

ณัฐกานต์ อัมรินทร์ารักษ์. 2552. การจับคู่อิลิเมนต์ระหว่างโครงสร้างเอกซ์เอ็มแอลแบบกึ่งอัตโนมัติ.

วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น.

อาจารย์ที่ปรึกษาวิทยานิพนธ์: ผศ. ดร. กานดา สายแก้ว

บทคัดย่อ

ปัจจุบันภาษาเอกซ์เอ็มแอลถูกใช้เป็นตัวกลางในการแลกเปลี่ยนข้อมูลระหว่างกันอย่างแพร่หลาย เพราะคุณสมบัติของโครงสร้างภาษาที่มีความยืดหยุ่น ทำความเข้าใจได้ง่าย และไม่ขึ้นกับแพลตฟอร์มใดแพลตฟอร์มหนึ่ง แต่ปัญหาที่สำคัญอย่างหนึ่งเมื่อต้องการแลกเปลี่ยนข้อมูลหรือใช้งานข้อมูลร่วมกันระหว่างข้อมูลเอกซ์เอ็มแอลที่ถูกสร้างจากแต่ละแหล่งนั้นคือ ผู้ปฏิบัติงานต้องจับคู่อิลิเมนต์ที่มีความสัมพันธ์กันระหว่างเอกสารเอกซ์เอ็มแอลที่ต้องการ ซึ่งกระบวนการจับคู่โดยคนนั้นที่ต้องใช้เวลาและมักจะสร้างความสับสนสำหรับผู้ปฏิบัติงานเป็นอย่างมากโดยเฉพาะอย่างยิ่งกับเอกซ์เอ็มแอลที่มีโครงสร้างข้อมูลขนาดใหญ่เนื่องจากผู้ปฏิบัติงานต้องเข้าใจรายละเอียดของโครงสร้างเอกสารเอกซ์เอ็มแอลที่ต้องการจับคู่ จึงจะสามารถทำการจับคู่ได้ วิทยานิพนธ์นี้นำเสนอวิธีการในการจับคู่อิลิเมนต์ระหว่างเอกสารเอกซ์เอ็มแอลสกีมาแบบกึ่งอัตโนมัติ รวมทั้งได้พัฒนาเครื่องมือเพื่อทดสอบวิธีการดังกล่าวซึ่งพิจารณาจากความหมายของชื่ออิลิเมนต์ ความคล้ายคลึงกันของชนิดของข้อมูล และความเหมือนระหว่างโครงสร้างของสกีมาผลลัพธ์ที่ได้จากการจับคู่ เป็นเซตของอิลิเมนต์ที่มีความสัมพันธ์กันระหว่างอิลิเมนต์ของโครงสร้างเอกซ์เอ็มแอลสกีมา. หลังจากการจับคู่แบบ-อัตโนมัติแล้ว ผู้ใช้งานสามารถตรวจสอบผลลัพธ์ของการจับคู่ และแก้ไข เพิ่ม หรือ ลด ผลการจับคู่ได้ผ่านการใช้งานโปรแกรมแบบกราฟิก เพื่อให้ได้ผลลัพธ์ที่เหมาะสม ผลลัพธ์ที่ได้จากการทดลองแสดงให้เห็นว่าวิธีการที่ใช้ในงานวิจัยนี้มีค่าในการตัดสินใจและ ค่าความถูกต้องสูงเมื่อเทียบกับวิธีการอื่น การใช้โปรแกรมเพื่อจับคู่อิลิเมนต์ด้วยวิธีการที่นำเสนอในวิทยานิพนธ์นี้จะทำให้ผู้ที่ต้องการใช้เอกสารเอกซ์เอ็มแอลร่วมกันสามารถหาอิลิเมนต์ที่มีความสัมพันธ์กัน ได้ง่ายและรวดเร็วมากขึ้น ลดความสับสนและลดระยะเวลาเมื่อเทียบกับการจับคู่ด้วยตนเองทั้งหมด

Nuttakan Amarintrarak. 2009. **Semi Automatic XML Schema mapping : SAXM.**

Master of Engineering Thesis in Computer Engineering, Graduate School,
Khon Kaen University.

Thesis Advisor: Asst. Prof. Dr. Kanda Saikaew

ABSTRACT

Currently XML has become widely essential in data integration technology because XML is flexible and portable across all platforms. However, one of important problems in XML data interchange arises when users share or interchange using XML documents from heterogeneous data sources. Users need to manually map related elements between XML documents. This task is time consuming. Moreover, it can be complicated as XML documents have a wider depth and more multiple data types.

This thesis proposes a new method called Semi-Automatic XML schema Mapping (SAXM). Semantic similarity is firstly used to separate unmatchable nodes. Structural similarity and data type compatibility are then applied to ascertain the matching of two nodes. The output of SAXM is a set of similarity relationship values between two XML schemas elements. Moreover, we can add, edit or delete the mapping results on GUI to obtains an appropriated result. Experimental results show that our proposed method provides the highest recall and comparable precision among well-known approaches. The semi-automatic mapping system which is proposed in this thesis can make XML mapping task take users less time and more convenience.