

ระบบสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบโดยใช้เทคโนโลยี XML

วราพร จิระพันธุ์ทอง

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

E-mail address: waraporn@it.dpu.ac.th

บทคัดย่อ

ปัจจุบันระบบสารสนเทศได้กลายมาเป็นพื้นฐานของการทำงานในองค์กร เช่น การเก็บข้อมูล การประมวลผลและการเก็บข้อมูล ข้อมูลจำนวนมากถูกจัดเก็บในรูปแบบของเท็กซ์ไฟล์ สิ่งที่เกิดขึ้นตามมาคือแม้ว่าเอกสารเหล่านั้นได้ถูกรวบรวมและจัดเก็บอยู่ในระบบและมีความเป็นระเบียบ การสืบค้นข้อมูลบนเอกสารเท็กซ์ไฟล์ ไฟล์จำนวนมากกลายเป็นงานที่ใช้เวลามากและเกิดความผิดพลาดได้บ่อยครั้ง ทั้งนี้โครงสร้างของข้อมูลในเอกสารต่างๆ มีความหลากหลายทำให้การเรียกค้นข้อมูลมีความยุ่งยากมากยิ่งขึ้นและเกิดความผิดพลาดได้ง่ายในการทำงาน บทความนี้เป็นกรนำเสนอต้นแบบของการค้นหาข้อมูลที่ถูกจัดเก็บในเท็กซ์ไฟล์ โดยมีการแปลงและสร้างข้อมูลที่มีลักษณะเป็นเท็กซ์ (text) ให้อยู่ในรูปแบบของเอกสาร XML [5] ซึ่งใช้เป็นภาษาที่สนับสนุนการแลกเปลี่ยนข้อมูลข่าวสารระหว่างกันโดยไม่จำกัดแพลตฟอร์มของระบบสารสนเทศ และการเรียกค้นเอกสารเหล่านั้นในระบบโดยประยุกต์ใช้งาน XQuery [6]

Abstract

Currently, an information system becomes a fundamental system for business processes in an organisation such as data collection, processing, and storage. Apparently, data is recorded in term of text files. Even though those files are well-defined and organized, the retrieval of data on those text files is still considering as a tedious and erroneous task. Particularly, it occurs that those files are documented by using various schemas. That will lead more difficulties and errors to the business processes. In this paper, the model for data retrieval on a large number of text-based documents is presented. The model includes a module for automatically transforming a text-based document and generating a document based on XML, which is a language supporting the information exchange on different platforms. Also, the work is applied with XQuery to support the retrieval on XML-based documents.

คำสำคัญ

เอ็กซ์เอ็มแอล, เอ็กซ์คิวรี่, การสืบค้นข้อมูล, การสืบค้นเอกสาร, XML, XQuery, Information Retrieval, Document

Retrieval

1. บทนำ

ระบบสารสนเทศบางครั้งถูกมองว่าเป็นระบบจัดเก็บและค้นคืนสารสนเทศ (Information and Retrieval System หรือใช้คำย่อว่า ISARS) หรือเรียกอีกชื่อว่า ระบบค้นคืนสารสนเทศ (Information Retrieval System หรือ IRS) โดยมีหลักการสำคัญ คือ การค้นคืน (Retrieval) หมายถึง การได้รับสิ่งที่ต้องการกลับคืนมา การค้นคืนสารสนเทศ (Information retrieval) จึงเป็นการกระทำใด ๆ ที่คัดเลือกสารสนเทศจากแหล่งเก็บเพื่อให้ได้รับสารสนเทศตามที่ต้องการ ซึ่งอาจเป็นข้อมูล หรือ รายการเอกสารซึ่งบรรจุเนื้อหาที่ต้องการ [1, 4] หลักการสำคัญของการค้นคืนสารสนเทศ คือ การค้นหาและนำสารสนเทศที่ตรงตามความต้องการ ส่งให้แก่ผู้ใช้อย่างรวดเร็วทันการ ระบบใดก็ตามที่ได้รับการออกแบบเพื่ออำนวยความสะดวกในการค้นข้อมูลจึงถูกเรียกว่า ระบบค้นคืนสารสนเทศ ตัวอย่างเช่น บัตรรายการ สิ่งพิมพ์ดรรชนี เป็นต้น ปัจจุบันลักษณะของข้อมูลปรากฏอยู่ในหลายรูปแบบ ตั้งแต่ข้อมูลที่อยู่ในรูปของเท็กซ์ (Text) ภาพ เสียง และวิดีโอ บทความนี้ได้นำเสนอแนวคิดและการพัฒนาต้นแบบของการสืบค้นข้อมูลภายในระบบสารสนเทศหนึ่งๆ ที่ได้ถูกจัดเก็บและเรียกใช้ในรูปแบบของเท็กซ์โดยมีจุดประสงค์เพื่อสนับสนุนการเรียกใช้ข้อมูลที่ถูกต้อง ให้มีความสะดวกรวดเร็ว มีประสิทธิภาพ ทำให้ลดความซ้ำซ้อนและความผิดพลาดในการค้นข้อมูล ทั้งนี้โดยทั่วไปการค้นหาข้อมูลจากเอกสารต่างๆ ต้องอาศัยการเปิดไฟล์ขึ้นมา แล้วค่อยค้นหาข้อมูลที่ต้องการ อาจจะเป็นการอ่านทีละบรรทัด หรือสุ่มข้อมูลที่ต้องการอ่าน หรือใช้ฟังก์ชันพื้นฐานของโปรแกรมทั่วไปช่วย อย่างเช่น ฟังก์ชันการค้นหา (find)

อย่างไรก็ดีปัญหาที่ตามมา คือ ในการค้นหาข้อมูลนั้นผู้ค้นหาข้อมูลระบุไฟล์ที่ค้นหาคลาดเคลื่อนก็อาจพลาดข้อความหรือเอกสารที่ต้องการ และสูญเสียเวลามากในกรณีที่มีเอกสารที่เกี่ยวข้องเป็นจำนวนมาก และสำหรับกรณีใช้ฟังก์ชัน find ผลลัพธ์ของการค้นหาที่ได้ก็อาจจะมีมากเกินไปเกินความต้องการหรือไม่ตรงกับความต้องการเลย เนื่องจากว่าผู้ใช้ระบุค่าที่ใช้ในการค้นหาไม่เจาะจงพอ

ในหัวข้อถัดไปเป็นการนำเสนอโครงสร้างระบบสำหรับสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบโดยประยุกต์ใช้เทคโนโลยี XML ในหัวข้อที่สามเป็นการนำเสนอเครื่องมือสนับสนุนระบบสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบโดยใช้ XML การประเมินผลการใช้งานระบบสำหรับสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบด้วยเครื่องมือที่ถูกพัฒนาขึ้น XML ถูกอธิบายในส่วนที่สี่

2. โครงสร้างระบบสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบโดยใช้เทคโนโลยี XML

ในส่วนนี้เป็นการนำเสนอต้นแบบสำหรับระบบการสืบค้นข้อมูลเอกสารหลากหลายรูปแบบด้วย XML โดยระบบประกอบด้วยองค์ประกอบหลัก 2 ส่วน คือ โมดูลสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบ XML และ โมดูลสำหรับการสืบค้นข้อมูลบนข้อมูลที่ได้ผ่านการแปลงรูปแบบเป็น XML

2.1 โมดูลสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบ XML

จากรูปที่ 1 เป็นภาพรวมของโมดูลสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบ XML ประกอบด้วยองค์ประกอบหลักดังนี้

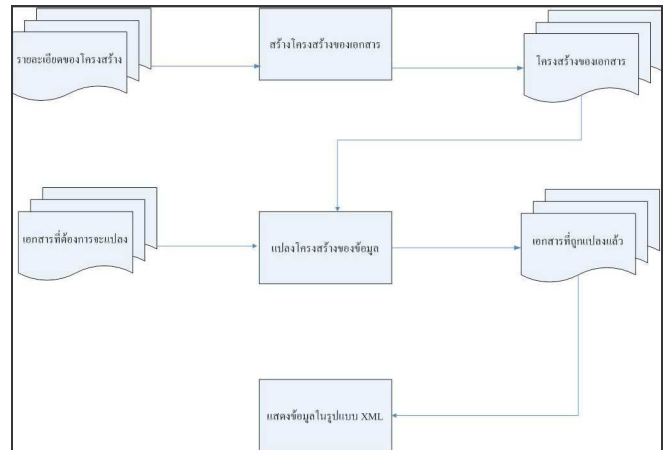
ก. **องค์ประกอบสำหรับการสร้างโครงสร้างของเอกสาร** ระบบจะสร้างโครงสร้างของเอกสารตามรายละเอียดของโครงสร้างที่ผู้ใช้งานระบุเบื้องต้น โครงสร้างของเอกสารจะถูกเก็บเป็นไฟล์ xml schema นามสกุล .xsd จากนั้นจะได้ไฟล์โครงสร้างของเอกสารเพื่อนำมาอ้างอิงในเอกสารที่ต้องการจะแปลงต่อไป

ข. **องค์ประกอบสำหรับการแปลงโครงสร้างของข้อมูล**

ระบบทำการแปลงข้อมูลจากเอกสารที่ระบุ ให้เป็นไปตามโครงสร้างที่กำหนด (ในรูปแบบของ xml) จากนั้นจะได้ไฟล์เอกสารในรูปแบบของโครงสร้างแบบ XML นามสกุล .xml

ค. **องค์ประกอบสำหรับการแสดงข้อมูลในรูปแบบของ XML**

องค์ประกอบนี้ใช้สำหรับแสดงข้อมูลของเอกสารในรูปแบบของ XML



รูปที่ 1 ภาพรวมของโมดูลสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบ XML

โมดูลสำหรับการแปลงข้อมูลให้อยู่ในรูปแบบ XML ประกอบด้วย class การทำงานหลัก 3 ส่วน คือ (ก) *Original Doc* เป็น class ที่เก็บรายละเอียดเกี่ยวกับเอกสารต้นฉบับแต่ละเอกสาร มี property คือ *Filename* เก็บข้อมูลของชื่อเอกสารต้นฉบับ, *Location* เก็บข้อมูลของที่อยู่ของเอกสารต้นฉบับ, *DocNo* เก็บข้อมูลของหมายเลขเอกสารต้นฉบับ

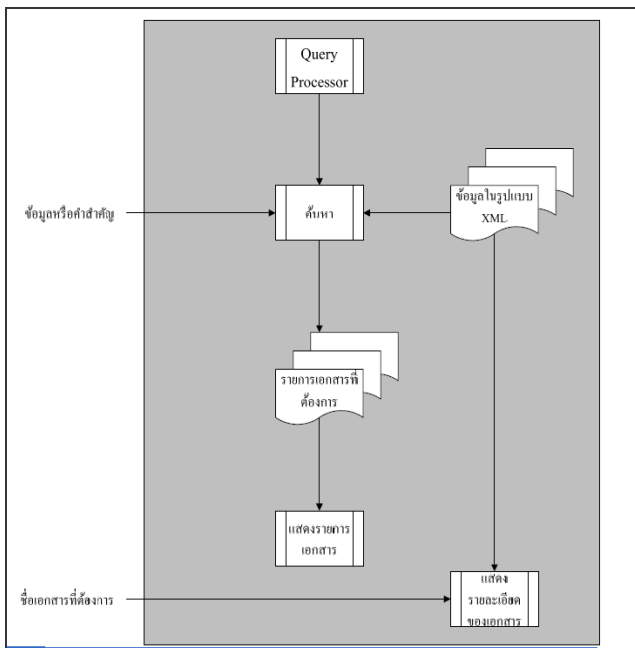
(ข) *Schema* เป็น class ที่เก็บรายละเอียดเกี่ยวกับโครงสร้างของเอกสาร XML มี property คือ *SchemaNo* เก็บข้อมูลของหมายเลขโครงสร้างของเอกสาร, *Filename* เก็บข้อมูลของชื่อของโครงสร้างเอกสาร, *Pathname* เก็บข้อมูลที่อยู่ที่ของไฟล์โครงสร้างของเอกสาร

(ค) *XML_Doc* เป็น class ที่เก็บรายละเอียดเกี่ยวกับเอกสาร XML มี property คือ *SchemaNo* เก็บข้อมูลของหมายเลขโครงสร้างของเอกสาร เพื่ออ้างอิงในไฟล์ของเอกสาร XML, *Filename* เก็บข้อมูลของชื่อไฟล์ของเอกสาร XML, *Pathname* เก็บข้อมูลที่อยู่ที่ของไฟล์เอกสาร XML, *DocNo* เก็บข้อมูลของหมายเลขเอกสารต้นฉบับ เพื่ออ้างอิงในไฟล์ของเอกสาร XML

2.2 โมเดลสำหรับการสืบค้นข้อมูลบนข้อมูลที่ได้ผ่านการแปลงรูปแบบเป็น XML

รูปที่ 2 แสดงภาพรวมของโมเดลการสืบค้นข้อมูลบนเอกสารที่ปรากฏในรูปแบบ XML ซึ่งประกอบด้วยองค์ประกอบย่อย 2 ส่วน คือ

ก. *องค์ประกอบค้นหา* องค์ประกอบนี้เป็นองค์ประกอบหลักที่ทำการประมวลการค้นหาและสืบค้นเอกสาร กระบวนการค้นหาเป็นกระบวนการที่ใช้ในการรับข้อมูล หรือคำสั่งจากผู้ใช้ระบบแล้วนำไปผ่านกระบวนการภายในของ Query Processor และข้อมูลที่อยู่ในรูปแบบของ XML เพื่อใช้ในการเรียกค้นข้อมูล หรือคำสั่งภายในเอกสาร XML ผู้ใช้ระบบในส่วนแรกจะทำหน้าที่ในการ ส่งข้อมูลหรือคำสั่งเข้าสู่ระบบเพื่อใช้เป็นส่วนประกอบในการค้นหาข้อมูล หรือสิ่งที่ใช้ระบบต้องการที่ค้นหา โดยจะส่งข้อมูล หรือคำสั่งเข้าสู่ระบบผ่าน *องค์ประกอบค้นหา*



รูปที่ 2 แสดงภาพรวมของโมเดลสำหรับการสืบค้นข้อมูลบนข้อมูลที่ได้ผ่านการแปลงรูปแบบเป็น XML

องค์ประกอบนี้จะประกอบด้วย Query Processor จะทำหน้าที่เป็นตัวกลางในการค้นหาข้อมูลต่างๆ ภายในเอกสาร ซึ่ง Query Processor นั้นจะทำหน้าที่ในการประมวลผลคำสั่งการเรียกค้นข้อมูลภายในเอกสาร XML ซึ่งข้อมูลได้ผ่านการแปลงให้อยู่ในรูปแบบ XML จากโมเดลสำหรับการแปลงข้อมูลให้อยู่

ในรูปแบบ XML ผลลัพธ์ที่ได้จะถูกแสดงเป็นรายการของเอกสารที่มีความเกี่ยวข้องกับคำสำคัญที่ผู้ใช้ระบบได้ใส่เข้ามา ข. *องค์ประกอบแสดงรายละเอียดของเอกสาร* องค์ประกอบนี้แสดงเอกสารต้นฉบับซึ่งถูกเลือกจากรายการที่เกี่ยวข้องกับคำที่ค้นหา ซึ่งเป็นผลลัพธ์จากองค์ประกอบค้นหา การออกแบบโครงสร้างภายในโมเดลสำหรับการสืบค้นข้อมูลบนข้อมูลที่ได้ผ่านการแปลงรูปแบบเป็น XML นั้นประกอบไปด้วย class หลักดังต่อไปนี้ คือ (ก) *Query Processor* ผู้แต่งได้ประยุกต์ใช้ Package ของ Saxon [2] ที่ใช้ในการเรียกค้นข้อมูลภายในเอกสาร XML ซึ่งภายใน Query Processor นี้ จะประกอบไปด้วยคลาสหลายๆ ตัวมาประกอบกัน อย่างเช่น *DynamicQueryContext*, *StaticQueryContext* เป็นต้น

(ข) *Query* เป็นคลาสที่เก็บรายละเอียดเกี่ยวกับคำสั่งการเรียกค้นแต่ละตัวเพื่อทำการเรียกค้นข้อมูลซึ่งประกอบด้วย *Query Statement* ใช้ในการเก็บข้อมูลหรือเงื่อนไขที่ใช้ในการเรียกค้นเอกสาร, *Query No.* ใช้ในการเก็บหมายเลขของข้อมูลหรือเงื่อนไขที่ใช้ในการเรียกค้นเอกสารแต่ละตัว

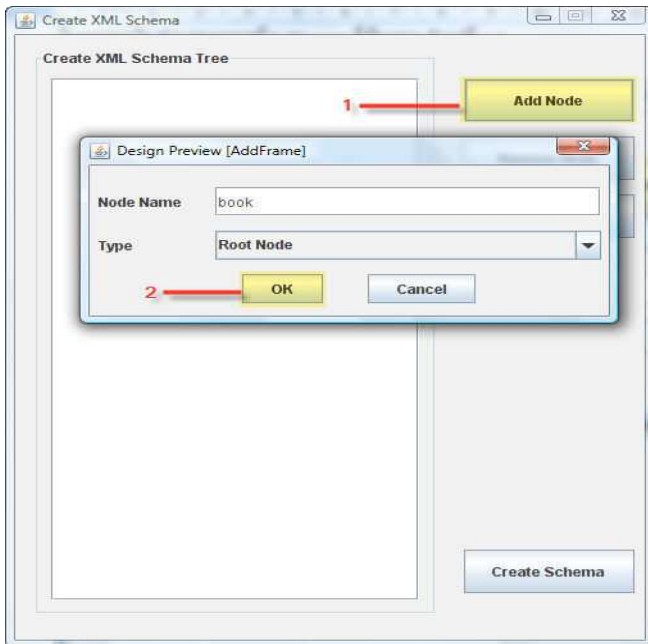
(ค) *Result* เป็นคลาสที่เก็บรายละเอียดของผลลัพธ์ที่ได้จากการเรียกค้นเอกสารแล้ว ซึ่งประกอบด้วย *Query No.* ใช้ในการเก็บหมายเลขของข้อมูลหรือเงื่อนไขที่ใช้ในการเรียกค้นเอกสารแต่ละตัว, *Data* ใช้ในการเก็บรายการของเอกสารที่ค้นเจอ

(ง) *Document* เป็นคลาสที่ใช้ในการเก็บรายละเอียดต่างๆ ของเอกสารที่ระบบมี ซึ่งประกอบด้วย *File Name* ใช้ในการเก็บชื่อของเอกสารที่มีอยู่ในระบบ, *Path Name* ใช้ในการเก็บที่อยู่ของเอกสารที่อยู่ภายในระบบ

(จ) *Display_Document* เป็นคลาสที่ใช้ในการเก็บรายละเอียดต่างๆ ของเอกสาร เพื่อนำไปใช้ในการแสดงผล ซึ่งประกอบด้วย *File Name* ใช้ในการเก็บชื่อของเอกสาร ที่จะนำไปใช้ในการแสดงผล, *Path Name* ใช้ในการเก็บที่อยู่ของเอกสาร ที่จะนำไปใช้ในการแสดงผล

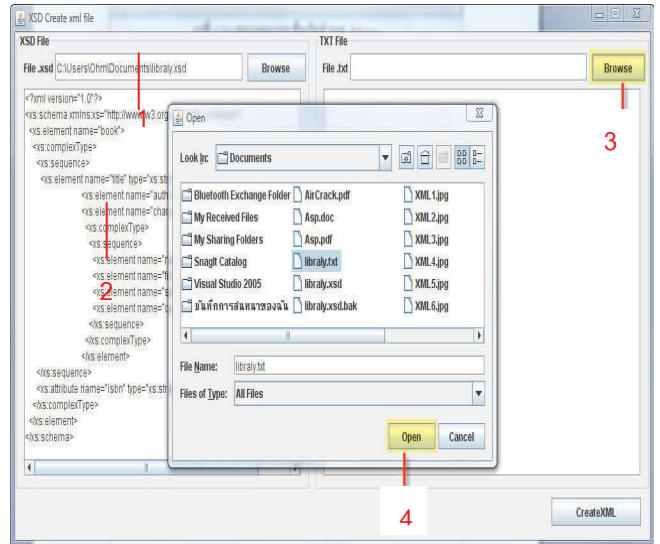
3. เครื่องมือสนับสนุนระบบสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบโดยใช้เทคโนโลยี XML

สำหรับการใช้เครื่องมือสนับสนุนระบบสืบค้นข้อมูลบนเอกสารหลากหลายรูปแบบด้วย XML ส่วนแรกเป็นการสร้างโครงสร้างของเอกสารเพื่อเป็นต้นแบบของเอกสาร XML ที่ต้องการ ตัวอย่างหน้าจอของเครื่องมือในส่วนนี้แสดงในรูปที่ 3 มีส่วนสำหรับการระบุเพื่อสร้าง element ใหม่ของเอกสาร XML ที่ต้องการ (หมายเลข 1) และส่วนที่ใช้ระบุชื่อและชนิดของ element (หมายเลข 2) การสร้าง element สามารถระบุเป็น root element, parent element, และ child element ได้ ผลลัพธ์ที่ได้จะปรากฏอยู่ในรูปของไฟล์ XML Schema



รูปที่ 3 ตัวอย่างหน้าจอสำหรับสร้างโครงสร้างเอกสาร XML

นอกจากนี้เครื่องมือยังประกอบด้วยส่วนหลักที่สอง คือ ส่วนที่ทำการสร้างเอกสาร XML จากเอกสารที่เป็นต้นฉบับซึ่งปรากฏอยู่ในรูปแบบของเท็กซ์ไฟล์ โดยกำหนดให้มีรูปแบบโครงสร้างที่อ้างอิงจาก XML Schema รูปที่ 4 เป็นตัวอย่างหน้าจอของเครื่องมือส่วนของการสร้างเอกสาร XML มีส่วนที่ระบุ XML Schema ที่เป็นโครงสร้างของเอกสารที่ต้องการสร้างเป็น XML (หมายเลขที่ 1) รายละเอียดของโครงสร้างจะปรากฏขึ้นเพื่อแสดงความซับซ้อนของโครงสร้าง (หมายเลข 2) จากนั้นระบุเอกสารต้นฉบับซึ่งปรากฏอยู่ในรูปของเท็กซ์ไฟล์ (หมายเลข 3 และ 4)



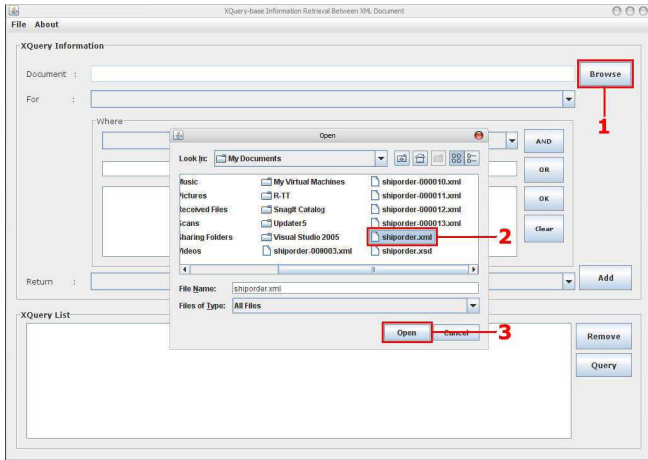
รูปที่ 4 ตัวอย่างหน้าจอสำหรับแปลงเอกสาร XML

ส่วนหลักที่สาม คือ การจัดการการสืบค้นข้อมูลบนเอกสาร ตัวอย่างของหน้าจอหลักในส่วนนี้อยู่ในรูปที่ 5 ประกอบไปด้วย 2 ส่วนหลัก คือ ส่วนที่ 1 *XQuery Information* คือ ส่วนที่ใช้ในการเลือกไฟล์ที่ต้องการจะค้นหาข้อมูล, เงื่อนไขที่ใช้ในการค้นหา, และสิ่งที่ต้องการจะให้เห็น โดยการระบุเงื่อนไขนั้น เครื่องมือจะทำการเข้าไปอ่านเอกสาร XML Schema ตามการอ้างอิงของเอกสาร XML ที่ถูกระบุเพื่อสืบค้น (หมายเลข 1 2 และ 3) โดยสามารถแสดงโครงสร้างของเอกสาร XML ได้โดยอัตโนมัติ ทำให้ผู้ใช้สามารถระบุเงื่อนไขการสืบค้นได้สะดวกและถูกต้องโดยอ้างอิงโครงสร้างของเอกสาร XML

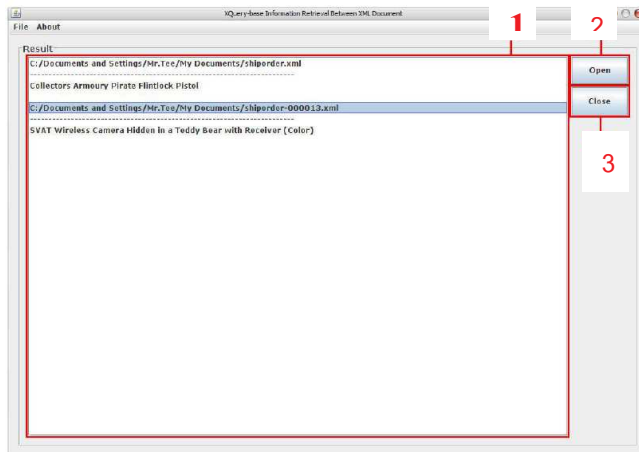
ส่วนที่ 2 *XQuery List* คือ ส่วนที่ใช้ในการแสดงรายการของเอกสารที่ต้องการจะค้นหา และในการค้นหาแต่ละครั้งสามารถที่จะทำการค้นหาได้แบบที่หลายๆเอกสารด้วยเงื่อนไขของการค้นหาหนึ่งๆ ซึ่งผลลัพธ์ที่ถูกค้นหาเจอตามเงื่อนไขจะถูกนำมาแสดง แต่การค้นหาที่ไม่ตรงตามเงื่อนไข ก็จะไม่ถูกนำมาแสดงเช่นกัน ผู้ใช้สามารถที่จะทำการเปิดอ่านเอกสารที่ค้นเจอข้อมูลตามเงื่อนไขได้จากการคลิกเลือกที่รายการที่ต้องการ

ส่วนหลักที่สี่ คือ เครื่องมือสามารถเรียกรายละเอียดของทั้งเอกสารในรูปแบบของต้นฉบับหรือ XML ก็ได้ รูปที่ 6 แสดงตัวอย่างของรายการเอกสารที่ค้นเจอ (หมายเลขที่ 1) ผู้ใช้

สามารถเรียกดูเอกสารต้นฉบับและเอกสารในรูปแบบ XML ได้ (หมายเลขที่ 2 และ 3)



รูปที่ 5 ตัวอย่างหน้าจอสำหรับสืบค้นข้อมูลเอกสาร XML



รูปที่ 6 ตัวอย่างหน้าจอแสดงรายการเอกสารที่เกี่ยวข้องกับคำที่สืบค้น

4. การประเมินและวิเคราะห์ผล

ผู้แต่งได้ออกแบบและสร้างรูปแบบของการทดลองระบบ การสร้างสถานการณ์จำลองขึ้นโดยจำลองเหตุการณ์การทำงานในระบบสารสนเทศของและมีความต้องการสืบค้นข้อมูล ได้แก่ ฝ่ายทะเบียนนักศึกษาของสถาบันอุดมศึกษาแห่งหนึ่ง ศูนย์พัฒนาเทคโนโลยีสารสนเทศของสถาบันอุดมศึกษาแห่งหนึ่ง แผนกจัดซื้อของบริษัทเอกชนแห่งหนึ่ง แผนกพัฒนาระบบสารสนเทศของบริษัทเอกชนแห่งหนึ่ง รวมเป็น 4 สถานการณ์จำลอง ในแต่ละสถานการณ์จำลองเหล่านี้มีผู้ร่วมทดลองและประเมินระบบ 5-10 คน ผู้ร่วมทดลองและประเมินผลได้ทำการสืบค้นตามเงื่อนไขที่แตกต่างกันโดยพิจารณาจากบทบาทและหน้าที่ประจำของผู้ร่วมทดลอง ผู้ร่วมทดลองและประเมินผลแต่ละท่านทำการสืบค้นตามเงื่อนไขที่

แตกต่างกันท่านละ 40 เงื่อนไข นอกจากนี้ผู้ร่วมทดลองและประเมินผลแต่ละท่านทำการค้นหาเอกสารที่ต้องการสองครั้ง คือ ครั้งแรกด้วยตนเองหรือใช้เทคนิคที่มีอยู่เดิม และครั้งที่สองโดยอาศัยเครื่องมือที่ได้ถูกพัฒนาขึ้น ผลลัพธ์ของการค้นหาแต่ละครั้ง ถูกพิจารณาโดยประยุกต์ใช้หลักการของ Precision และ Recall [3]

Precision หมายถึงอัตราส่วนระหว่างจำนวนข้อมูลที่ตรงความต้องการและจำนวนข้อมูลทั้งหมดที่สืบค้นได้จากฐานข้อมูล Recall หมายถึงอัตราส่วนระหว่างจำนวนข้อมูลที่สืบค้นได้และจำนวนข้อมูลทั้งหมดที่มีในฐานข้อมูล กล่าวคือจำนวนของเอกสารที่เมื่อผู้ร่วมทดลองและประเมินผลค้นเจอ ถูกเปรียบเทียบกับจำนวนเอกสารที่เกี่ยวข้องกับเงื่อนไขสำหรับการค้นหานั้น

$$\text{Precision} = \frac{b}{m}$$

$$\text{Recall} = \frac{b}{h}$$

โดย b = จำนวนของเอกสารที่เมื่อผู้ร่วมทดลองและประเมินผลค้นเจอและเป็นเอกสารที่เกี่ยวข้องกับเงื่อนไขสำหรับการค้นหานั้นอย่างถูกต้อง; m = จำนวนของเอกสารที่ได้จากการค้นหานั้น; h = จำนวนของเอกสารที่เกี่ยวข้องกับเงื่อนไขสำหรับการค้นหานั้นในฐานข้อมูล

จากตารางที่ 1 ผลการทดลองครั้งที่หนึ่งซึ่งผู้ร่วมทดลองและประเมินผลได้ทำการสืบค้นตามเงื่อนไขที่แตกต่างกันด้วยตนเองหรือใช้เทคนิคที่มีอยู่เดิมพบว่า ค่าเฉลี่ยของ precision ของทั้งสี่กลุ่ม (0.65, 0.55, 0.58, และ 0.6 ตามลำดับ) มีค่าเฉลี่ยรวม 0.595 และค่าเฉลี่ยของ recall ของทั้งสี่กลุ่ม (0.65, 0.58, 0.68, และ 0.55 ตามลำดับ) มีค่าเฉลี่ยรวม 0.615 ในขณะที่ผลการทดลองครั้งที่สอง ผู้ร่วมทดลองและประเมินผลได้ทำการสืบค้นตามเงื่อนไขที่แตกต่างกันโดยอาศัยเครื่องมือที่ได้ถูกพัฒนาขึ้น ค่าเฉลี่ยของ precision ของทั้งสี่กลุ่ม (0.8, 0.78, 0.7, และ 0.85 ตามลำดับ) มีค่าเฉลี่ยรวม 0.7825 และค่าเฉลี่ยของ recall ของทั้งสี่กลุ่ม (0.7, 0.8, 0.9, และ 0.77 ตามลำดับ) มีค่าเฉลี่ยรวม 0.7925 ทำให้สามารถสรุปได้ว่าการ

ประยุกต์ใช้ระบบสนับสนุนการสืบค้นที่มีเครื่องมือประยุกต์ใช้เทคโนโลยี XML สามารถสนับสนุนงานการสืบค้นได้อย่างถูกต้อง รวดเร็ว และมีประสิทธิภาพมากกว่าการไม่มีระบบสนับสนุนนี้หรือการใช้เทคนิคแบบดั้งเดิม

ตารางที่ 1 ผลการทดลองครั้งที่หนึ่งและสอง

ผลการทดลองครั้งที่	precision		recall	
	1	2	1	2
กลุ่มที่หนึ่ง	0.65	0.8	0.65	0.7
กลุ่มที่สอง	0.55	0.78	0.58	0.8
กลุ่มที่สาม	0.58	0.7	0.68	0.9
กลุ่มที่สี่	0.6	0.85	0.55	0.77

ค่าเฉลี่ยของ precision และ recall ที่ได้จากการทดลองในสถานการณ์จำลองทั้งสี่กลุ่ม แต่ละกลุ่มประกอบด้วยจำนวนผู้ทดลองที่แตกต่างกัน กลุ่มที่หนึ่งฝ่ายทะเบียนนักศึกษาของสถาบันอุดมศึกษาแห่งหนึ่งมีผู้ร่วมทดลอง 6 คน ผู้ทดลองมีความถนัดทางด้านเทคโนโลยีสารสนเทศปานกลาง กลุ่มที่สองศูนย์พัฒนาเทคโนโลยีสารสนเทศของสถาบันอุดมศึกษาแห่งหนึ่งมีผู้ร่วมทดลอง 8 คน ผู้ทดลองมีความถนัดทางด้านเทคโนโลยีสารสนเทศมาก กลุ่มที่สามแผนกจัดซื้อของบริษัทเอกชนแห่งหนึ่งมีผู้ร่วมทดลอง 10 คน ผู้ทดลองมีความถนัดทางด้านเทคโนโลยีสารสนเทศน้อย กลุ่มที่สี่แผนกพัฒนาระบบสารสนเทศของบริษัทเอกชนแห่งหนึ่ง ผู้ทดลองมีความถนัดทางด้านเทคโนโลยีสารสนเทศสูง ผลทดลองพบว่าค่าเฉลี่ยของ precision และ recall ในกลุ่มที่มีความถนัดทางด้านเทคโนโลยีสารสนเทศสูง มีค่าใกล้เคียงกันมาก ขณะที่ค่าเฉลี่ยของ precision และ recall ในกลุ่มที่มีความถนัดทางด้านเทคโนโลยีสารสนเทศปานกลางและน้อย มีค่าต่างกัน

อย่างไรก็ดี ค่าเฉลี่ยของ precision ของทั้งสี่กลุ่ม (0.8, 0.78, 0.7, และ 0.85 ตามลำดับ) มีค่าเฉลี่ยรวม 0.7825 และค่าเฉลี่ยของ recall ของทั้งสี่กลุ่ม (0.7, 0.8, 0.9, และ 0.77 ตามลำดับ) มีค่าเฉลี่ยรวม 0.7925 สามารถสรุปได้ว่าการประยุกต์ใช้ระบบสนับสนุนการสืบค้นที่มีเครื่องมือประยุกต์ใช้เทคโนโลยี XML สามารถสนับสนุนงานการสืบค้นได้อย่าง

ถูกต้อง รวดเร็ว และมีประสิทธิภาพมากกว่าการไม่มีระบบสนับสนุนนี้หรือการใช้เทคนิคแบบดั้งเดิม

5. บทสรุป

การค้นหาข้อมูลจากเอกสารจำนวนมาก และมีโครงสร้างของข้อมูลในเอกสารหลากหลาย กลายมาเป็นงานที่ใช้เวลามากและเกิดความผิดพลาดได้บ่อยครั้ง แม้ว่าได้ถูกรวบรวมและจัดเก็บไว้อย่างเป็นระเบียบ การเรียกค้นข้อมูลที่ปรากฏอยู่ในไฟล์ข้อมูลจึงยังคงเป็นเรื่องที่ยุ่งยากและใช้เวลานาน นอกจากนี้สามารถเกิดความผิดพลาดได้ บทความนี้เป็นกรนำเสนอต้นแบบของการค้นหาข้อมูลที่ถูกจัดเก็บในไฟล์ โดยมีการแปลงข้อมูลที่มีลักษณะเป็นเท็กซ์และสร้างให้อยู่ในรูปแบบของเอกสาร XML และสามารถสนับสนุนการสืบค้นที่เป็นระบบมากขึ้นโดยประยุกต์ใช้งาน XQuery เครื่องมือได้ถูกสร้างเพื่อสนับสนุนความคิดต้นแบบ การทดลองได้ถูกสร้างและพัฒนาเพื่อสังเกตและรวบรวมข้อมูลเปรียบเทียบการสืบค้นตามเงื่อนไขที่ต่างกันด้วยตนเองหรือใช้เทคนิคที่มีอยู่เดิมพบว่า การสืบค้นที่อาศัยเครื่องมือที่ได้ถูกพัฒนาขึ้น จากการทดลอง ระบบสามารถสนับสนุนงานการสืบค้นได้อย่างถูกต้อง รวดเร็ว และมีประสิทธิภาพโดยผู้ใช้งานไม่จำเป็นต้องเป็นผู้เชี่ยวชาญทางด้านเทคโนโลยีสารสนเทศ

6. กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จได้ด้วยความร่วมมือของนายวิรัช แม้นญาติและ นายธีรชัย รัตติกุล ผู้ร่วมทดลองทุกท่าน รวมถึงหน่วยงานต้นสังกัดของผู้ร่วมทดลอง จึงขอขอบคุณมา ณ โอกาสนี้

7. เอกสารอ้างอิง

[1] Cleveland & Cleveland. Introduction to indexing and abstracting. 2nd Ed. Colorado: Libraries Unlimited, 1990.

[2] Sourceforge Saxon. 2008. The XSLT and XQuery Processor. [Online] Available: <http://saxon.sourceforge.net/>

[3] Strzalkowski, Tomek. Natural language information retrieval. Information Processing & Management. 3(10): May/June, 1995. P. 397-417.

[4] Wiley, Deborah Lynne. Can the traditional abstracting and indexing services survive?. Database 17, Dec, 1994. P. 18-24.

[5] XML. 2008. Extensible Markup Language. [Online] Available: <http://www.w3.com/XML/>

[6] XQuery. 2007. XQuery 1.0: An XML Query Language. [Online] Available: <http://www.w3.org/TR/xquery/>.