้ของโปรแกรมเมอร์ในการปรับแต่งระบบ ดังนั้นการศึกษาและแนะนำโปรแกรมสำเร็จรูปอื่นๆ จึงเป็นแนวทางหนึ่งที่ฝ่ายบริการความรู้ทางวิทยาศาสตร์และเทคโนโลยี สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ สนใจศึกษา

Drupal ซอฟต์แวร์บริหารจัดการเนื้อหาเว็บ

Drupal เป็นซอฟต์แวร์บริหารจัดการเนื้อหาเว็บ หรือ Content Management System (CMS) ในกลุ่มซอฟต์แวร์เปิดเผยรหัสต้นฉบับ (Open Source Software) ที่มีโปรแกรมเสริม (Modules) จำนวนมาก ทำให้ Drupal เป็นซอฟต์แวร์บริหาจัดการเนื้อหาเว็บที่สามารถนำไปประยุกต์ใช้พัฒนาระบบงานออนไลน์ได้หลากหลายระบบ ทั้งการพัฒนาเว็บไซต์ทั่วไป เว็บไซต์องค์กร ฐานข้อมูลออนไลน์ คลังเอกสารดิจิทัล คลังเอกสารจดหมายเหตุ คลังภาพถ่ายดิจิทัล ทั้งนี้โปรแกรมหลัก Drupal และโปรแกรมเสริมของ Drupal สามารถเข้าถึงได้ผ่านเว็บไซต์ http://www.drupal.org



รูปแสดงหน้าเว็บของ Drupal.org



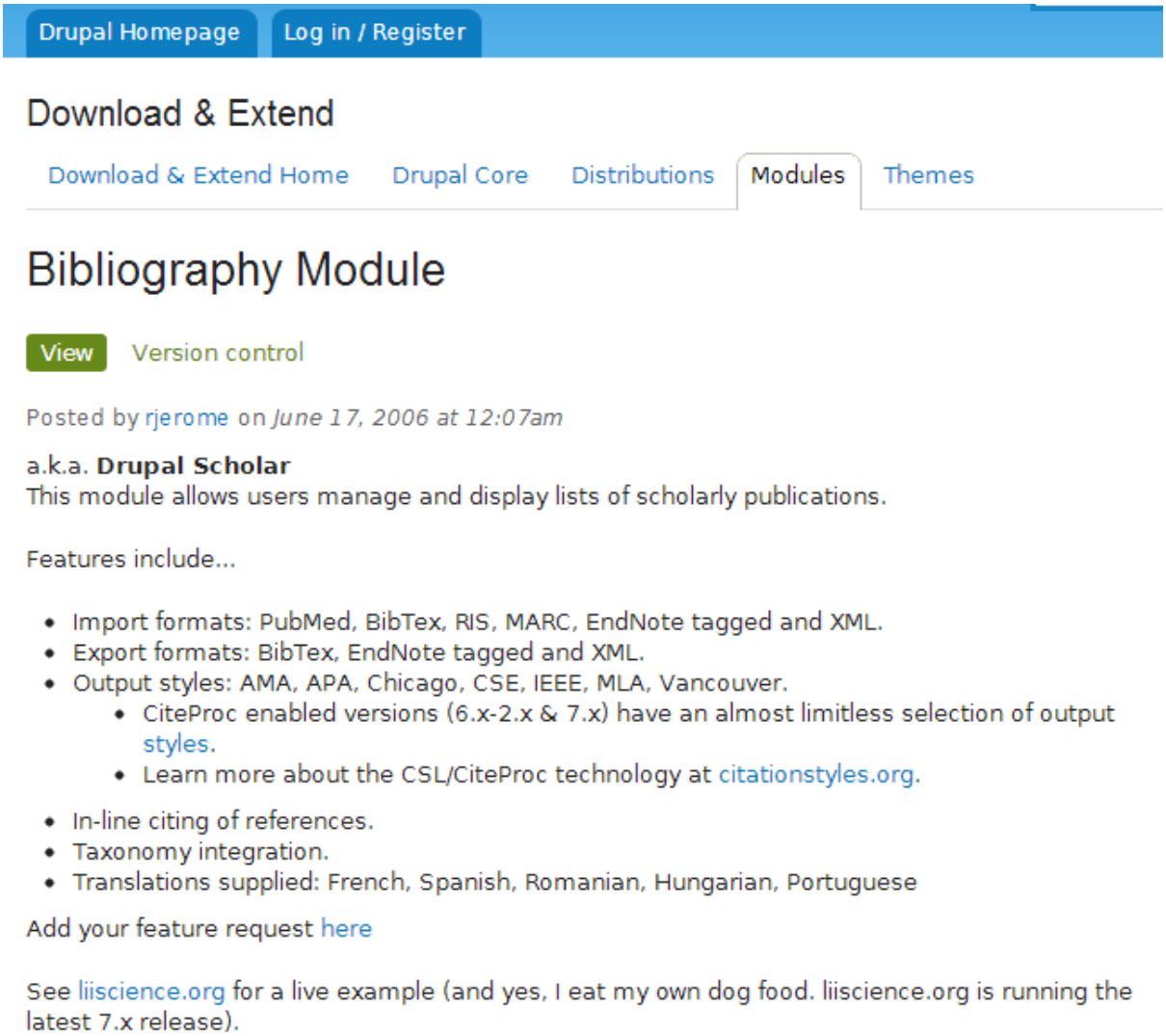
รูปแสดงโปรแกรมเสริมที่มีถึงหนึ่งหมื่นสามพันโปรแกรม

เหตุผลสำคัญที่การพัฒนาคลังเอกสารดิจิทัลระบบเปิด มักจะเลือกใช้ซอฟต์แวร์ในกลุ่มเปิดเผยรหัสต้นฉบับ ก็เพราะซอฟต์แวร์ดังกล่าว

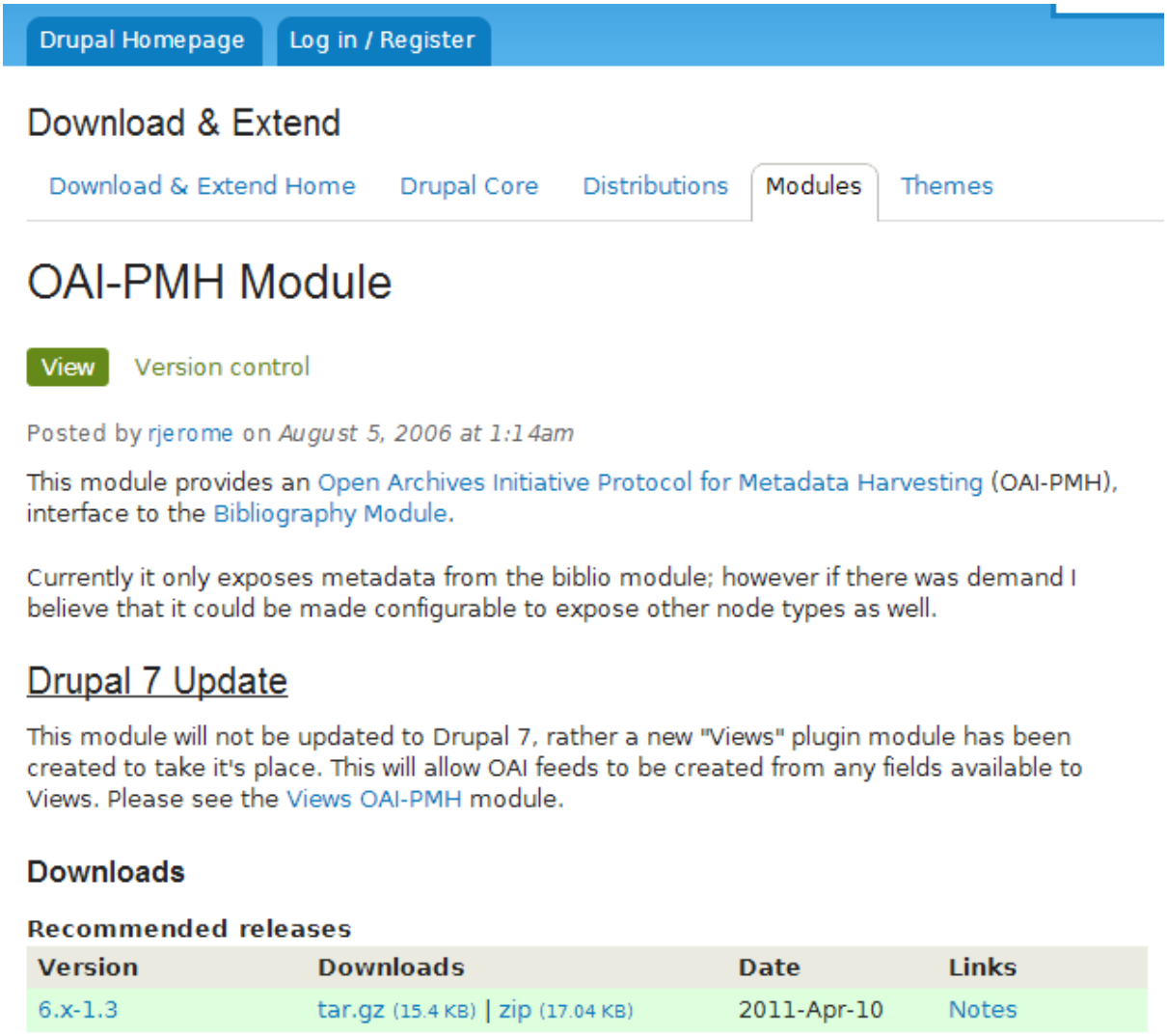
1) ไม่ต้องเสียใช้จ่ายใดในการได้มาซึ่งโปรแกรม อันเป็นการช่วยประหยัดค่าใช้จ่าย
2) สามารถใช้ในวัตถุประสงค์ใดก็ได้ อันเป็นการลดความเสี่ยงจากปัญหาด้านลิขสิทธิ์
3) สามารถเผยแพร่และแจกจ่ายโปรแกรมที่ทำการปรับแก้แล้วได้ อันเป็นการส่งเสริมให้เกิดสังคมแห่งการเรียนรู้

นอกจากนี้เนื่องจากซอฟต์แวร์ดังกล่าวอยู่ในกลุ่มเปิดเผยรหัสต้นฉบับ อันเป็นหนึ่งในลักษณะสำคัญของเทคโนโลยีระบบเปิด (Open Technology) จึงทำให้กระบวนการพัฒนาระบบเชื่อมข้อมูลแบบเปิดสามารถทำได้ง่าย สะดวกกว่าซอฟต์แวร์ในกลุ่มเทคโนโลยีระบบปิด

การพัฒนาคลังเอกสารดิจิทัลระบบเปิดด้วย Drupal โดยพื้นฐานแล้วจะใช้โปรแกรมหลัก Drupal ผสมผสานด้วยโปรแกรมเสริมชื่อ Bibliography Module และ OAI-PMH Module ก็สามารถพัฒนาเว็บไซต์คลังเอกสารดิจิทัลระบบเปิดได้แล้ว และหากต้องการพัฒนาปรับปรุงในส่วนแสดงผลก็ยังสามารถเพิ่มโปรแกรมเสริมอื่นๆ ตามที่ต้องการได้



รูปแสดงข้อมูลของ Bibliography module



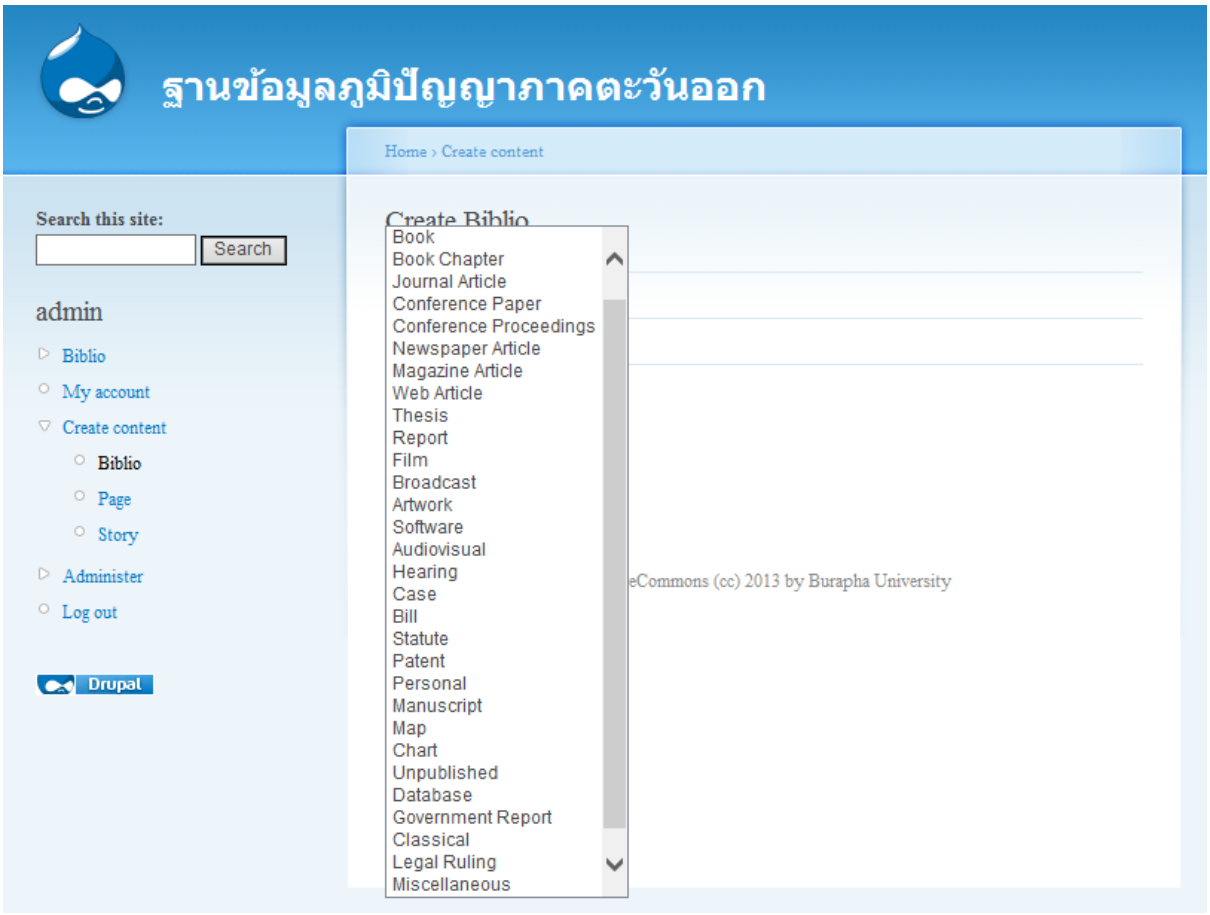
รูปแสดงรายละเอียดของ OAI-PMH Module

ความสามารถของ Drupal กับการพัฒนาคลังเอกสารดิจิทัลระบบเปิด

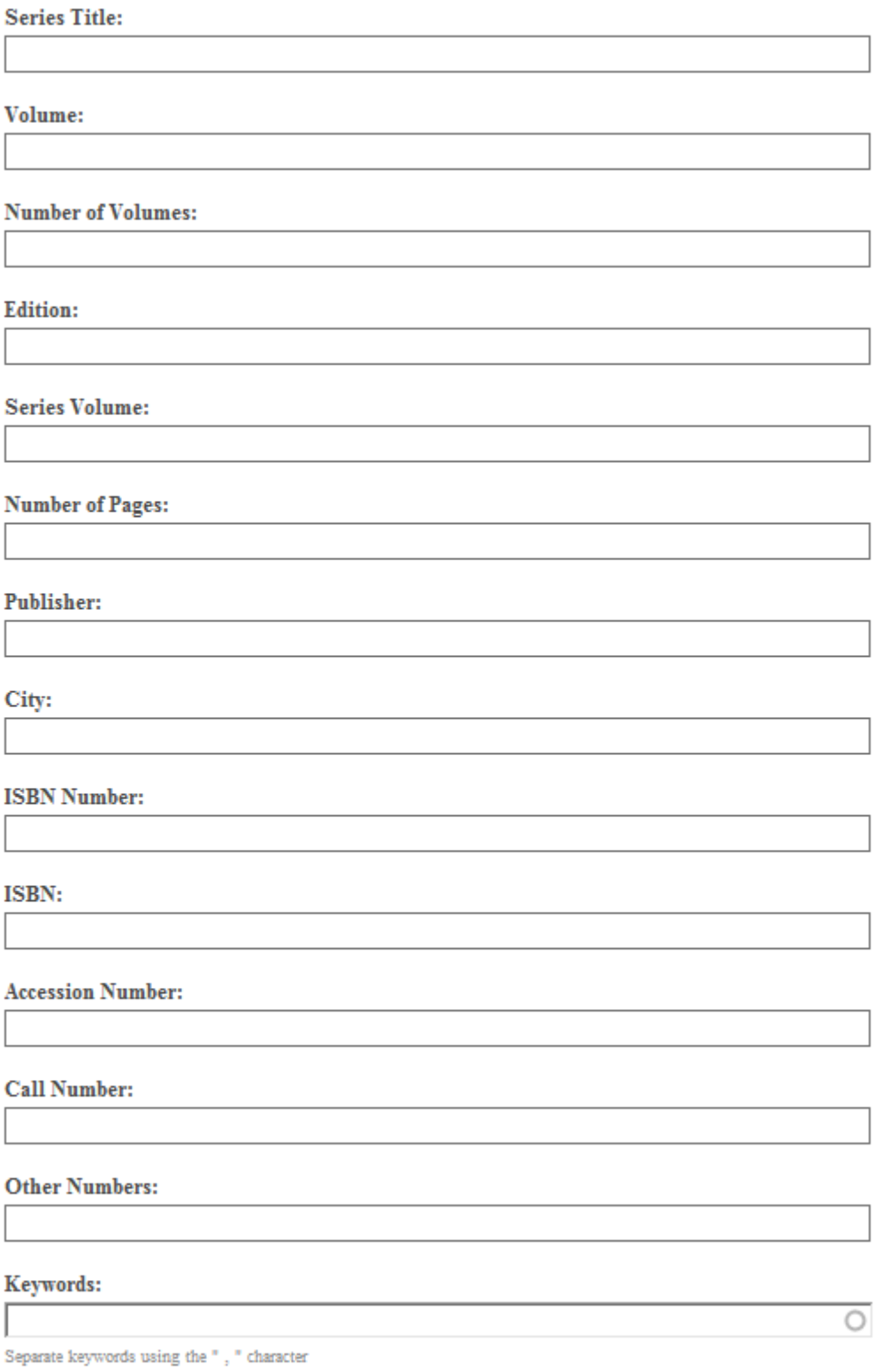
การพัฒนาคลังเอกสารดิจิทัลระบบเปิดด้วย Drupal จะเหมาะสมกับหน่วยงานที่มีความต้องการความพร้อม ดังต่อไปนี้

บุคลากรที่ดูแลงานด้านระบบ งานเทคนิค มีความรู้ความเชี่ยวชาญในภาษาโปรแกรมมิ่ง PHP โดยพบว่าบุคลากรด้านนี้ในประเทศไทย มีพื้นฐานความรู้และทักษะในโปรแกรมนี้มากกว่าโปรแกรมภาษาอื่น ทำให้การพัฒนาปรับปรุงความสามารถของโปรแกรม Drupal เป็นไปได้โดยง่าย สะดวก รวดเร็ว

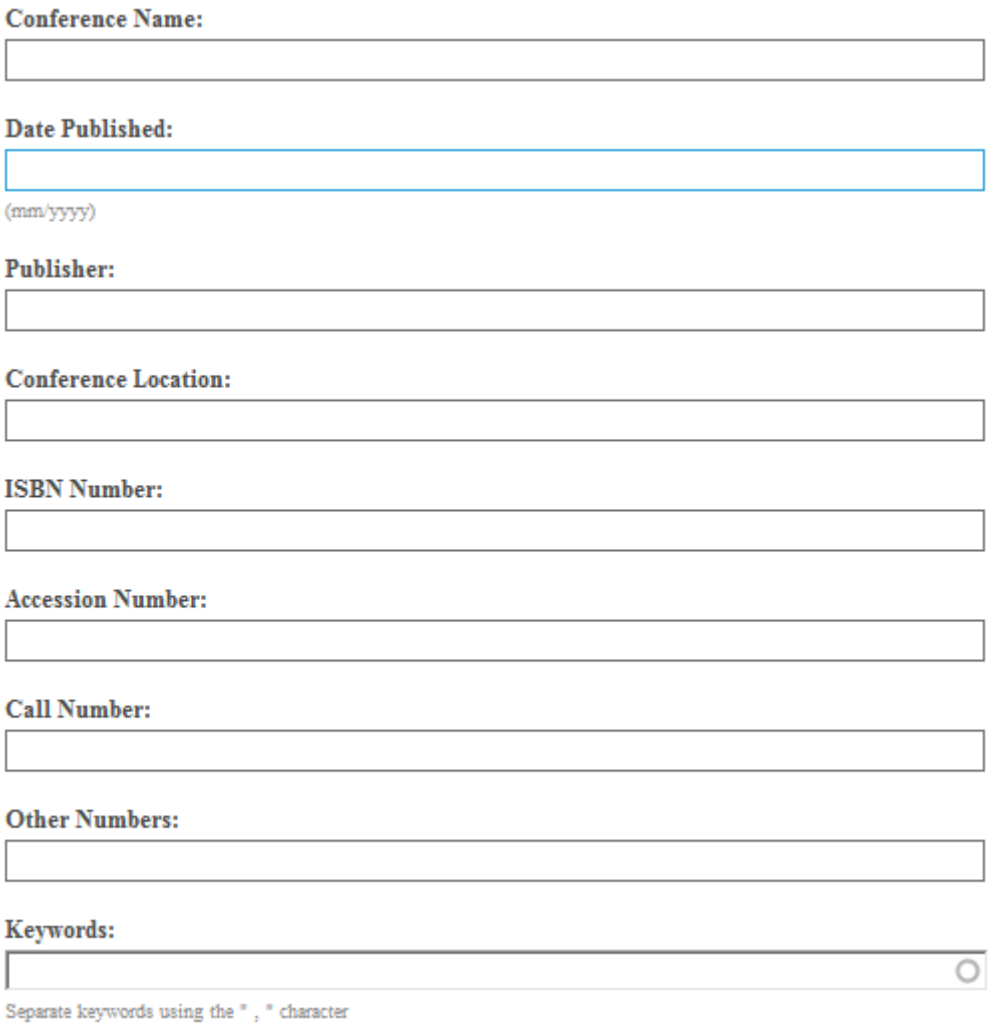
คลังเอกสารดิจิทัลระบบเปิดที่สนับสนุนการลงรายการ การนำเข้าสารสนเทศหลายประเภท ด้วยความสามารถของ Bibliography module ของ Drupal ทำให้สามารถเลือกประเภทของทรัพยากรสารสนเทศได้ตรงกับความต้องการ โดยโปรแกรมจะปรับปรุงรายละเอียดของฟิลด์ข้อมูล (Field) ให้เหมาะสมในแต่ละประเภททรัพยากร



รูปแสดงประเภทของทรัพยากรสารสนเทศที่สามารถนำเข้าระบบ



รูปแสดงส่วนรายละเอียดการลงรายการของทรัพยากรประเภทหนังสือ

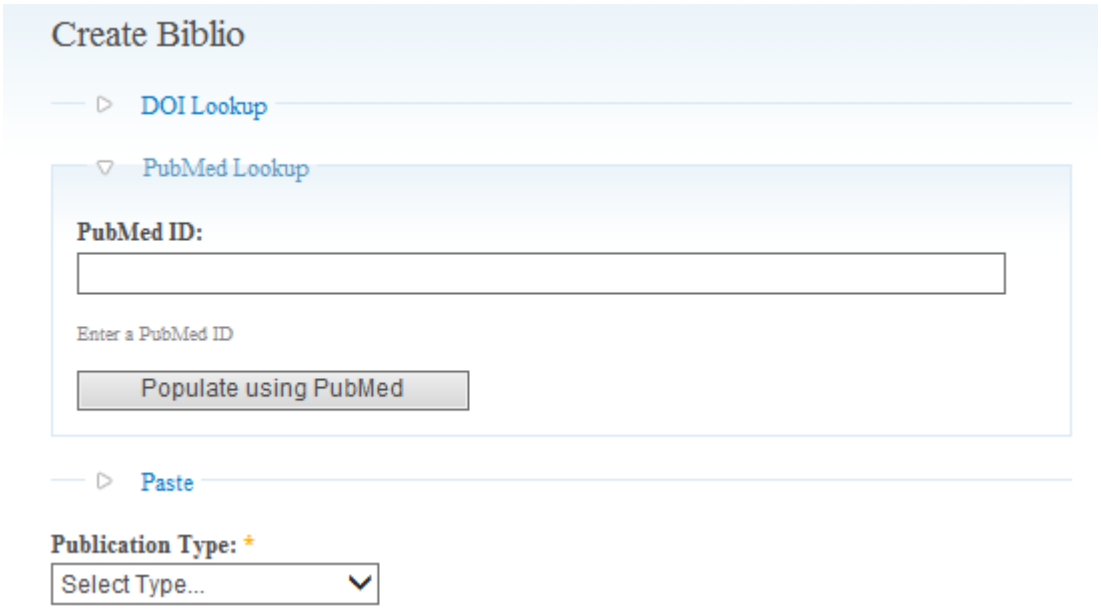


รูปแสดงส่วนรายละเอียดการลงรายการของทรัพยากรประเภท Conference

ทั้งนี้ประเภทของทรัพยากรสารสนเทศ และรายละเอียดของฟิลด์ต่างๆ โปรแกรม Bibliography กำหนดให้โดยอัตโนมัติ สามารถใช้งานได้ทันทีที่ติดตั้งโปรแกรมโดยไม่ต้องเสียเวลามาสร้าง อย่างไรก็ตาม โปรแกรมก็อนุญาตให้ปรับแก้ไขประเภท ชื่ออธิบาย และวิธีการลงข้อมูลได้ตามที่ผู้ใช้งานต้องการ

หน่วยงานต้องการลดความซ้ำซ้อนจากการนำเข้าข้อมูล การนำเข้าข้อมูลเป็นปัญหาสำคัญปัญหาหนึ่งของหน่วยงาน เพราะหน่วยงานมักจะมีการลงรายการข้อมูลไว้ก่อนหน้าอยู่แล้ว หรือมีข้อมูลปรากฏในแหล่งข้อมูลต่างๆ อยู่ก่อน เช่น ฐานข้อมูลออนไลน์ แต่ยังต้องมานำเข้าข้อมูลดังกล่าวสู่คลังเอกสารดิจิทัลระบบเปิดซ้ำ ส่งผลต่องบประมาณ เวลา และกำลังที่ต้องเสียไป Drupal ได้ออกแบบส่วนการนำเข้าให้

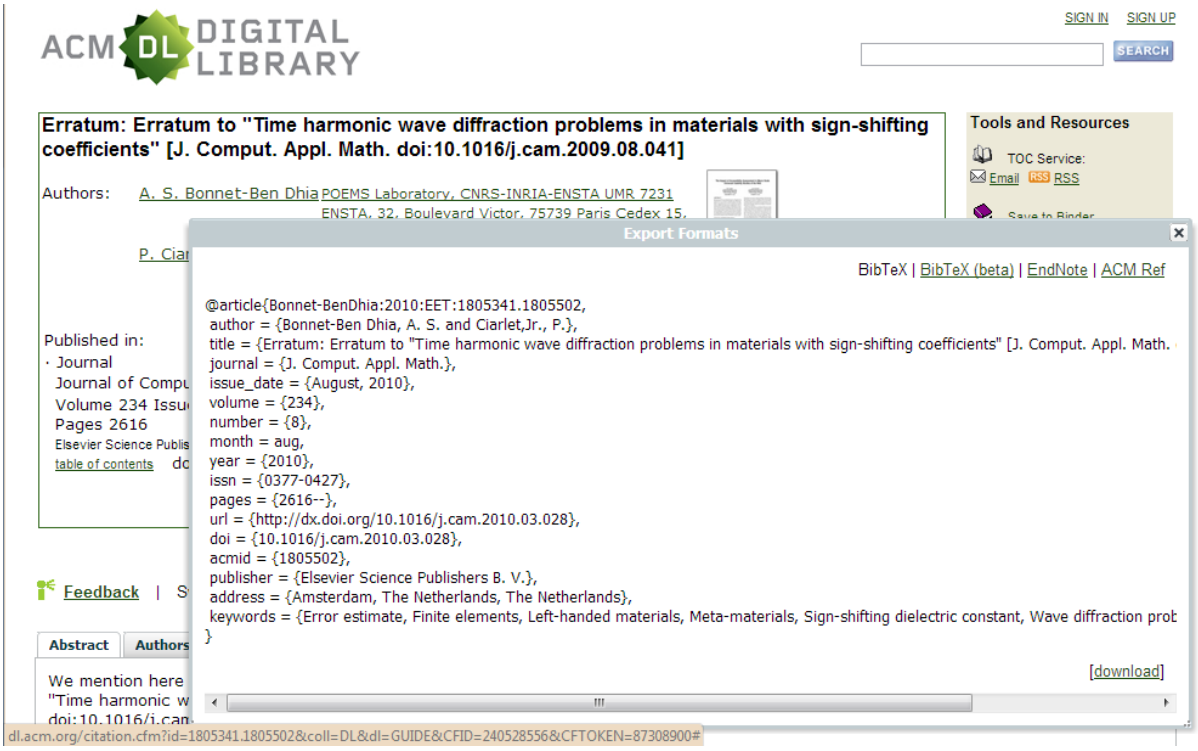
สะดวกยิ่งขึ้น โดยนอกจากรองรับการนำเข้าด้วยการพิมพ์ทีละรายการ ยังรองรับการนำเข้าด้วยโปรแกรมเลขประจำเอกสาร อันประกอบด้วย DOI: Digital Object Identifier หรือตัวระบุวัตถุดิจิทัล (ดีโอไอ) และ PubMed ID



รูปแสดงการนำเข้ารายการทรัพยากรสารสนเทศ

ทั้งนี้ DOI และ PubMed ID เป็นรหัสที่ออกแบบมาเพื่อประจำบทความวารสารวิชาการ และเมื่อป้อนรหัสดังกล่าวโปรแกรม Drupal จะเชื่อมต่ออินเทอร์เน็ตและค้นหาข้อมูลของทรัพยากรที่มีรหัสดังกล่าวและนำเข้าโดยอัตโนมัติ

นอกจากนี้ยังสามารถนำเข้าได้โดยการวาง (Paste) รหัสคำสั่งชุดที่เรียกว่า BibTeX ซึ่งเป็นรหัสคำสั่งที่ฐานข้อมูลออนไลน์ ระบบห้องสมุดอัตโนมัติได้สร้างจากบรรณานุกรมของทรัพยากรได้อีกวิธีหนึ่ง



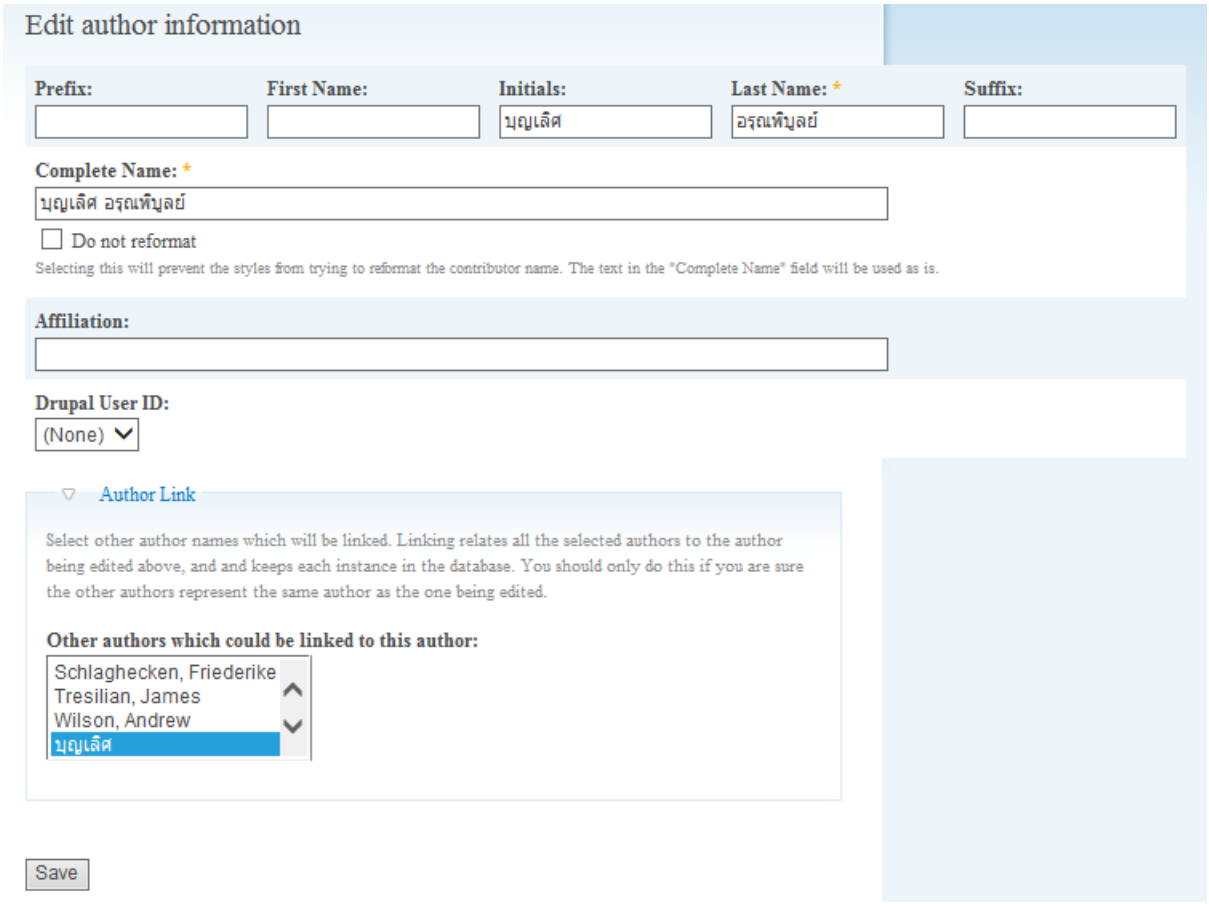
รูปแสดงชุดรหัส BibTeX จากฐานข้อมูล ACM Digital Libraries

ความสามารถการทำความสะอาดข้อมูล ปัญหาหนึ่งของฐานข้อมูล คลังเอกสารต่างๆ ก็คือ ข้อมูลที่ลงรายการอาจจะมีความแตกต่าง ความผิดพลาด เช่น ชื่อบุคคลมีหลายรูปแบบ คำค้นมีความแตกต่างกัน บุคคลมีการเปลี่ยนชื่อ นามสกุล ดังนั้นการประมวลผลข้อมูลก็ผิดพลาดตามไปด้วย จุดเด่นของ Drupal คือ ความสามารถที่เรียกว่า Global Change และ Author Merge ทำให้การทำความสะอาดข้อมูลโดยเฉพาะชื่อบุคคล และคำค้นเป็นไปได้อย่างสะดวกรวดเร็ว

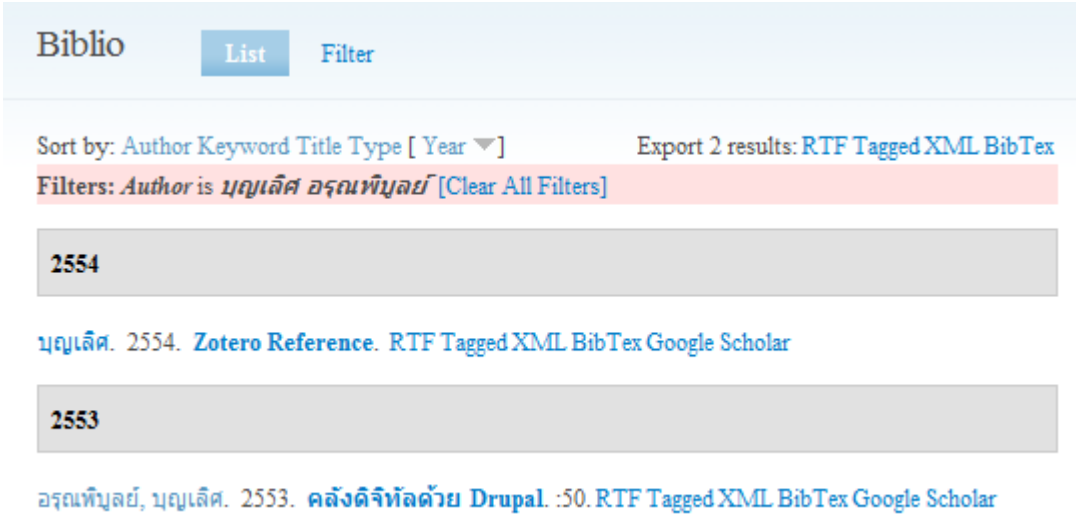


รูปแสดงการนำชื่อ และค้นคำมาประมวลผลโดยอัตโนมัติ

จากรูปข้างต้น จะพบว่ามีการนำเข้าข้อมูลชื่อบุคคลแตกต่างกัน คือ บุญเลิศ อรุณพิบูลย์ และบุญเลิศ โดยสองชื่อนี้เป็นบุคคลเดียวกัน ซึ่งการปรับแก้ไขทำได้โดยใช้ความสามารถ Global Change หรือจะใช้วิธีการรวมชื่อ (Author Merge) ก็ได้



รูปแสดงการทำความสะอาดข้อมูลของ Drupal



รูปแสดงผลลัพธ์การค้นที่ชื่อมีการรวมกันแล้ว

การใช้ Drupal ในประเทศไทย

แม้ว่า Drupal จะเป็นซอฟต์แวร์ใหม่ของประเทศไทยที่ถูกเลือกใช้เพื่อพัฒนาคลังเอกสารดิจิทัลระบบเปิด แต่หน่วยงานระดับหน่วยงานให้ทุนวิจัยในประเทศไทยหลายหน่วยก็เลือกใช้ Drupal เพื่อพัฒนาฐานข้อมูลผลงานวิชาการของตนเอง อันได้แก่ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ สำนักงานกองทุนสนับสนุนการวิจัย และสำนักงานคณะกรรมการวิจัยแห่งชาติ และได้ดำเนินการเชื่อมฐานข้อมูลด้วยมาตรฐานระบบเปิดกับระบบคลังข้อมูลงานวิจัยไทย ซึ่งปัจจุบันเข้าถึงได้ที่เว็บ http://www.tnrr.in.th



รูปแสดงหน้าเว็บไซต์ระบบคลังข้อมูลงานวิจัยไทย

ภาคสรุป

การเลือกใช้ซอฟต์แวร์ในกลุ่มเปิดเผยรหัสต้นฉบับพัฒนาคลังจัดเก็บเอกสารดิจิทัลระบบเปิดจำเป็นต้องพิจารณาในหลายๆ องค์ประกอบ เนื่องจากซอฟต์แวร์ก็มีความแตกต่างกันในเรื่องคุณสมบัติ การใช้งานและความสามารถ ดังนั้นการเลือกใช้งานจะต้องเลือกใช้โปรแกรมให้เหมาะสมกับรูปแบบ และแนวทางการจัดการที่วางแผนไว้

Drupal เป็นเพียงหนึ่งทางเลือกที่อาจจะเหมาะสมกับหน่วยงานบางหน่วยงาน และอาจจะไม่เหมาะสมกับบางหน่วยงานได้เช่นกัน ดังนั้นการศึกษาบนฐานการพิจารณาที่รอบคอบถึงความเหมาะสมกับชนิดของทรัพยากรสารสนเทศที่ต้องการจัดเก็บ ระยะเวลาในการเก็บ ความสามารถของเทคโนโลยีที่มีอยู่ในการรองรับ ค่าใช้จ่ายในการติดตั้งและดูแลระบบ และความสามารถของบุคลากร จึงเป็นประเด็นสำคัญประเด็นหนึ่งที่หน่วยงานต้องใส่ใจเป็นพิเศษ

เอกสารอ้างอิง

บุญเลิศ อรุณพิบูลย์. (2556). การจัดทำคลังเอกสารดิจิทัลด้วย Drupal 7.19. Library & Information Science Knowledge Sharing. Retrieved August 02, 2013, from http://www.thailibrary.in.th/2013/03/08/drupal-digital-collections/

บุญเลิศ อรุณพิบูลย์. (2554). การพัฒนา IR – Institutional Repository ของ สวทช. Library & Information Science Knowledge Sharing. Retrieved August 12, 2013, from http://www.thailibrary.in.th/2011/09/14/nstda-ir/

บุญเลิศ อรุณพิบูลย์. (2554). OAI-URL ของระบบงานวิจัยหน่วยงานต่างๆ. Retrieved August 12, 2013, from http://www.tnrr.in.th/index.php/knowledges/48-oai-url

ประสิทธิชัย เลิศรัตนเคหกาล. (2552). การพัฒนาโปรแกรมรหัสเปิดระบบห้องสมุดอัตโนมัติ OpenBiblio สำหรับหอสมุดแห่งชาติรัชมังคลาภิเษก เชียงใหม่. มหาวิทยาลัยเชียงใหม่.

วัลยา ภู่สว่าง. (2554). ภารกิจในการเข้าถึงเอกสารแบบเปิด : ระดับมหาวิทยาลัยและแหล่งทุน. Library & Information Science Knowledge Sharing. Retrieved August 12, 2013, from http://www.thailibrary.in.th/2013/04/30/oa-article/

สำนักงานคณะกรรมการวิจัยแห่งชาติ. (n.d.). ตัวระบุวัตถุดิจิทัล (ดีโอไอ) | DOI. Retrieved August 12, 2013, from http://www.nrct.go.th/th/doi.aspx#.UhaLsT9byuI

Suber, P. (2009). Open access in 2008. Journal of Electronic Publishing, 12(1). Retrieved from http://quod.lib.umich.edu/cgi/t/text/idx/j/jep/3336451.0012.104/--open-access-in-2008?rgn=main;view=fulltext