

การพัฒนาระบบตรวจสอบการเชื่อมโยงแหล่งทรัพยากรสากล (URL) แบบอัตโนมัติ

ด้วย Azure AI Service

อภิช เหรียญวิวัฒน์*, เพ็ญสุภา แก้วพิทยาภรณ์

สำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์

50 ถ.งามวงศ์วาน ลาดยาว จตุจักร กรุงเทพฯ 10900

Development of Automated URL Correction using Azure AI Service

Apiyos Rienvipattana*, Pensupa Kaewpitthayaporn

Kasetsart University Library

50 Ngam Wong Wan Rd. Lat Yao Chatuchak Bangkok 10900

E-mail: apiyos.r@ku.th

► รับบทความ 19 มกราคม 2567 ► แก้ไขบทความ 01 เมษายน 2567 ► ตอรับบทความ 22 เมษายน 2567

บทคัดย่อ

การพัฒนาระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยงแหล่งทรัพยากรสากลแบบอัตโนมัติ ด้วย Azure AI Service มีวัตถุประสงค์สำคัญในการดำเนินการพัฒนาระบบเพื่อเพิ่มประสิทธิภาพการเข้าถึงแหล่งทรัพยากรสารสนเทศให้แก่ผู้ใช้บริการสามารถเข้าถึงได้อย่างถูกต้องและรวดเร็ว ด้วยการปรับรูปแบบการทำงานการตรวจสอบการเชื่อมโยง URL ของทรัพยากรสารสนเทศที่ให้บริการบนฐานข้อมูลทรัพยากรห้องสมุด (WebOPAC) จากกระบวนการแบบ Manual) เป็นการทำงานด้วยระบบอัตโนมัติ ส่งผลให้ลดระยะเวลาและขั้นตอนการปฏิบัติงาน รวมถึงการเพิ่มประสิทธิภาพการบริการ ซึ่งผู้พัฒนาระบบใช้หลักการและขั้นตอนตามวิธีการของวงจรการพัฒนา (SDLC: System Development Life Cycle) โดยดำเนินการด้วย Azure AI Service และใช้ Python 3 ในการพัฒนาโปรแกรม จากผลการดำเนินการพบว่าระบบช่วยลดขั้นตอนกระบวนการและระยะเวลาในการปฏิบัติงานในส่วนของการตรวจสอบความถูกต้องของรูปแบบการเชื่อมโยง ซึ่งเกิดจากความผิดพลาดของผู้กรอกข้อมูล 2) การตรวจสอบสถานะของการเชื่อมโยงและปรับปรุงแหล่งการเชื่อมโยงใหม่ โดยกระบวนการทั้งหมดใช้เวลาเฉลี่ยที่ 1-2 วินาที ต่อ 1 รายการ ทั้งนี้ ระบบสามารถนำไปประยุกต์ใช้กับการทำงานหรือระบบอื่น ๆ โดยไม่จำกัดแค่ระบบห้องสมุดอัตโนมัติเพียงอย่างเดียว เช่น ระบบ Digital Repositories ระบบ Journal System รวมถึงแหล่งเชื่อมโยงของเว็บไซต์ที่เชื่อมกับหน้าเว็บต่าง ๆ โดยไม่จำเป็นต้องจัดสรรงบประมาณเพื่อการจัดซื้อหรือจ้างพัฒนาระบบ

คำสำคัญ

การตรวจสอบการเชื่อมโยง, ลิงก์เสียหาย, บริการปัญญาประดิษฐ์

Abstract

Development of the Automated URL Validation, Status Monitoring and Correction using Azure AI Service aims to enhance the efficiency of end-user access to library resources. This allows users to access correctly and swiftly by adjusting the URL verifying process of library resources on WebOPAC from manual

ผลงานนี้ได้ผ่านการนำเสนอในการประชุมวิชาการระดับชาติ PULINET ครั้งที่ 14 วันที่ 10-12 มกราคม 2567
จัดโดยศูนย์บรรณสารและสื่อการศึกษา มหาวิทยาลัยแม่ฟ้าหลวง และหน่วยงานห้องสมุดมหาวิทยาลัยสวนภูมิภาค (PULINET)

to automatic operation. As a result, it is reducing time and steps in the operational process. Such optimizations collectively serve to enhance overall service effectiveness. The system developers employ principles and processes in accordance with the System Development Life Cycle (SDLC), employing Azure AI Service and utilizing Python3 for program development. From the operational results, it was found that the system helps reduce the steps and time in the operational process in the following aspects: 1. Verifying the correctness of the linking format, which arises from errors made by data entry personnel. 2. Checking the status of the URL and updating new sources, the entire process takes an average of 1-2 seconds per item. Furthermore, this system is not limited to library systems but may be applied to a variety of operations or systems. As an example, Digital Repositories, Journal Systems, and even website links to various web pages can be benefited from this system without the necessity allocating budgets for system development or procurement.

Keywords

URL Checker, Broken Link, AI Service

บทนำ (Introduction)

สำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์ให้บริการเอกสารฉบับเต็มบนฐานข้อมูลทรัพยากรห้องสมุด (WebOPAC) ด้วยทรัพยากรสารสนเทศหลากหลายประเภท ไม่ว่าจะเป็นวิทยานิพนธ์ ปัญหาพิเศษ หนังสือ วารสาร และฐานข้อมูลต่าง ๆ ซึ่งมีแหล่งที่จัดเก็บแตกต่างกันไป ตามลักษณะที่มาของทรัพยากรสารสนเทศ เช่น วิทยานิพนธ์หรือปัญหาพิเศษของมหาวิทยาลัยเกษตรศาสตร์ มีการจัดเก็บในเซิร์ฟเวอร์ (Server) ของสำนักหอสมุด ส่วนหนังสือ วารสาร และฐานข้อมูลต่าง ๆ มีที่มาทั้งจากการจัดซื้อจากตัวแทนจำหน่าย หรือสำนักพิมพ์ต่าง ๆ การได้รับอนุญาตจากผู้ผลิต ผู้จัดพิมพ์ หรือแหล่งข้อมูลวิชาการแบบเปิด (Open Access Scholarly Resources) ข้อมูลดังกล่าวมีผลให้ผู้ใช้บริการสามารถเข้าถึงข้อมูลที่ต้องการได้อย่างสะดวกรวดเร็ว

อย่างไรก็ตาม การเข้าใช้งานเอกสารฉบับเต็มหรือทรัพยากรสารสนเทศอิเล็กทรอนิกส์ดังกล่าว อาจเกิดปัญหาแก่ผู้ใช้ที่สำคัญคือ การไม่สามารถเข้าใช้งานได้ อันเนื่องมาจากสาเหตุหลายประการ เช่น มีการเปลี่ยนแปลงสิทธิ์การเข้าใช้ การเปลี่ยนแปลงสถานที่จัดเก็บโดยมีการเปลี่ยนแปลงแหล่งทรัพยากรสากล (URL: Universal Resource Locator) หรือการสูญหายของไฟล์ จากปัญหาข้างต้น ผู้เกี่ยวข้องที่รับผิดชอบต้องดำเนินการตรวจสอบ URL เป็นประจำอย่างต่อเนื่อง เพื่อป้องกันไม่ให้เกิดปัญหาดังกล่าว ในลักษณะการตรวจสอบทีละรายการแบบแมนนวล (Manual) ทำให้การดำเนินงานล่าช้า ไม่เป็นปัจจุบัน

ดังนั้น จึงมีการดำเนินการศึกษาเพื่อวิเคราะห์ ออกแบบ พัฒนาระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service และเพื่อเพิ่มประสิทธิภาพการเข้าถึงแหล่งทรัพยากรสารสนเทศให้แก่ผู้ใช้บริการสามารถเข้าถึงได้อย่างถูกต้องและรวดเร็ว ซึ่งเรียกว่า “ระบบตรวจสอบการเชื่อมโยงแหล่งทรัพยากรสากล (URL) แบบอัตโนมัติ” หรือ “ระบบ URL Checker” โดยคาดหวังว่า ระบบดังกล่าวจะสามารถลดขั้นตอน ประหยัดเวลาการตรวจสอบและเผยแพร่ข้อมูล ที่สำคัญผู้ใช้บริการได้รับข้อมูลเอกสารฉบับเต็มได้บนฐานข้อมูลทรัพยากรห้องสมุดอย่างสะดวกรวดเร็ว ถูกต้องครบถ้วนต่อไป

วัตถุประสงค์ (Objective)

1. เพื่อวิเคราะห์ ออกแบบ พัฒนาระบบตรวจสอบการเชื่อมโยงแหล่งทรัพยากรสากล (URL) แบบอัตโนมัติ
2. เพื่อเพิ่มประสิทธิภาพการเข้าถึงแหล่งทรัพยากรสารสนเทศให้แก่ผู้บริการสามารถเข้าถึงได้อย่างถูกต้อง

วิธีการดำเนินการ (Methodology)

ได้ดำเนินการพัฒนาระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service โดยใช้หลักการและขั้นตอนตามวิธีการของวงจรการพัฒนา (SDLC: System Development Life Cycle) ซึ่งเป็นวงจรที่แสดงถึงกิจกรรมต่าง ๆ ในแต่ละขั้นตอน ตั้งแต่เริ่มจนกระทั่งสำเร็จ ดังนี้



ภาพที่ 1 ขั้นตอนตามวิธีการของวงจรการพัฒนา (SDLC: System Development Life Cycle)

1. วิเคราะห์ปัญหา

ผู้บริการไม่สามารถเข้าถึงบริการเอกสารฉบับเต็มบนฐานข้อมูลทรัพยากรห้องสมุด เนื่องจากการเปลี่ยนแปลง URL และผู้เกี่ยวข้องที่รับผิดชอบดำเนินการตรวจสอบในลักษณะการตรวจสอบที่ละรายการแบบ Manual ทำให้การดำเนินงานล่าช้า ไม่เป็นปัจจุบัน

2. ออกแบบระบบ

การวิเคราะห์และออกแบบระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service เพื่อให้มีความเหมาะสมสอดคล้องกับหลักการและทฤษฎีการออกแบบ มีการทำงานเป็น 3 ฟังก์ชัน ประกอบด้วย

2.1 ฟังก์ชันตรวจสอบรูปแบบของ URL ทำหน้าที่ตรวจสอบข้อมูลของ URL ที่จัดเก็บอยู่ในเขตข้อมูล (Tag) 856 ในรูปแบบ MARC21 หากพบ Tag 856 ระบบจะทำงานอัตโนมัติ โดยดึงข้อมูลหมายเลขรายการบรรณานุกรม และข้อมูลใน Tag 856 ออกมาตรวจสอบรูปแบบของ URL ซึ่งใช้คำสั่งนิพจน์ปกติ (Regular Expression) ซึ่งเป็นคำสั่งที่ใช้ในการเลือกข้อความหรือรูปแบบอักษร ซึ่ง Regular Expression สามารถใช้ในการอธิบายหรือเลือกข้อความที่มีความซับซ้อนได้โดยการกำหนดกฎเกณฑ์ด้วยข้อความสั้น ๆ โดยมีคำสั่งในการทำงาน ซึ่งสามารถอธิบายได้ ดังนี้

```
'^(https?|ftp):\/\/([a-zA-Z0-9.-]+)(?:\[a-zA-
```

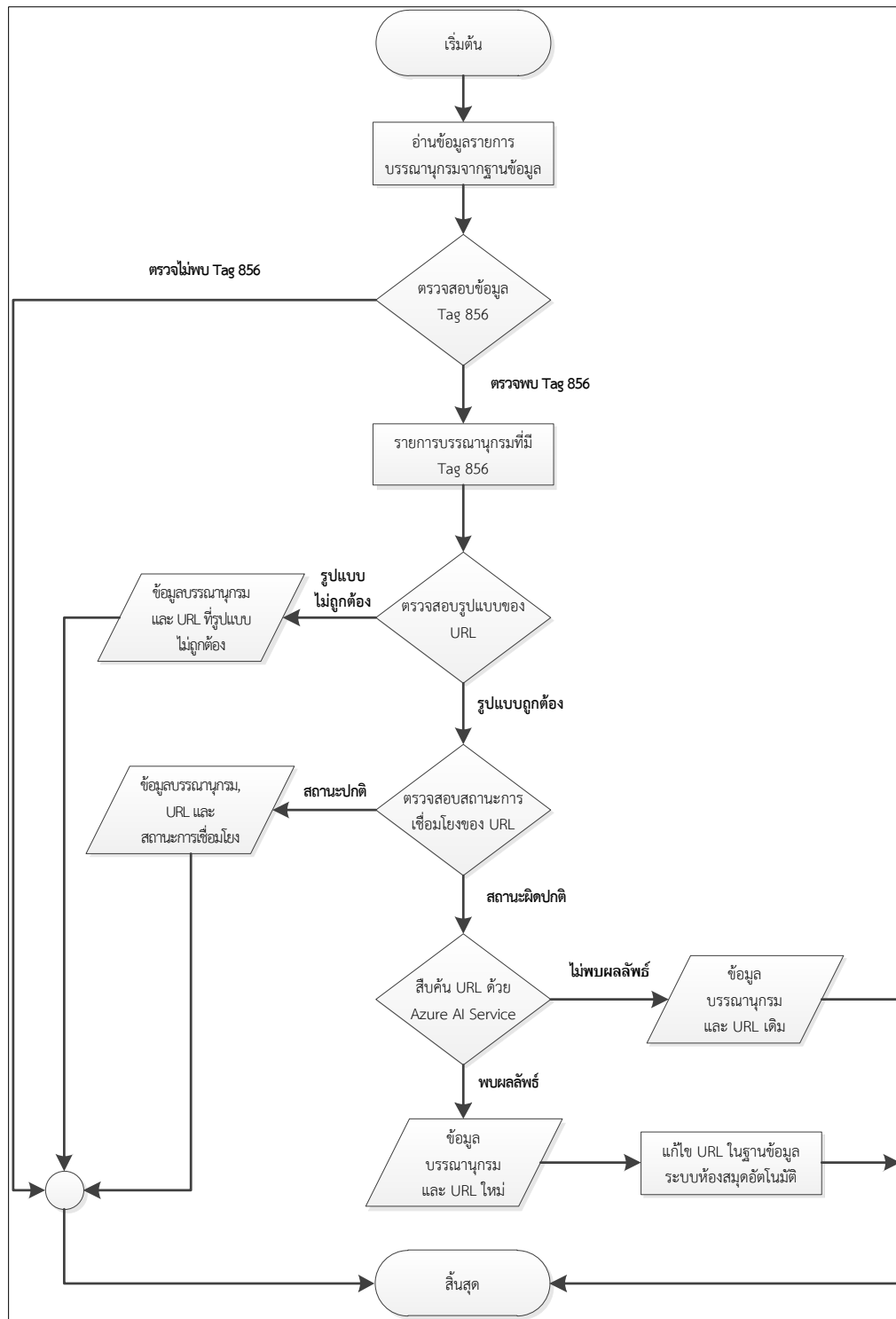
- https?|ftp ตรวจสอบแบบแผนโปรโตคอล (http, https, ftp)
- :\/\/ ตรวจสอบอักขระคำเต็มหลังต้องมีลักษณะเป็น :\/\/

- ([a-zA-Z0-9.-]+(?:\.[a-zA-Z]{2,})+) ตรวจสอบอักขระของชื่อโดเมนเนม (Domain name) โดยจะอนุญาตให้มีอักขระที่เป็นตัวอักษร ตัวเลข เครื่องหมายจุดและขีดเท่านั้น และตรวจสอบประเภทของโดเมนเนมซึ่งต้องตามด้วยเครื่องหมายจุดและอักขระค่าเต็มหลังตัวอักษรและตัวเลขเท่านั้น
- (?:[/^\\s]*)? ตรวจสอบพาร์ธ (Path) หรือ เส้นทางที่นำไปสู่หน้าของเว็บไซต์ ซึ่งต้องมีอักขระขึ้นต้นเป็น '/' และต้องไม่มีช่องว่าง รวมถึง tab, \n, \r

เมื่อทำการตรวจสอบรูปแบบของ URL แล้ว ระบบจะทำการจัดส่งรายงานผลการตรวจสอบให้ผู้ใช้งานทราบ ผ่าน Google Sheet โดยอัตโนมัติ

2.2 ฟังก์ชันตรวจสอบสถานะการเชื่อมโยงของ URL ทำหน้าที่ตรวจสอบสถานะของการเชื่อมโยง โดยนำรายการบรรณานุกรมที่มีรูปแบบของ URL ที่ถูกต้อง พร้อมกับข้อมูล MARC21 นำมาตรวจสอบสถานะการตอบสนองด้วย HTTP Status Code ที่ใช้ในการบ่งบอกสถานการณ์ทำงานของเว็บไซต์นั้น ๆ โดยใช้ไลบรารีของ Python ที่ชื่อ Requests สำหรับการส่ง HTTP request ไปยัง URL ที่ต้องการตรวจสอบ และดึงข้อมูลสถานะของการตอบสนองที่ถูกส่งกลับมา ซึ่งจะประกอบด้วยกลุ่มตัวเลข 3 ตัว และทำการกรองข้อมูลสถานะเฉพาะในกลุ่มตัวเลข 4xx (Client Error) และ 5xx (Server Error) เท่านั้น จากนั้นระบบจะทำการจัดส่งรายงานผลการตรวจสอบให้ผู้ใช้งานทราบ ผ่าน Google Sheet โดยอัตโนมัติ

2.3 ฟังก์ชันปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service ทำหน้าที่นำข้อมูลบรรณานุกรมที่มีสถานะการเชื่อมโยงที่เกิดความผิดพลาด มาทำการสืบค้น URL ใหม่ด้วย Azure AI Service โดยนำข้อมูล MARC21 ได้แก่ Tag 008 ส่วนภาษาของเนื้อหา, Tag 245 ชื่อเรื่อง, Tag 260 ส่วนสำนักพิมพ์ และส่วนปีพิมพ์ มาสร้างเป็น Query Parameter เพื่อนำไปใช้ค้นหา URL ใหม่ ผ่านโปรโตคอล REST (Representational State Transfer) ซึ่งเป็นวิธีการแลกเปลี่ยนข้อมูลกันระหว่าง Server-Client รูปแบบหนึ่งซึ่งอยู่บนพื้นฐาน HTTP โดยทำการส่ง Query Parameter ไปที่ Endpoints ของ Bing Web Search API v7 และรับข้อมูลตอบกลับเป็นข้อมูล URL ใหม่ เมื่อได้ URL ใหม่แล้ว ระบบจะนำไปปรับปรุงข้อมูลบรรณานุกรมในฐานข้อมูลระบบห้องสมุดอัตโนมัติโดยอัตโนมัติ จากนั้นระบบจะทำการจัดส่งรายงานผลการตรวจสอบให้ผู้ใช้งานทราบ ผ่าน Google Sheet โดยอัตโนมัติ



ภาพที่ 2 ผังงานระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service

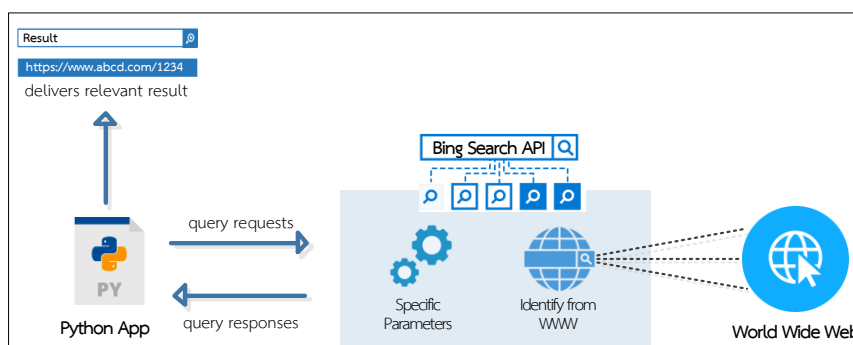
3. พัฒนาระบบ

การพัฒนาาระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service โดยใช้โปรแกรมภาษา Python สำหรับพัฒนาระบบ และเชื่อมโยงกับระบบห้องสมุดอัตโนมัติ โดยมีขั้นตอนดังนี้

3.1 ผู้พัฒนาศึกษาทฤษฎีที่เกี่ยวข้องเพื่อใช้ในการประกอบการพัฒนาระบบ

3.1.1 รูปแบบการจัดเก็บข้อมูลบรรณานุกรมรูปแบบ MARC21

- 3.1.2 หลักการทำงานของฐานข้อมูลเชิงสัมพันธ์
- 3.1.3 หลักการทำงานของโปรแกรม Sierra DNA ของระบบห้องสมุดอัตโนมัติ Sierra
- 3.1.4 หลักการทำงานของโปรโตคอลสื่อสารผ่านอินเทอร์เน็ต (ระบบเครือข่ายอินเทอร์เน็ตและอินเทอร์เน็ตโปรโตคอล)
- 3.1.5 หลักการทำงานของ Azure AI Service ผ่าน Python SDK
- 3.2 ผู้พัฒนาระบบติดตั้งโปรแกรมภาษา Python และไลบรารีที่จำเป็นต่อการพัฒนาระบบที่เครื่องคอมพิวเตอร์แม่ข่ายระบบห้องสมุดอัตโนมัติ
- 3.3 สมัครใช้บริการ Azure AI Service ที่ <https://azure.microsoft.com/en-us/products/ai-services> เพื่อใช้บริการ Bing Search Services ซึ่งเป็นแพลตฟอร์มการให้บริการสืบค้นข้อมูลด้วยปัญญาประดิษฐ์ (AI) มีการทำงานที่รวดเร็ว เหมาะกับผู้พัฒนาระบบที่ต้องการนำฟังก์ชันการสืบค้นไปประยุกต์ใช้ระบบที่พัฒนาขึ้นเอง เมื่อสมัครใช้บริการแล้ว ให้นำ API Key ที่ได้รับมาใช้ในการพัฒนาระบบเพื่อเชื่อมต่อกับ Azure AI Service



ภาพที่ 3 แสดงขั้นตอนการพัฒนาระบบ

- 3.4 พัฒนาระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service ด้วยโปรแกรมภาษา Python
- 3.5 ติดตั้งระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service ที่เครื่องคอมพิวเตอร์แม่ข่ายระบบห้องสมุดอัตโนมัติ

4. ทดสอบ

การทดสอบระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service ทำโดยการติดตั้งและทำการเปิดใช้งานระบบเพื่อเป็นการทดสอบการทำงาน ดังนี้ การเชื่อมต่อกับระบบห้องสมุดอัตโนมัติ ความถูกต้องของการดึงข้อมูลรายการบรรณานุกรม การเชื่อมต่อเครือข่ายอินเทอร์เน็ต การเชื่อมต่อกับ Azure AI Service และการปรับปรุงข้อมูลรายการบรรณานุกรมในระบบห้องสมุดอัตโนมัติ

5. ประเมินผลและปรับปรุงระบบ

กระบวนการประเมินผลการทำงานของระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service โดยการสอบถาม รับฟังปัญหาและข้อเสนอแนะจากผู้เกี่ยวข้อง เพื่อประเมินและทำการปรับปรุงระบบให้ตอบสนองความต้องการของผู้ใช้ทุกเดือน ซึ่งเป็นการเพิ่มประสิทธิภาพการทำงานของระบบ และทำการบำรุงรักษาระบบอย่างสม่ำเสมอเพื่อให้ระบบสามารถใช้งานได้อย่างปกติ

ผลการดำเนินการและอภิปรายผล (Result and Discussion)

ผลการพัฒนาระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service ได้ทำการแบ่งส่วนของฟังก์ชันการทำงานของระบบออกเป็น 3 ส่วน เพื่อลดความซับซ้อนและเพิ่มความคล่องตัวในการทำงาน โดยใช้โปรแกรมภาษา Python 3 ในการพัฒนาโปรแกรม ซึ่งมีรายละเอียดดังต่อไปนี้

1. ฟังก์ชันตรวจสอบรูปแบบของ URL มีหลักการทำงาน คือ เมื่อเปิดใช้งานแล้ว ระบบจะเชื่อมต่อกับฐานข้อมูล PostgreSQL ซึ่งเป็นฐานข้อมูลของระบบห้องสมุดอัตโนมัติ ทำการดึงข้อมูลรายละเอียดของรายการบรรณานุกรมที่มีข้อมูลใน Tag 856 มาทำการตรวจสอบรูปแบบของ URL ดังภาพที่ 5 และระบบจะทำการจัดส่งรายงานผลการตรวจสอบในรูปแบบของ Google Sheet โดยอัตโนมัติ ซึ่งจะแสดงข้อมูลประกอบด้วย BIBNO และ URL ที่ไม่ผ่านการตรวจสอบ

```
URI Validate-Verify-reVise Tools.
by Apiyos Rienvipattana
Kasetsart University Library
apiyos.r@ku.th

#####
START Validating BIB with marc tag 856 for non url content.
Connecting to Sierra Database.
Retrieving data.

Validating BIB with marc tag 856 for non url content completed!
107 row(s) for non url content.
Validating Process: 0:00:36.930009

#####
```

ภาพที่ 4 แสดงการตรวจสอบรูปแบบ URL

BIBNO	URL
b1302460	www.loc.gov/catdir/bios/jhu052/2003006211.html
b1369048	0http://www.loc.gov/catdir/enhancements/fy0707/2001271711-b.html
b1344339	+http://www.loc.gov/catdir/enhancements/fy0806/2007047402-d.html
b1169176	uhttp://www.tandfonline.com/toc/lesb20/current#.VREx603NviU
b1563647	http://bidyalib.eco.ku.ac.th/2020/IS%209821%20%E0%B8%98%E0%B8%B5%E0%B8%A3%
b1450452	www.gmw.rug.nl/~millssurvialeha/
b1557611	http://bidyalib.eco.ku.ac.th/2020/February/IS%209437.pdf
b1557600	http://bidyalib.eco.ku.ac.th/2020/February/IS%209431.pdf
b1563415	http://bidyalib.eco.ku.ac.th/2020/IS%209783%20%E0%B8%AD%E0%B8%B1%E0%B8%87%
b1557759	http://bidyalib.eco.ku.ac.th/2020/March/IS%209488.pdf
b1557582	http://bidyalib.eco.ku.ac.th/2020/February/IS%209416.pdf
b1560439	http://bidyalib.eco.ku.ac.th/2020/IS%209772.pdf
b1504549	http://bidyalib.eco.ku.ac.th/CD%20IS/is%208827.pdf
b1275332	ttp://www.loc.gov/catdir/description/mh031/2003535497.html
b1563268	http://bidyalib.eco.ku.ac.th/2020/IS%209773%20%E0%B8%95%E0%B9%89%E0%B8%AD%
b1517480	http://bidyalib.eco.ku.ac.th/2019/IS%202560/IS%209038.pdf
b1280799	www.avrdc.org/pdf/TB34.pdf
b1563413	http://bidyalib.eco.ku.ac.th/2020/IS%209781%20%E0%B9%80%E0%B8%88%E0%B8%A9%
b1558851	http://bidyalib.eco.ku.ac.th/2022/June/IS%209633.pdf
b1517797	http://bidyalib.eco.ku.ac.th/2019/IS%202560/IS%209066.pdf

ภาพที่ 5 แสดงรายการระเบียบข้อมูลที่ไม่ผ่านการตรวจสอบ

2. ฟังก์ชันตรวจสอบสถานะการเชื่อมโยงของ URL มีหลักการทำงาน คือ ระบบจะทำการดึงข้อมูลใน Tag 856 พร้อมกับ Tag 020 ISBN/ISSN, Tag 040 ส่วนภาษาการลงรายการ, Tag 245 ชื่อเรื่อง, Tag 260 ส่วนสำนักพิมพ์ และปีพิมพ์ โดยนำข้อมูลใน Tag 856 มาทำการตรวจสอบสถานะการเชื่อมโยงของ URL และแสดงผลการตรวจสอบสถานะ ดังภาพที่

7 และ ภาพที่ 8 และระบบจะทำการจัดส่งรายงานผลการตรวจสอบในรูปแบบของ Google Sheet โดยอัตโนมัติ ซึ่งจะแสดงข้อมูลประกอบด้วย BIBNO, Title, URL และสถานะการตรวจสอบ

```
#####
START Verifying BIB with marc tag 856 for HTTP Status.
Connecting to Sierra Database.
Checking for HTTP status.
Bib#: 420907795032
URL: http://atoz.ebsco.com/Home.asp?KW=economic&botany&sid=124209858&id=2688&SF=Titles&ST=Exact&cmdSearchSubmit=Search
Status: 200

Bib#: 420907795038
URL: http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2656
Status: ERROR: 403 Client Error: Forbidden for url: https://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2656

Bib#: 420907795116
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2531/SuthiCha/index.html
Status: 200

Bib#: 420907795117
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2528/AumnuaysartHas/index.html
Status: 200

Bib#: 420907795397
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2530/YaowalakCha/index.html
Status: 200

Bib#: 420907796087
URL: http://www.openbase.in.th/files/tbpj075.pdf
Status: 200

Bib#: 420907796311
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2538/ThanayaNua/index.html
Status: 200

Bib#: 420907796328
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2526/SaranPra/index.html
Status: 200

Bib#: 420907796381
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2526/KasimdaVir/index.html
Status: 200

Bib#: 420907796388
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2529/ThanasakAud/index.html
Status: 200

Bib#: 420907796397
URL: https://kasetsart.idm.oclc.org/login?url=https://www.lib.ku.ac.th/KUthesis/2525/PraneetSon/index.html
Status: 200
```

ภาพที่ 6 การตรวจสอบสถานะการเชื่อมโยงของ URL

BIBNO	Title	URL	STATUS
4.21E+11	The Journal of Animal Ecology	http://onlinelibrary.wiley.com/journal/10.1111/[ISSN]1365-2656	ERROR: 403 Client Error: Forbidden for url: https://onlinelibrary.wiley.com/journal/10.1111
4.21E+11	ผู้บัญชาการข่าวพุทธ :	http://www.openbase.in.th/files/tbpj075.pdf	ERROR: HTTPConnectionPool(host='www.openbase.in.th'
4.21E+11	วารสารวิชาการสงเสริมและนเทศศาสตร	http://www.aec.agr.ku.ac.th/ejournal.aspx	ERROR: 404 Client Error: Not Found for url: http://aec.agr.ku.ac.th/ejournal.aspx
4.21E+11	สรุปงานงานกาเารวังโครปจำปี /	http://203.157.15.110/boeng/annual.php	ERROR: HTTPConnectionPool(host='203.157.15.110'
4.21E+11	ขุนนางอยุธยา =	http://www.openbase.in.th/files/tbpj022.pdf	ERROR: HTTPConnectionPool(host='www.openbase.in.th'
4.21E+11	นิยสาร สลทพ. =	http://portal.ipt.ac.th/portal/dr/?x-p=ipst/d/service/magazine	ERROR: HTTPConnectionPool(host='portal.ipt.ac.th'
4.21E+11	Journal of Molecular Structure	http://www.sciencedirect.com/science/journal/00222860	ERROR: 403 Client Error: Forbidden for url: https://www.sciencedirect.com/science/journal
4.21E+11	การเงินธนาคาร =	http://www.karngern.com/	ERROR: HTTPConnectionPool(host='www.karngern.com'
4.21E+11	คอมพิวเตอร์ ทีวี =	http://www.lib.ku.ac.th/content_online/t_cmnr.pdf	ERROR: 404 Client Error: Not Found for url: https://www.lib.ku.ac.th/content_online/t_co
4.21E+11	คอมพิวเตอร์ อีเลทรอนนิคส์ เวิลด์ =	http://www.cewmedia.com	ERROR: HTTPConnectionPool(host='ww38.cewmedia.com'
4.21E+11	จุฬาลงกรณ์เศรฐกิจปริทัศน์ =	http://cbsreview.acc.chula.ac.th/Home.aspx	ERROR: HTTPConnectionPool(host='cbsreview.acc.chula.ac.th'
4.21E+11	จุฬาลงกรณ์วารสาร =	http://library.acc.chula.ac.th/Thai_Journal/ChulaReview/Index_ChulaRev	ERROR: HTTPConnectionPool(host='library.acc.chula.ac.th'
4.21E+11	เทียวรอบโลก =	http://www.tiewrobok.com	ERROR: 403 Client Error: Forbidden for url: http://www.tiewrobok.com/
4.21E+11	เนชั่นสุดสัปดาห์	http://www.nationweekend.com	ERROR: 404 Client Error: Not Found for url: https://www.bangkokbiznews.com/category/
4.21E+11	บ้านและสวน	http://www.baanlaesuan.com/index.aspx	ERROR: 404 Client Error: Not Found for url: https://www.baanlaesuan.com/index.aspx
4.21E+11	Buyer's News Engineering Products	http://www.lib.ku.ac.th/content_online/t_ben.pdf	ERROR: 404 Client Error: Not Found for url: https://www.lib.ku.ac.th/content_online/t_be
4.21E+11	ผู้คิดการ	http://www.gotomanager.com/magazine/default.aspx?menu=magazine	ERROR: 404 Client Error: Not Found for url: http://gotomanager.com/magazine/default.as
4.21E+11	แม่และเด็ก	http://www.motherandchildmagazine.com	ERROR: HTTPConnectionPool(host='www.motherandchildmagazine.com'
4.21E+11	รัฐสภาสาร	https://www.parliament.go.th/	ERROR: 403 Client Error: Forbidden for url: https://www.parliament.go.th/%20%20
4.21E+11	What Hi-Fi	http://www.lib.ku.ac.th/content_online/t_wfh.pdf	ERROR: 404 Client Error: Not Found for url: https://www.lib.ku.ac.th/content_online/t_wl
4.21E+11	วารสารศรัทธาปริทัศน์ ฉบับมนุษยศาส	http://sripatum-review.spu.ac.th/content_show.cfm?content_id=51	ERROR: HTTPConnectionPool(host='sripatum-review.spu.ac.th'
4.21E+11	วารสารเกษตร =	http://web.agri.cmu.ac.th/agjournal/journal_home.asp	ERROR: 404 Client Error: Not Found for url: https://web.agri.cmu.ac.th/agjournal/journal
4.21E+11	วารสารนักบัญชี =	http://www.lib.ku.ac.th/content_online/t_nj.pdf	ERROR: 404 Client Error: Not Found for url: https://www.lib.ku.ac.th/content_online/t_nj
4.21E+11	ว.ส.ล.	http://www.wfb-hq.org	ERROR: 403 Client Error: Forbidden for url: https://www.wfb-hq.org/
4.21E+11	วารสารพัฒนาบริหารศาสตร =	http://journal.nida.ac.th/journal/index.php?option=com_content&view=	ERROR: HTTPConnectionPool(host='journal.nida.ac.th'
4.21E+11	วารสารพัฒนาชุมชน	http://www.lib.ku.ac.th/content_online/t_pnp.pdf	ERROR: 404 Client Error: Not Found for url: https://www.lib.ku.ac.th/content_online/t_pnp
4.21E+11	วารสาร มกค.	http://www.utcc.ac.th/thesis/index.php	ERROR: 404 Client Error: Not Found for url: https://www.utcc.ac.th/thesis/index.php
4.21E+11	วารสารราชบัณฑิตยสถาน =	http://www.royin.go.th/2ebok-category=journal-of-the-royal-institute	ERROR: HTTPConnectionPool(host='www.royin.go.th'
4.21E+11	วารสารวิจัยสภากาชาดสยาม =	http://www.eric.chula.ac.th/journal/fer/	ERROR: HTTPConnectionPool(host='www.eric.chula.ac.th'

ภาพที่ 7 รายงานแสดงผลลัพธ์การตรวจสอบสถานะ


```

varfield_id      BIBNO      URL content      HTTP STATUS
0      11002048      420908223005      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPSConnectionPool(host='idp.springer...
1      11002049      420908223006      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPSConnectionPool(host='link.springer...
2      11002128      420908223015      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPSConnectionPool(host='idp.springer...
3      11002706      420908223027      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPSConnectionPool(host='link.springer...
4      11002534      420908223028      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPConnectionPool(host='link.springer...
80      11036012      420908222433      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPSConnectionPool(host='idp.springer...
81      11036087      420908224434      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPSConnectionPool(host='link.springer...
82      11036363      420908224444      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPConnectionPool(host='link.springer...
83      11036390      420908224445      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPConnectionPool(host='link.springer...
84      11036388      420908224446      ...      KU Network connect to SpringerLink      (0000000000)      ERROR: HTTPConnectionPool(host='link.springer...

[85 rows x 11 columns]
varfield_id: 11002048
New URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5922600/
Keyword: "Gastrointestinal oncology"

"Gastrointestinal oncology"
varfield_id: 11002049
New URL: https://link.springer.com/book/10.1007/978-3-642-13312-1
Keyword: "Classification and multivariate analysis for complex data structures"

"Classification and multivariate analysis for complex data structures"
varfield_id: 11002128
New URL: https://link.springer.com/book/10.1007/978-3-642-13535-4
Keyword: "Pediatric airway surgery"

"Pediatric airway surgery"
varfield_id: 11002706
New URL: https://th.hnote.asia/orgdevelopment/190702-design-thinking/
Keyword: "Design thinking"

"Design thinking"
varfield_id: 11002534
New URL: https://www.cancer.org/cancer/types/leukemia-in-children.html
Keyword: "Childhood leukemia"

"Childhood leukemia"
varfield_id: 11002530
New URL: https://www.google.com/preferences%3Fhl%3Dth
Keyword: "Raising of microvascular flaps"

"Raising of microvascular flaps"
varfield_id: 11002645
New URL: https://link.springer.com/book/10.1007/978-3-642-13832-4
Keyword: "Decision making in radiation oncology"

```

BIBNO	URL	NEW_URL
4.20908E+11	http://dx.doi.org/10.1007/978-0-387-85952-1	https://link.springer.com/book/10.1007/978-0-387-85952-1
4.20908E+11	http://dx.doi.org/10.1007/978-0-387-87363-3	https://link.springer.com/book/10.1007/978-0-387-87363-3
4.20908E+11	http://dx.doi.org/10.1007/978-0-387-89518-5	https://psycnet.apa.org/record/2006-21200-000
4.20908E+11	http://dx.doi.org/10.1007/978-0-387-92168-6	https://link.springer.com/book/10.1007/978-3-319-75708-7
4.20908E+11	http://dx.doi.org/10.1007/978-1-4020-8286-3	https://link.springer.com/referencework/10.1007/978-1-4020-8286-3
4.20908E+11	http://dx.doi.org/10.1007/978-1-4020-9210-7	https://link.springer.com/book/10.1007/978-1-4020-9210-7
4.20908E+11	http://dx.doi.org/10.1007/978-1-4020-9362-3	https://link.springer.com/book/10.1007/978-1-4020-9362-3
4.20908E+11	http://dx.doi.org/10.1007/978-1-4020-9389-0	https://link.springer.com/book/10.1007/978-1-4020-9389-0
4.20908E+11	http://dx.doi.org/10.1007/978-0-387-09548-6	https://www.car.chula.ac.th/display7.php?bib=b2027219
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2745	https://www.cambridge.org/core/journals/journal-of-tropical-ecology
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1529-8817	https://onlinelibrary.wiley.com/journal/15298817
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1098-2736	https://onlinelibrary.wiley.com/journal/10982736
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1745-459X	https://onlinelibrary.wiley.com/journal/1745459X
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1748-5827	https://onlinelibrary.wiley.com/journal/17485827
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-9841	https://onlinelibrary.wiley.com/journal/14679841
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1745-4603	https://onlinelibrary.wiley.com/journal/17454603
4.20908E+11	http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2885	https://onlinelibrary.wiley.com/journal/13652885
4.20908E+11	http://www.sciencedirect.com/science/book/9780123741813	https://www.sciencedirect.com/book/9780128018132/wine-tasting
4.20908E+11	http://www.sciencedirect.com/science/book/9780123741448	https://www.sciencedirect.com/book/9780123741448/encyclopedia
4.20908E+11	http://www.sciencedirect.com/science/book/9780123746283	https://www.sciencedirect.com/book/9780123746283/bioactive-food
4.20908E+11	http://www.sciencedirect.com/science/book/9780123742278	https://www.sciencedirect.com/book/9780123742278/single-molecule

PULINET Journal Vol. 11, No. 1, January-June 2024: pp.68-80
<https://so14.tci-thaijo.org/index.php/PJ>
 Published by Provincial University Library Network, THAILAND

จากการตรวจสอบรายการบรรณานุกรมทั้งหมด 553,904 ระเบียน (ข้อมูล ณ เดือนกรกฎาคม 2566) พบว่า มีรายการบรรณานุกรมที่มีข้อมูล Tag 856 (Electronic Location and Access) ซึ่งเป็นข้อมูล URL ทั้งสิ้น 121,149 รายการ การตรวจสอบรูปแบบของ URL พบรูปแบบที่ไม่ถูกต้องทั้งสิ้น 103 รายการ แบ่งเป็นลงรายการบรรณานุกรมผิด Tag จำนวน 65 รายการ และใส่รูปแบบ URL ไม่ถูกต้อง เช่น พิมพ์ผิด (http://, +http://) พิมพ์ตกหล่น (tp://) ใส่ข้อมูล Subfield ลงไป (uhttp://) จำนวน 32 รายการ

การตรวจสอบสถานะการเชื่อมโยงของ URL พบสถานะการเชื่อมโยงดังนี้ สถานะเชื่อมโยงสำเร็จ (Status 2xx) จำนวน 107,221 รายการ สถานะผิดพลาดที่ฝั่ง Client (Status 4xx) จำนวน 12,057 รายการ สถานะถูกปฏิเสธจาก Server เพราะมีการส่ง Request เกินที่ Server กำหนด (Max Retries) จำนวน 1,830 รายการ สถานะผิดพลาดที่ฝั่ง Server (Status 5xx) จำนวน 22 รายการ สถานะการเชื่อมต่อมีปัญหา (Connection Reset Error 10054) จำนวน 14 รายการ สถานะไม่พบ Protocol (File Transfer Protocol) จำนวน 5 รายการ (ตารางที่ 1) โดยสามารถจำแนกตาม Top-level Domain ได้ดังนี้ .com 7,999 รายการ .th 2520 รายการ .org 1554 รายการ .gov 1493 รายการ .us 64 รายการ .uk 48 รายการ .au 19 รายการ .ca 18 รายการ .edu 18 รายการ .jp 16 รายการ อื่น ๆ 170 รายการ การปรับปรุงแหล่งการเชื่อมโยงด้วย Azure AI Service ได้ทำการปรับปรุงแหล่งการเชื่อมโยงของ URL ทั้งสิ้น 13,928 รายการ (ตารางที่ 2) และได้ทำการเลือกกลุ่มตัวอย่างเพื่อตรวจสอบความถูกต้องของแหล่งการเชื่อมโยงกับข้อมูลบรรณานุกรม โดยใช้สูตรการคำนวณขนาดของกลุ่มตัวอย่างตามหลักการของทาโรยามาเน่ (Taro, 1973) ที่ระดับความน่าเชื่อถือ ร้อยละ 97.5 ความคลาดเคลื่อนที่ร้อยละ ± 2.5 ดังนี้

สูตรคำนวณ

$$n = \frac{N}{1 + Ne^2}$$

แทนค่า

$$n = \frac{13,928}{1 + (13,928 \times 0.25^2)}$$

$$n = 1,392 \text{ รายการ}$$

จากการคำนวณข้างต้น ได้กลุ่มตัวอย่าง เท่ากับ 1,392 รายการ และได้ดำเนินการสุ่มตามหลักการสุ่มอย่างง่าย (Simple Random Sampling) จากการตรวจสอบด้วยเจ้าหน้าที่ โดยการนำข้อมูลบรรณานุกรมไปสืบค้นเพื่อตรวจสอบแหล่งการเชื่อมโยง พบว่าแหล่งการเชื่อมโยงมีความถูกต้องตรงกับข้อมูลบรรณานุกรม จำนวน 1,239 รายการ และแหล่งการเชื่อมโยงที่ไม่ถูกต้องกับข้อมูลบรรณานุกรม จำนวน 153 รายการ คิดเป็นความถูกต้องเท่ากับร้อยละ 88.49 แสดงให้เห็นว่าการปรับปรุงแหล่งการเชื่อมโยงด้วย Azure AI Service มีความถูกต้องแม่นยำเป็นที่ยอมรับและสามารถนำมาใช้งานได้ และหลังจากทำการปรับปรุงแหล่งการเชื่อมโยงของ URL แล้ว ไม่พบข้อร้องเรียนจากการแจ้งการเชื่อมโยงข้อมูลผิดพลาด นับตั้งแต่เดือนสิงหาคม-กันยายน พ.ศ. 2566

ตารางที่ 1 จำนวนร้อยละของจำนวนรายการการเชื่อมโยงของ URL แยกตามประเภทสถานะการเชื่อมโยง

ประเภทสถานะการเชื่อมโยง	จำนวนรายการ	
	รายการ	ร้อยละ
เชื่อมโยงสำเร็จ (Status 2xx)	107,221	88.5
ผิดพลาดที่ฝั่ง Client (Status 4xx)	12,057	9.95
ถูกปฏิเสธจาก Server เพราะมีการส่ง Request เกินที่ Server กำหนด (Max Retries)	1,830	1.51
ผิดพลาดที่ฝั่ง Server (Status 5xx)	22	0.02
สถานะการเชื่อมต่อมีปัญหา (Connection Reset Error 10054)	14	0.01
ไม่พบ Protocol (File Transfer Protocol)	5	0.004
รวม	121,149	100

ตารางที่ 2 จำนวนร้อยละของจำนวนรายการการเชื่อมโยงของ URL ที่ไม่สำเร็จ แยกตามประเภท Top-level Domain

ประเภท Top-level Domain	จำนวนรายการ	
	รายการ	ร้อยละ
.com	7,999	57.43
.th	2,520	18.09
.org	1,554	11.15
.gov	1,493	10.71
.us	64	0.45
.uk	48	0.34
.au	19	0.13
.ca	18	0.12
.edu	18	0.12
.jp	16	0.11
อื่นๆ	170	1.22
รวม	13,928	100

บทสรุป (Conclusion)

จากการใช้งานระบบตรวจสอบความถูกต้อง ความพร้อมใช้งานและปรับปรุงการเชื่อมโยง URL ด้วย Azure AI Service พบว่า ระบบช่วยลดขั้นตอนกระบวนการและระยะเวลาในการปฏิบัติงานในส่วนของ 1) การตรวจสอบความถูกต้องของรูปแบบการเชื่อมโยง ซึ่งเกิดจากความผิดพลาดของผู้กรอกข้อมูล 2) การตรวจสอบสถานะของการเชื่อมโยงและปรับปรุงแหล่งการเชื่อมโยงใหม่ ซึ่งจากเดิมต้องทำโดยการเปิดดูแหล่งเชื่อมโยงในแต่ละรายการว่าสามารถเข้าใช้งานได้หรือไม่ ถ้าใช้งานไม่ได้ ต้องทำการสืบค้นข้อมูลใหม่ แล้วนำมาปรับปรุงข้อมูลในระบบห้องสมุดอัตโนมัติ โดยหลังจากนำระบบมาใช้งานนั้น ได้มีการปรับปรุงกระบวนการในการตรวจสอบสถานะและปรับปรุงแหล่งการเชื่อมโยงใหม่ด้วย Azure AI Service ระบบจะทำ

การตรวจสอบสถานะให้อย่างอัตโนมัติ ถ้าสถานะของแหล่งเชื่อมโยงไม่สามารถใช้งานได้ ระบบจะดำเนินการสืบค้นข้อมูล โดยนำคำค้นจากรายการบรรณานุกรมนั้น เช่น ชื่อเรื่อง ISBN ISSN ปีพิมพ์ สำนักพิมพ์ เป็นต้น มาสืบค้นและคัดเลือกแหล่งเชื่อมโยงที่สอดคล้องกับคำค้น และนำแหล่งเชื่อมโยงไปปรับปรุงข้อมูลในระบบห้องสมุดอัตโนมัติทันที หลังจากนั้น ระบบจะส่งผลการตรวจสอบทั้งหมดเป็นรายงานให้ผู้รับผิดชอบได้ทราบ โดยกระบวนการทั้งหมดใช้เวลาเฉลี่ยที่ 1-2 วินาที ต่อ 1 รายการ และเป็นการเพิ่มประสิทธิภาพในการใช้งานแก่ผู้ใช้บริการ ผู้ใช้บริการสามารถเข้าถึงแหล่งเชื่อมโยงทรัพยากรสารสนเทศที่ถูกต้องครบถ้วนสมบูรณ์

ทั้งนี้ ระบบสามารถนำไปประยุกต์ใช้กับการทำงานหรือระบบอื่น ๆ โดยไม่จำกัดแค่ระบบห้องสมุดอัตโนมัติเพียงอย่างเดียว เช่น ระบบ Digital Repositories ระบบ Journal System รวมถึงแหล่งเชื่อมโยงของเว็บไซต์ที่เชื่อมกับหน้าเว็บต่าง ๆ และไม่จำเป็นต้องจัดสรรงบประมาณเพื่อการจัดซื้อหรือจ้างพัฒนาระบบ

การพัฒนาระบบได้พบปัญหาและอุปสรรค ดังนี้

1. ระบบจำเป็นต้องเชื่อมต่อเครือข่ายอินเทอร์เน็ตทุกครั้งเมื่อมีการทำงาน ถ้าเครือข่ายที่ใช้งานอยู่มีปัญหาหรือเกิดความไม่เสถียร ระบบจะไม่สามารถทำงานได้
2. การตรวจสอบสถานะการเชื่อมโยง เมื่อมีการส่งคำขอ (Request) ไปยังผู้ให้บริการเว็บไซต์ บางเว็บไซต์จะจำกัดการร้องขอไว้ ทำให้ต้องขยายระยะเวลาในการส่ง Request ในแต่ละรายการออกไป ทำให้ระบบต้องใช้ระยะเวลาในการทำงานมากขึ้น
3. การใช้ Azure AI Service จำกัดปริมาณการร้องขอ (Request) อยู่ที่ 1,000 Transactions ต่อเดือน สำหรับการใช้งานแบบ Free Tier ซึ่งถ้าต้องการใช้งานปริมาณมากกว่านี้ จะต้องทำการบอกรับในรูปแบบ Subscription โดยจะมีค่าใช้จ่าย อยู่ที่ \$15-\$25 ต่อ 1,000 Request (ข้อมูลราคาเดือน กรกฎาคม พ.ศ. 2566 จาก <https://www.microsoft.com/en-us/bing/apis/pricing>) โดยได้ทำการแบ่งการตรวจสอบเดือนละ 1,000 Transactions จนครบจำนวนรายการบรรณานุกรมทั้งหมด
4. ในการทดสอบครั้งแรก มีจำนวนแหล่งการเชื่อมโยงที่เชื่อมโยงไม่สำเร็จเป็นจำนวนมาก ทำให้การปรับปรุงแหล่งการเชื่อมโยงด้วย Azure AI Service ไม่สามารถดำเนินการให้เสร็จภายในครั้งเดียว เนื่องจากติดปัญหาด้านการจำกัดปริมาณการร้องขอ (Request) อยู่ที่ 1000 Transactions ต่อเดือน (ข้อ 3) โดยหลังจากทำการตรวจสอบแหล่งที่เชื่อมโยงไม่สำเร็จครั้งแรกครบทุกรายการบรรณานุกรมแล้ว ได้ทำการตั้งให้ระบบทำงานทุกเดือน โดยจำนวนแหล่งที่เชื่อมโยงไม่สำเร็จได้มีจำนวนลดลงจากการทดสอบครั้งแรก ทำให้ระบบสามารถทำงานได้โดยไม่ถูกจำกัดปริมาณการร้องขอที่ 1,000 Transactions ต่อเดือน
5. ข้อมูลบางรายการ มีรายละเอียดทางบรรณานุกรมไม่ถูกต้องครบถ้วน เช่น ชื่อเรื่อง ISBN ISSN สำนักพิมพ์ เป็นต้น ซึ่งเป็นผลจากการที่ผู้ปฏิบัติงานทำการลงรายการบรรณานุกรมไม่ถูกต้องหรือครบถ้วน ทำให้การสืบค้นข้อมูลโดยใช้ Azure AI Service ได้ผลลัพธ์ของแหล่งเชื่อมโยงที่ไม่สอดคล้องกันกับข้อมูลบรรณานุกรม

การปรับปรุงและพัฒนาระบบ

1. การตรวจสอบสถานะการเชื่อมโยง ซึ่งได้กำหนดให้ตรวจสอบเฉพาะ HTTP Status ในหมวด 4xx และ 5xx เท่านั้น ควรกำหนดขอบเขตการตรวจสอบที่ละเอียดมากยิ่งขึ้น เพราะบางสถานะในหมวด ไม่ได้แสดงถึงสถานะที่เชื่อมโยงไม่สำเร็จ เช่น 403 (ไม่ได้รับอนุญาตให้เข้าถึง) 429 (มีการร้องขอมากเกินไป) 503 (ไม่สามารถดำเนินการได้ชั่วคราว) 504 (ไม่ได้รับการตอบสนองจาก Gateway) เป็นต้น เพื่อผลลัพธ์ในการตรวจสอบสถานะมีความถูกต้องและแม่นยำยิ่งขึ้น
2. เพิ่มขั้นตอนการตรวจสอบสถานะการเข้าถึงเว็บไซต์หรือข้อมูลภายในเว็บไซต์ ในกรณีที่เว็บไซต์นั้นต้องมีการยืนยันตัวตน (Authorization) ผ่านการใช้ชื่อผู้ใช้และรหัสผ่าน หรือหมายเลข IP Address เป็นต้น เพื่อการอนุญาตให้เข้าถึงข้อมูล

หรือเพิ่มการตรวจสอบเนื้อหาหรือรายละเอียดของเว็บไซต์ เพื่อให้ทราบสถานะของข้อมูลหรือรายละเอียดที่อยู่ภายในเว็บไซต์ได้ถูกต้องยิ่งขึ้น

3. การสืบค้นด้วย Azure AI Service เป็นการสืบค้นด้วยวิธี Simple Query ซึ่งเป็นการค้นหาแบบง่าย ให้ผลลัพธ์ที่ตรงไปตรงมา ใช้รูปแบบการพัฒนาระบบที่ไม่ซับซ้อน เหมาะสำหรับการค้นหาทั่ว ๆ ไป ซึ่งควรปรับปรุงวิธีการสืบค้นที่มีรูปแบบซับซ้อน เช่น Fuzzy Search, Proximity Search, Field Search เป็นต้น เหมาะสำหรับการสืบค้น ที่ต้องการผลลัพธ์ที่ถูกต้อง แม่นยำ มีประสิทธิภาพมากยิ่งขึ้น แต่ต้องใช้ระยะเวลาในการทำงานนานขึ้น

4. เพิ่มขั้นตอนการตรวจสอบสถานะและปรับปรุงแหล่งการเชื่อมโยงใหม่โดยใช้ Azure OpenAI ร่วมกับ โมเดล ChatGPT โดยทำการฝึกฝน AI ด้วยข้อมูลภายในของสำนักหอสมุด (Bring Your Own Data) เช่น ข้อมูลการบอกรับฐานข้อมูล ข้อมูลวารสารออนไลน์ เป็นต้น เพื่อให้ได้ผลลัพธ์ที่ถูกต้องมากยิ่งขึ้น

รายการอ้างอิง (References)

- B.T. Sampath Kumara, D. Vinay Kumarb. (2012). HTTP 404-page (not) found: Recovery of decayed URL citations. *Journal of Informetrics*, 7(1), 145-157
- Iterative Inc. (2023). *PyDrive2's documentation*. <https://docs.iterative.ai/PyDrive2>
- Kenneth Reitz. (2023). *Requests III: HTTP for Humans and Machines, alike*. <https://3.python-requests.org>
- Microsoft. (2023). *Bing Search API documentation*. <https://learn.microsoft.com/en-us/azure/cognitive-services/bing-web-search>
- Mohamed Sami (2012). *Software Development Life Cycle Models and Methodologies*. <https://melsatar.blog/2012/03/15/software-development-life-cycle-models-and-methodologies/>
- กอบเกียรติ สระอุบล. (2563). *เรียนรู้ Data Science และ AI : Machine Learning ด้วย Python*. มีเดีย เนทเวิร์ค.