

รัฐสภาไทยนำ Google Search Appliance ช่วยยกระดับประสิทธิภาพการค้นหาในห้องสมุดดิจิทัล



ภาพรวม

เป้าหมายของโครงการ

- พัฒนาระบบการค้นหาข้อมูลให้ถูกต้องมากขึ้น เพื่อให้ผู้ใช้ค้นหาเอกสารนิติบัญญัติที่ต้องการได้ง่าย
- รวมผลการค้นหาจากแหล่งข้อมูลต่างๆ หลากหลายรูปแบบ ให้ผลลัพธ์ที่เป็นหนึ่งเดียว และถูกต้อง
- ลดการทำงานซ้ำซ้อนและผลลัพธ์ที่ซ้อนกัน ระหว่างการค้นหาข้อมูล

วิธีการดำเนินการ

- นำ Google Search Appliance (GSA) มาใช้เพื่อค้นหาข้อมูลจากสื่อชนิดต่างๆ ได้อย่างรวดเร็ว แก้ไขปัญหาความยุ่งยากที่เกิดจากอักขระภาษาไทย

ผลลัพธ์

- เพิ่มความเร็วกระบวนการค้นหาข้อมูลได้ถึง 10 เท่า
- เพิ่มปริมาณการใช้งานห้องสมุดดิจิทัลได้ถึง 200%
- ยกระดับการแลกเปลี่ยนองค์ความรู้ในแวดวงวิชาการ เนื่องจากจำนวนผู้ใช้รายใหม่ๆ จากสถานศึกษาและแวดวงกฎหมายที่เข้าใช้ระบบนี้ เพื่อการเรียนรู้และการทำงาน

“ผมเชื่อมั่นว่าในตลาดขณะนี้ไม่มีโซลูชันอื่นใดเลยที่เทียบได้กับข้อเสนอของ Google ในเรื่องของความสามารถในการใช้งานและความต่อเนื่องของรอบอายุผลิตภัณฑ์”

—ดร. อติศักดิ์ สุกุล จากสาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภูมิหลัง

รัฐสภาไทยมีการจัดเก็บเอกสารทางราชการไว้เป็นจำนวนมหาศาลกว่า 350,000 ฉบับ ตั้งแต่ พ.ศ. 2475 หรือนับตั้งแต่รัฐธรรมนูญไทยถือกำเนิดขึ้น

ด้วยจำนวนและขนาดของเอกสารที่เพิ่มขึ้นอย่างมากในแต่ละปี รัฐสภาไทยจึงได้มีการแสวงหาระบบการค้นหาที่น่าเชื่อถือและสามารถรองรับข้อมูลมหาศาล เพื่อจัดการกับการค้นหาคลังข้อมูลขนาดใหญ่ นี่คือการเริ่มต้นของโครงการสร้างห้องสมุดดิจิทัลสำหรับข้อมูลนิติบัญญัติที่ใหญ่ที่สุดในประเทศ ประชาชนทั่วไปสามารถเข้าถึงคลังข้อมูลนี้ทางเว็บไซต์ <http://dl.parliament.go.th/> โครงการนี้มีวัตถุประสงค์ที่จะแปลงเอกสารสนเทศทั้งหมดให้เป็นข้อมูลดิจิทัลซึ่งมีการจัดเก็บในแพลตฟอร์มเดียว โดยเปิดให้ภาครัฐ ภาควิชาการ และบุคคลทั่วไปเข้าใช้ข้อมูลได้อย่างอิสระ นอกจากนี้ยังเป็นการอนุรักษ์เอกสารสำคัญในระยะยาว พร้อมทั้งลดพื้นที่จัดเก็บเอกสารในรูปแบบสิ่งพิมพ์อีกด้วย

ความท้าทาย

ห้องสมุดรัฐสภาไทยเป็นที่ยึดเก็บข้อมูลนิติบัญญัติที่มีคุณค่าไว้เป็นจำนวนมาก โดยมีเอกสารนิติบัญญัติหลากหลายประเภท ในรูปแบบต่างๆ ไม่ว่าจะเป็น เอกสาร วารสาร หนังสือพิมพ์ หนังสืออิเล็กทรอนิกส์ และวิดีโอ ในการจัดเก็บเอกสารนิติบัญญัติแต่ละประเภท มีการใช้ข้อมูลเมตา Dublin Core ซึ่งเป็นชุดคำศัพท์ตามมาตรฐานอุตสาหกรรมในการกำหนดมาตรฐานและทำให้ประสิทธิภาพในการค้นหาดียิ่งขึ้น

แต่เนื่องจากความซับซ้อนและความหลากหลายของโครงสร้างของเอกสารนิติบัญญัติ การดึงข้อมูลผลลัพธ์ให้ออกมาแม่นยำจึงเป็นเรื่องที่ยากลำบาก ถึงแม้จะมีการสร้างระบบเพิ่มประสิทธิภาพของเครื่องมือค้นหา โดยใช้ซอฟต์แวร์คลังข้อมูลสารสนเทศของรัฐสภาไทยอยู่แล้วก็ตาม ที่เป็นเช่นนั้นก็เนื่องจากการอ้างอิงเอกสารจำนวนมากที่มีความเกี่ยวพันกันสูง ทำให้กระบวนการค้นหาทำได้ยากลำบากขึ้น

ดร. อติศักดิ์ สุกุล จากสาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ซึ่งเป็นหนึ่งในคณะทำงานที่เป็นที่ปรึกษาด้านการพัฒนาคลังข้อมูลดิจิทัล อธิบายว่า “เนื่องจากความแตกต่างของโครงสร้างข้อมูลเมตาสำหรับเอกสารแต่ละประเภท จึงเป็นเรื่องยากมากที่จะใช้ระบบค้นหาเพียงระบบเดียวเพื่อสร้างผลการค้นหาที่แม่นยำ อัลกอริทึม Enterprise PageRank ซึ่งเป็นส่วนหนึ่งของ Google Search Appliance มีประโยชน์มากช่วยให้เราได้ผลลัพธ์ที่แม่นยำ และอีกหนึ่งเหตุผลก็คือความซับซ้อนของการรวมอักขระภาษาไทยในการประกอบคำ ข้อด้อยประการหนึ่งของเครื่องมือค้นหาก่อนหน้านี้นี้คือไม่สามารถรับรู้อักขระไทยได้อย่างถูกต้อง ทำให้ผลการค้นหาไม่แม่นยำเท่าที่ควร หรือบางครั้งก็ไม่ได้ผลลัพธ์มาเลยด้วยซ้ำ”

ทางออก

รัฐสภาไทยต้องการระบบคลังสารสนเทศที่สามารถปรับแต่งได้ และสามารถให้ประสิทธิภาพในการค้นหาที่เรียบง่ายแต่มีประสิทธิภาพ โซลูชันนี้จะต้องทำงานร่วมกับโมดูลในส่วนแบ็คเอนด์ได้ดี นั่นคือซอฟต์แวร์คลังข้อมูล DSpace ซึ่งเป็นระบบโอเพนซอร์สขององค์กรที่ได้ได้รับความนิยมในห้องสมุดทางวิชาการของประเทศไทย แม้ว่าซอฟต์แวร์คลังข้อมูล DSpace จะสามารถจัดเก็บเอกสารทางวิชาการไว้ในแพลตฟอร์มเดียว แต่ข้อด้อยประการหนึ่งของระบบนี้ก็คือฟังก์ชันการค้นหาสำหรับอักขระภาษาไทยที่ค่อนข้างจำกัด

“เราเลือก GSA เป็นส่วนหน้าของเครื่องมือค้นหาข้อมูลในห้องสมุดดิจิทัลเนื่องจาก Google นั้นมีชื่อเสียงในตลาดเครื่องมือค้นหาอยู่แล้ว” อาจารย์อติศักดิ์กล่าว “ผมเชื่อมั่นว่าในตลาดขณะนี้ไม่มีโซลูชันอื่นใดเลยที่เทียบได้กับข้อเสนอของ Google ในเรื่องของความสามารถในการใช้งานและความต่อเนื่องของรอบอายุผลิตภัณฑ์”

เกี่ยวกับ Google Search Appliance

Google.com นำโลกแห่งสารสนเทศอันกว้างใหญ่ไพศาลมาอยู่ในมือคุณ Google Search Appliance ก็มีคุณสมบัติเช่นเดียวกันในการจัดการข้อมูลในองค์กรของคุณ Google Search Appliance เป็นผลิตภัณฑ์ฮาร์ดแวร์และซอฟต์แวร์ที่ผสานรวมเป็นหนึ่งเดียว ได้รับการออกแบบมาเพื่อให้องค์กรได้ใช้พลังการค้นหาโดย Google ช่วยเพิ่มประสิทธิภาพการทำงาน นี่คือโซลูชันการค้นหาขององค์กรที่ให้ความเรียบง่าย มีประสิทธิภาพ และครอบคลุมครบวงจรเหมือนกับ Google เอง เวอร์ชันล่าสุดประกอบด้วย การประเมินความเกี่ยวข้องที่ดีขึ้น การเข้าถึงเนื้อหาได้มากขึ้น และมีตัวเลือกการรักษาความปลอดภัยที่ดีขึ้น

ท่านสามารถอ่านข้อมูลเพิ่มเติมได้ที่
www.google.com/enterprise/search

“การทำงานกับภาษาไทยในเรื่องการเพิ่มประสิทธิภาพเครื่องมือค้นหานั้นเป็นเรื่องที่ทำหายนามาก แต่ทีมงานด้านการค้นหาที่ทุ่มเทและมีความสามารถของ Google ก็ทำงานอย่างหนักเพื่อให้เราก้าวข้ามอุปสรรคไปได้”

—ดร. อติศักดิ์ สุกุล จากสาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

หลังจากเปลี่ยนมาใช้ GSA ในปี 2554 การค้นหาข้อมูลสำหรับผู้ใช้งานสมุดดิจิทัลมีผลดีขึ้นอย่างมาก การรวมเนื้อหาไว้ในแพลตฟอร์มเดียวที่ขับเคลื่อนโดยเครื่องมือค้นหาของ Google ช่วยให้การค้นหาเร็วขึ้นถึง 10 เท่า และ GSA ก็มีส่วนติดต่อที่ใช้งานง่าย ทำให้ผู้ใช้รายใหม่ๆ สนใจเข้ามาใช้ห้องสมุดดิจิทัลมากขึ้นเรื่อยๆ ตั้งแต่ที่เราเปิดให้บุคคลทั่วไปสามารถเข้ามาใช้ระบบออนไลน์ได้

“ในอดีตระบบนี้จะอนุญาตให้ใช้ได้เฉพาะสมาชิกของรัฐสภาเท่านั้น แต่เนื่องจากเราได้รับคำร้องขอจากบุคคลทั่วไปในการเข้าถึงข้อมูลนิติบัญญัติเป็นจำนวนเพิ่มมากขึ้น ทางรัฐสภาไทยจึงได้ตัดสินใจเปิดให้บุคคลทั่วไปสามารถเข้าถึงเอกสารที่ไม่เป็นความลับได้ทางโดเมนสาธารณะ ขณะนี้มีการใช้งานเพิ่มขึ้นมากจากผู้ใช้งานใหม่ๆ เช่น นักกฎหมาย นักศึกษา และยังได้รับความนิยมอย่างมากจากองค์กรพัฒนาเอกชน” อาจารย์อติศักดิ์กล่าว

GSA ยังสามารถก้าวข้ามอุปสรรคของภาษาไทย ด้วยการใช้พจนานุกรมที่ติดตั้งมากระบบค้นหาและฟังก์ชันการเรียนรู้วิธีที่สามารถอ่านและแยกคำได้อย่างชาญฉลาด

ผลลัพธ์

การนำ GSA มาใช้งานทำให้กระบวนการค้นหามีประสิทธิภาพยิ่งขึ้น เนื่องจากสามารถรวบรวมข้อมูลจากหลายๆ แพลตฟอร์มได้มากขึ้น ทำให้ห้องสมุดดิจิทัลนำเข้ามาใช้สำหรับผู้ใช้งาน และจนถึงขณะนี้ ห้องสมุดดิจิทัลมีปริมาณการใช้งานเพิ่มขึ้นถึง 200%

“และเนื่องจากตอนนี้เราใช้แค่แพลตฟอร์มเดียว ดังนั้นการค้นหาจะยิ่งดีกว่าเดิมอีกถึงสิบเท่า” อาจารย์อติศักดิ์กล่าว “นอกจากนี้เรายังเห็นว่าการค้นหาที่ทำใน Google.com หรือเครื่องมือค้นหาอื่นๆ จะมีการเชื่อมโยงไปที่เอกสารในห้องสมุดดิจิทัลโดยตรง ซึ่งหมายความว่าขณะนี้เราสามารถเข้าถึงคนมากขึ้น และมีผู้ใช้ที่ได้รับประโยชน์จากคลังข้อมูลในห้องสมุดดิจิทัลของเราในวงกว้างมากขึ้น”

ฟังก์ชันป้อนข้อมูลอัตโนมัติและการตรวจตัวสะกดของ GSA มีประโยชน์สำหรับผู้ใช้งานซึ่งจะช่วยให้ได้ผลการค้นหาที่แม่นยำยิ่งขึ้นในเวลาที่ใช้รวดเร็วขึ้น “การป้อนข้อมูลอัตโนมัติและการตรวจสอบการสะกดนั้นมีประโยชน์มากจริงๆ สำหรับโครงสร้างภาษาของไทย มีการสะกดผิดกันบ่อยๆ แต่ GSA เข้าใจคำไทยที่ใช้กันบ่อยๆ และสามารถให้คำแนะนำที่ถูกต้องได้ ทำให้การค้นหาง่ายและเร็วมาก” อาจารย์อติศักดิ์กล่าว

ความท้าทายหลากหลายอย่าง เนื่องมาจากขนาดของโครงการ ข้อมูลเมตาที่ซับซ้อน และลักษณะของภาษาไทยที่ละเอียด ทำให้รัฐสภาไทยและทีมงานของ Google ต้องทำงานร่วมกันอย่างใกล้ชิด

“การทำงานกับภาษาไทยในเรื่องการเพิ่มประสิทธิภาพเครื่องมือค้นหานั้นเป็นเรื่องที่ทำหายนามาก แต่ทีมงานด้านการค้นหาที่ทุ่มเทและมีความสามารถของ Google ก็ทำงานอย่างหนักเพื่อให้เราก้าวข้ามอุปสรรคไปได้ ผมพอใจมากกับความสำเร็จของโครงการห้องสมุดนี้ ซึ่ง Google มีส่วนสำคัญอย่างมากทีเดียว” อาจารย์อติศักดิ์กล่าว

